

**Protein Interaction Analysis of a Genetic Engineered,  
Stem Cell-Derived Model of Cardiomyopathy**

by

**Juan Alfredo Pérez-Bermejo**

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biological and Medical Informatics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Protein Interaction Analysis of a Genetic Engineered,  
Stem Cell-Derived Model of Cardiomyopathy

by

Juan Alfredo Pérez-Bermejo

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biological and Medical Informatics

in the

GRADUATE DIVISION

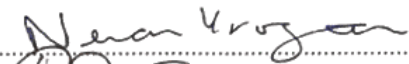
of the

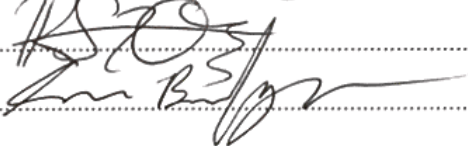
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:



Chair





Committee in Charge

Copyright 2017

By

Juan Alfredo Pérez-Bermejo

A mis padres.

# Acknowledgements

I thank the “la Caixa” Foundation for their support and advice. Without them this would have never happened. I also thank the American Heart Association for their continued trust in my project. I thank UCSF and the Gladstone Institutes for giving me this opportunity and providing a work environment that allowed me to grow both scientifically and personally.

Thanks to my advisor Bruce Conklin for his tireless faith in me, which translated into hours of advice, aimless conversation, and some arguments. I also thank my co-mentor Nevan Krogan for the rare but productive meeting hours. Their supervision style allowed me to explore, get lost and find myself again in Science multiple times, learning many unique lessons on the way. I also thank my thesis committee members, Sourav Bandyopadhyay and Ryan Hernandez, for their advice and directness.

I have been lucky to enjoy the help and company of the members of the Conklin and Krogan labs, all of whom I thank thoroughly. I especially would like to thank Po-Lin So and Luke Judge for being a role model of scientific wisdom and real-life sensitivity that I aspire to emulate someday. I also thank Christina Jensen and Annie Truong for their exceptional help, and for the invaluable lessons they taught me about myself and about teamwork. I must thank Sara ‘Sarita’ Prophet and Damon Williams for putting up with me as a novice mentor. Thanks to Ruth Huttenhain, Jeffrey Johnson and especially Robyn Kaaake for their infinite patience and advice. I also owe a lot to Mercedes and Julio, Gladstone janitors and night companions.

I very deeply thank Graham Heimberg, Kyle Barlow and James Webber for their continued friendship, that prevented me from drowning in the waters of American culture, and for accepting me in their families as one of their own. I feel truly lucky. I also thank Diego Garrido, carnal y compañero, for all the conversations and the Mondays in the Sun.

I must give thanks to my Spanish family in San Francisco. To Ana Rodríguez, first friend and support. To Luis Luna, master of Science and of laughs. And most especially thanks to Marta Franco, friend and headache through all these years, partner in joy and in struggles, and always great confidant. Thank you.

Thanks to my Spanish Friends in Spain, a beacon of stability and support. Thanks especially to my cousins Pedro and Paco, lifetime partners in crime, and to Miriam Carrión.

Very special thanks to Victoria Taffe for her indefatigable love, patience and support. For her enthusiastic admiration for my work, which helped me push through difficulties. And for the countless coquiadventures exploring San Francisco and part of world. I also apologize for all the nights we had to spend in lab, ‘for Science’. I am very lucky to have you in my life.

Finally, it is most important I thank my family in Spain. I thank my brothers, Javi and José, for always making me feel like if I was there with them, and I apologize for not actually being there for them through all these years of growth. Most importantly, I thank my parents, José Pérez Méndez and Francisca Bermejo Molina, for their unconditional, lifetime love. For the tales in bed, the sand and the sea, the family trips, los Reyes Magos, the almond trees, the evenings of homework together, the poetry, even for the Sunday Mass; for each and every laugh and tear. Any contributions and sacrifices anyone has done for this thesis dwarf compared to those made by them. I will never find enough words to express how thankful I am, and how sorry I feel that I once left to go so far. You made me the person I am proud to be. Thank you. This thesis is entirely dedicated to you.

*“We shall not cease from exploration,  
And the end of all our exploring  
Will be to arrive where we started  
And know the place for the first time.”*

*--T.S. Eliot, Little Gidding*

# Contributions

Most of the chapters in this dissertation contain unpublished material. The main exception is Chapter 3, which is a reprint of the following published article: *Judge LM, Perez-Bermejo JA, Truong A, Ribeiro AJS, Yoo JC, Jensen CL, Mandegar MA, Huebsch N, et al (2017) BAG3 chaperone complex maintains cardiomyocyte function during proteotoxic stress. JCI Insight 2. doi:10.1172/jci.insight.94623.*

In addition, the generation of the BAG3<sup>C151R</sup> (which then was used to produce BAG3<sup>C151R-3xFLAG</sup>) in Chapter 2 was published as a fragment in: *Miyaoka Y, Chan AH, Judge LM, Yoo J, Huang M, Nguyen TD, Lizarraga PP, So P-L, Conklin BR (2014) Isolation of single-base genome-edited human iPS cells without antibiotic selection. Nat Methods 11, 291–3.*

Other people helped the work in this dissertation by providing help or performing some of the experiments described. Luke Judge, assisted by Jennifer Yoo and Ann Truong made the BAG3<sup>P209L</sup> cell line. Christina Jensen and Ann Truong provided assistance in making all the 3xFLAG-containing cell lines from Chapter 5. Christina Jensen generated the western blot image used in Chapter 6.1. Sarah Prophet helped with performing the affinity purification steps in Chapter 5. Jennifer Rauch in Jason Gestwicki's lab performed the FCPIA analysis and luciferase refolding assays in Chapter 5. Robyn Kaake and Jeffrey Johnson ran the samples from Chapter 5 and 6 on the Mass Spectrometry instrument.

# Abstract:

## **Protein Interaction Analysis of a Genetic Engineered, Stem Cell-Derived Model of Cardiomyopathy**

Juan Alfredo Pérez-Bermejo

Genetic association studies have yielded a wealth of information on specific variants that lead to the development of disease, but our mechanistic understanding for most of them is lacking. Unbiased, comparative study of protein-protein interactions (PPI) for protein coding variants is a very promising approach to fill this gap. To yield data that can inform us on tissue-specific pathologies (e.g. cardiovascular, neurological disease) we need to study PPIs in the cell types affected by the disease. In a similar fashion, studying proteins expressed endogenously would allow us to overcome issues related to gene expression or dose-dependent protein function, which is a feature of many of these pathologies.

Here I present my work trying to unravel the mechanism mutations in the BAG3 chaperone gene lead to disease. To address this, I used a combination of induced pluripotent stem (iPS) cell technology (to generate cardiac myocytes), genome engineering (to induce specific disease-modifying mutations and protein fusions) and affinity purification coupled to mass spectrometry (to compare PPIs between variants). The results show that disease-related variants of BAG3 display a different profile of protein partners, some of them cardiac-specific, pointing towards interactions that potentially underlie the disease mechanism and tissue



specificity. We also observed a cell culture phenotype for BAG3 deficiency and for some of the protein-coding variants, which sets the ground for the exploration of the role of BAG3 interactions in disease.

Overall, this piece of work proves the value of studying cardiac disease-related genetic variability using a disease-relevant iPS-derived model and genome engineering to explore multiple variants. In particular, we provide new clues on the role of BAG3 in the heart, and how that role is compromised by disease-related mutations. We hope that the knowledge gained on BAG3 genetics and mechanism will be useful for the development of therapeutic strategies in the future.

# Table of Contents

<b>Chapter 1 : Introduction .....</b>	<b>1</b>
1.1 Protein interactions allow us to unravel the functioning of the cell .....	3
1.2 BAG3 variants cause heart disease: protein quality control in the heart .....	7
1.3 Stem cell technology and genome engineering enable the study of heart disease genetics .....	13
<b>Chapter 2 - Genome engineering and protein interaction mapping: new opportunities and remaining challenges .....</b>	<b>19</b>
Introduction.....	19
2.1 Limitations of current protein-protein interaction mapping datasets .....	20
2.1.1 1.-Bait protein overexpression .....	20
2.1.2 2.-Use of immortalized cell lines .....	22
2.2 Genome engineering allows us to overcome these issues.....	24
2.3 Outlook / Conclusion .....	27
<b>Chapter 3 - A BAG3 Chaperone Complex Maintains Cardiomyocyte Function During Proteotoxic Stress.....</b>	<b>31</b>
3.1 Abstract.....	32
3.2 Introduction .....	32
3.3 Results .....	34
3.4 Discussion .....	47
3.5 Materials and Methods.....	50
3.5.1 Human iPSC culture .....	50
3.5.2 TALEN and CRISPR construct design.....	51

3.5.3 Transfection of iPSCs for gene targeting .....	52
3.5.4 Isolation of modified iPSC clones .....	52
3.5.5 Karyotyping.....	53
3.5.6 Pluripotency Staining for iPSCs.....	54
3.5.7 iPS-CM Differentiation and Culture.....	54
3.5.8 Immunofluorescent staining of iPS-CMs .....	55
3.5.9 Flow cytometry.....	56
3.5.10 Sarcomere scoring in fixed iPS-CMs.....	56
3.5.11 Analysis of micro-patterned iPS-CMs.....	57
3.5.12 Live sarcomere labeling and scoring in iPS-CMs.....	58
3.5.13 Drug treatment and contractility assays in iPS-CMs.....	59
3.5.14 Resazurin viability assay in iPS-CMs .....	60
3.5.15 Western blot analysis .....	60
3.5.16 Affinity Purification – Mass Spectrometry of BAG3 interacting partners .....	62
3.5.17 Software and statistical analysis .....	64
3.5.18 Study approval.....	65
3.6 Supplementary Figures .....	66
<b>Chapter 4 - genome engineering of the endogenous copy of the BAG3 gene on induced pluripotent stem cells .....</b>	<b>77</b>
4.1 Introduction .....	77
4.2 Materials and Methods.....	78
4.2.1 iPS cell culture .....	78
4.2.2 TALEN and CRISPR targeting sequence design and cloning.....	78
4.2.3 Single nucleotide genome editing, sib-selection .....	79

4.2.4	Generating the TetOn-BAG3 <sup>3xFLAG</sup> .....	80
4.2.5	Differentiation of iPS into iPS-CM.....	80
4.2.6	Insertion of the 3xFLAG sequence in the endogenous BAG3 locus .....	80
4.2.7	On-plate freezing, thawing and genomic DNA extraction of iPS cell populations.....	81
4.2.8	iPS cell colony cloning.....	82
4.2.9	Cell genotyping.....	82
4.2.10	Karyotyping .....	84
4.2.11	Phasing assay .....	84
4.2.12	Western Blot.....	85
4.2.13	Immunofluorescence staining.....	86
4.3	Results and discussion .....	87
4.3.1	Droplet digital PCR enables identification of low frequency genome editing events and enables scarless single nucleotide genome editing.....	87
4.3.2	Transgenes inserted in the AAVS1 safe harbor locus express unevenly across cells.....	90
4.3.3	Genome editing tools allow for the insertion of long nucleotide sequences in target sites in the genome and the expression of fusion proteins at physiological levels.....	92
4.3.4	Droplet digital PCR allows for the successful phasing of chromosomal variants .....	95
4.3.5	Homologous recombination replaces kilobase-length regions of the genome and enables for the simultaneous insertion of multiple modifications .....	97
4.4	Conclusion.....	99
<b>Chapter 5 : Affinity Purification - Mass Spectrometry the functional study of two families of proteins and the comparison of disease-relevant BAG3 protein variants in an overexpression system.....</b>		<b>105</b>

5.1 Introduction .....	105
5.2 Materials and Methods.....	107
5.2.1 Gene expression data analysis.....	107
5.2.2 HEK293T cell culture and transfection.....	107
5.2.3 Affinity purification of 3xFLAG protein fusions .....	107
5.2.4 Mass Spectrometry analysis of protein samples .....	108
5.2.5 Analysis of APMS data.....	110
5.2.6 Sequence analyses .....	111
5.2.7 Western blot and immunocytochemistry.....	111
5.2.8 Flow Cytometry Protein Interaction Assay (FCPIA) .....	113
5.2.9 Luciferase Refolding Assay.....	113
5.3 Results and discussion .....	114
5.3.1 Data processing for the discrimination of true interactors from nonspecific interactions.....	114
5.3.2 Unbiased analysis of the stable interactions of the BAG family of co- chaperones reveals member-specific functions.....	119
5.3.3 Interactors of the small heat shock protein family provide new insights in the functional diversity of the co-chaperone network.....	130
5.3.4 AP-MS study of BAG3 interactors for different disease-causing mutations in an overexpression system on immortalized cells .....	140
<b>Chapter 6 - Comparative phenotypic and protein interaction analysis of disease-             associated BAG3 variants in iPS-derived cardiomyocytes.....</b>	<b>149</b>
6.1 Introduction .....	149
6.2 Materials and Methods.....	150
6.2.1 Differentiation of iPS into iPS-CM and iPS-CM cell culture .....	150
6.2.2 Flow cytometry.....	151

6.2.3 Bortezomib treatment. Contractility and viability assay .....	151
6.2.4 Affinity Purification coupled to Mass Spectrometry analysis.....	152
6.2.5 AP-MS Data Analysis .....	154
6.2.6 Western blotting.....	155
6.3 Results .....	155
6.4 Discussion .....	159
6.5 Conclusion.....	163
<b>Chapter 7 : Concluding Remarks.....</b>	<b>164</b>
<b>Bibliography .....</b>	<b>166</b>

# List of Tables

<b>Table 2.1</b> - Advantages and disadvantages of the main approaches for protein interaction analysis using AP-MS. ....	27
<b>Table 2.2</b> - Limitations of genome engineering and AP-MS technologies for the mapping of Protein-protein interactions.....	29
<b>Table 4.1</b> - Sequences of primers for genotyping reactions .....	100
<b>Table 4.2.</b> - Cell lines generated in this study.....	101
<b>Table 4.3</b> - Sequences of TALEN proteins.....	103
<b>Supplementary Table 3.1</b> – Genotyping primers used in this study. ....	76

# List of Figures

<b>Figure 1.1</b> - Affinity purification – mass spectrometry (APMS).....	4
<b>Figure 1.2</b> - Protein-protein interactions bridge the gap between genotype and phenotype...6	
<b>Figure 1.3</b> – Dilated cardiomyopathy has a strong genetic component. ....	7
<b>Figure 1.4</b> - Variants in the BAG3 co-chaperone are associated with DCM. ....	10
<b>Figure 1.5</b> - Two different approaches to disease modeling using iPS cells. ....	15
<b>Figure 1.6.</b> – Genome engineering using targeted endonucleases. ....	17
<b>Figure 2.1</b> - Comparison of the main approaches for protein interaction analysis using AP-MS. ....	26
<b>Figure 3.1</b> - Genome engineering an isogenic series of BAG3 mutations leads to sarcomeric disarray in human iPS-derived cardiomyocytes. ....	37
<b>Figure 3.2</b> - BAG3 mutations produce contractile deficits in iPS-CMs cultured on micro-patterned substrates.....	39
<b>Figure 3.3</b> - BAG3 is required to prevent severe cardiotoxicity from the proteasome inhibitor bortezomib.. ....	41
<b>Figure 3.4</b> - Genetic and pharmacologic specificity of the interaction between BAG3 and proteasome inhibitors. ....	43
<b>Figure 3.5</b> - Bortezomib induces a cardiac chaperone stress-response and increases autophagy flux, the latter of which does not require BAG3.....	45
<b>Figure 3.6</b> - Network representing cardiac-specific BAG3 protein interactions.....	47
<b>Figure 4.1</b> - Scarless introduction of single nucleotide polymorphisms in the BAG3 gene..	89
<b>Figure 4.2</b> - BAG3 <sup>3xFLAG</sup> cassette inserted in the AAVS1 safe harbor gets silenced in iPS-CM. ....	91



<b>Figure 4.3</b> - Insertion of the 3xFLAG epitope tag fusion sequence in the endogenous BAG3 gene.....	94
<b>Figure 4.4</b> – ddPCR-based assay for phasing of BAG3 variants and the 3xFLAG sequence.....	96
<b>Figure 4.5</b> – Karyotyping of the cell lines obtained that were used in further experiments.	98
<b>Figure 5.1</b> - Effect of normalization method on protein counts and on recovery of known interactions.....	116
<b>Figure 5.2</b> – The BAG family of co-chaperones.....	120
<b>Figure 5.3</b> - APMS identifies putative protein interaction partners for the BAG family of co-chaperones in HEK293T.....	130
<b>Figure 5.4</b> – The small Heat Shock Protein (sHSP/HSPB) family.....	133
<b>Figure 5.5</b> - APMS identifies putative protein interaction partners for the small Heat Shock Protein family in a HEK293T overexpression system.....	140
<b>Figure 5.6</b> - APMS is able to capture differences between variants of the BAG3 protein.	145
<b>Figure 6.1</b> - Cardiomyocytes differentiated from BAG3 isogenic lines express BAG3-3xFLAG and display similar differentiation efficiencies.....	156
<b>Figure 6.2</b> – Heart disease-associated cell Bag3 gene variants display a different response to proteotoxic stress.....	158
<b>Figure 6.3.</b> – BAG3 interaction partners that bind differentially for the BAG3 <sup>E455K</sup> variant.....	159
<b>Supplementary Figure 3.1</b> - Expression of BAG3 protein in human iPSC and iPS-CMs..	66
<b>Supplementary Figure 3.2</b> - Genotyping iPSC lines with targeted mutations.....	67
<b>Supplementary Figure 3.3</b> - Karyotype of mutant iPSC lines.....	68

<b>Supplementary Figure 3.4</b> - Mutant iPSC lines express pluripotency factors and efficiently differentiate into cardiomyocytes. ....	69
<b>Supplementary Figure 3.5</b> - Quantification of myofibrillar disarray in BAG3-mutant cardiomyocytes. ....	70
<b>Supplementary Figure 3.6</b> - BAG3 mutations produce contractile deficits in iPSC-CMs cultured on micro-patterned substrates with increased stiffness. ....	71
<b>Supplementary Figure 3.7</b> - Effect of proteasome inhibitors on WT iPSC-CM contractility. ....	72
<b>Supplementary Figure 3.8</b> - BAG3 mutant iPSC-CMs have a positive chronotropic response to adrenergic stimulation without loss of contractility. ....	73
<b>Supplementary Figure 3.9</b> - Dose-response assay for cardiomyocyte viability after chemotherapy drug exposure. ....	74
<b>Supplementary Figure 3.10</b> - Genome engineering a 3X-FLAG affinity tag in the endogenous BAG3 locus of WTc human iPSC. ....	75

# Chapter 1 : Introduction

The cell is a fascinatingly complex system. After nearly four centuries since its discovery, and several decades of modern cellular and cellular biology, we still only understand a fraction of how the cell organizes and coordinates its thousands of components to function. Beneath this complexity also lies the root of human disease - a state where the delicate balance inside the cell is perturbed with more or less dramatic consequences. Understanding disease is one of the biggest challenges of humankind, and unraveling the mechanisms by which the cell operates would allow us to tackle it much more effectively.

During my PhD work, I strove to foster our knowledge of human disease by tackling some of the biggest problems of studying disease using molecular biology. First, the cell is not just a concoction of loose molecules, but an exquisitely organized system of interacting molecules – suggesting that mere cataloging of components is not enough. Second, not every cell in the human body is made equal - they all participate in different, specialized functions, suggesting a different molecular organization and composition inside. Third, the intrinsic variability in between humans imposes confounding factors that make it very difficult to draw a direct cause-effect of how genetic variants lead to disease.

The work described here is a journey on using the study of protein-protein interactions and stem cell-derived cardiac muscle cells to unravel the consequences of genetic variants in a specific gene, BAG3. In Chapter 2, I perform a critical analysis of the literature and justify the

importance of the study of protein interactions in a disease-relevant cell type and working with proteins expressed at endogenous levels. Chapter 3 contains already published work describing the elucidation of a disease relevant phenotype in stem cell-derived cardiomyocytes after knocking out the BAG3 gene using genome engineering. This work validated the importance of the stem cell plus genome engineering combination system for the modeling of heart disease, and set the ground for functional analyses that could be used to validate hits from protein interaction analyses. In Chapter 4 I describe the use of genome engineering on stem cells to develop multiple cell lines that enable the study of BAG3 variants in heart cells. In Chapter 5, I apply a technique to characterize protein interactions to study the function of the members of two different protein families associated with BAG3, and also to analyze the interactors of several protein variants for BAG3. This work is done in an immortalized cancer cell line using protein overexpression, and served as a proof-of-concept for the usefulness of comparing protein-protein interactions across proteins variants. Finally, in Chapter 6 I describe the results obtained from using the cell lines from Chapter 4 to study of heart-specific interactions of BAG3 variants and how cells expressing them respond to proteotoxic stress.

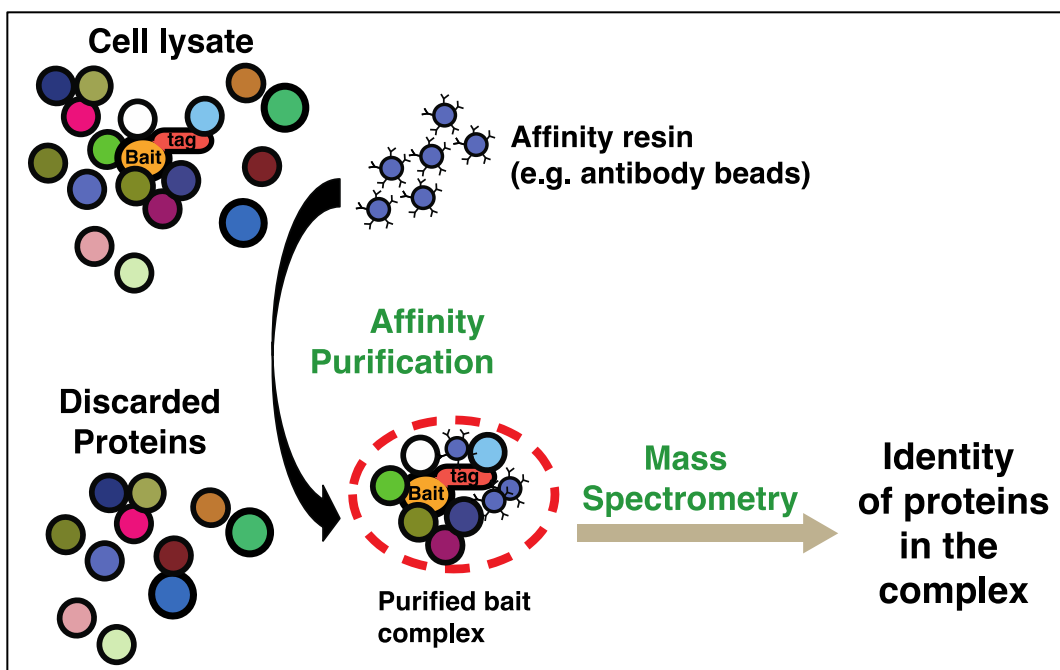
During this journey I learned a lot about protein interactions and how they are affected in disease, particularly in the protein quality control network of the cell. In the process I got to master stem cell work and analysis of protein interaction datasets, and pushed the limits of genome engineering to make their genome pliable to our will. Among many other personal and scientific lessons, I also witnessed first-hand the explosion of the genome engineering field, the maturation of the stem cell technology into a lab standard and translational applications, and the award of a Nobel prize to a neighboring lab.

## 1.1 Protein interactions allow us to unravel the functioning of the cell

Most processes and functions inside the cell are carried out by interactions between the biological molecules<sup>1</sup>. To reach an understanding of how the cell works, it is not enough to simply catalog and measure the different components. We also need to describe (“map”) how these interact with each other to orchestrate the cell functioning, and how these interactions fulfill the different functions of the cell in response to internal and external clues. This concept has given rise to the concept of ‘network biology’, which aims to describe how molecules interact with each other and use this knowledge to describe their function<sup>2</sup>. The whole set of molecular interactions in a cell is often referred to as the “interactome”.

Since proteins are commonly accepted to be the functional workhorse of the cell<sup>3</sup>, the study of protein-protein interactions (PPI) is of special relevance to our study of the system. Although methods for *in silico* PPI prediction exist<sup>4</sup>, network biology still needs of the generation of empirical, high-quality, information-rich data on interactions. This data should ideally be collected using unbiased methods. Multiple experimental approaches exist for the unbiased analysis of PPIs, but arguably the most popular one is the use of affinity purification followed by tandem mass spectrometry (AP-MS) (Figure 1.1). In AP-MS, cells are lysed in relatively mild conditions so all its components can be retrieved. Then, a protein of interest (“bait”) is purified by using a specific affinity matrix (usually antibodies that bind to the bait per se or an affinity tag the protein has been fused to). The bait protein will be captured along with the other proteins (“prey”) that interact with it in a stable manner. After a few washes to

remove nonspecific interactions, the protein solution is eluted from the affinity matrix and processed to be loaded in a mass spectrometer. The mass spectrometer identifies the proteins present in the mix, thus allowing us to get a list of the proteins that co-precipitate (interact) with our bait protein of interest. Unlike other methods for PPI analysis, AP-MS has the advantages of studying interactions in the cellular environment where they happen, being high-throughput (multiple proteins identified per run without prior knowledge), and being relatively easy to perform. Comprehensive discussion on these can be found elsewhere<sup>4-6</sup>.

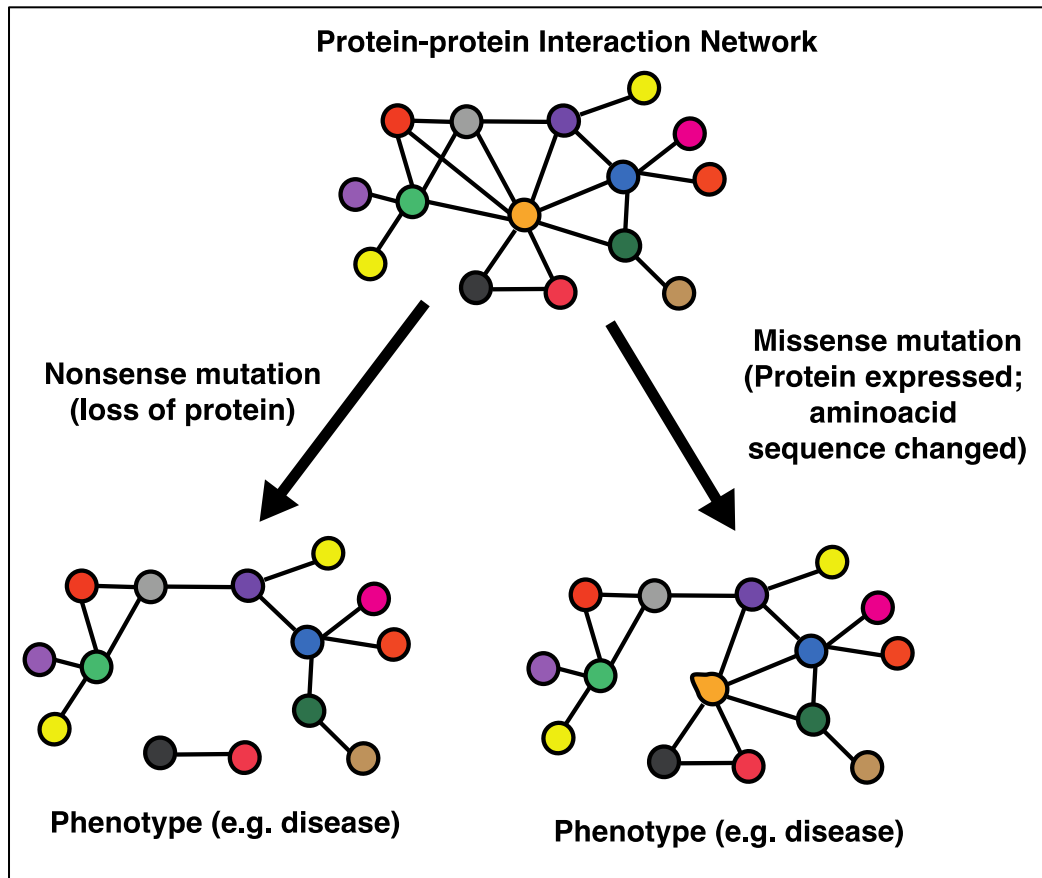


**Figure 1.1** - Affinity purification – mass spectrometry (APMS). To elucidate protein interactions, APMS uses an affinity purification resin (usually antibody-based) to isolate a specific target protein (“bait”) and all the other proteins stably interacting with it (“preys”). The bait protein is often fused with an epitope tag that allows for more robust isolation and better control (this is discussed in more detail in Chapter 2). Once the protein complex is isolated, the solution is submitted to mass spectrometry to identify the components. Note that some proteins bind to the epitope tag or to the affinity resin. These are considered false positives and need to be teased out later.

An important feature of AP-MS that needs to be taken into account is that it will not provide information on whether an interaction is direct (by physical protein-protein contact) or indirect (a whole complex being co-precipitated or another protein interacting strongly with one of the direct partners). This is a limitation if we are to study the topology of the PPI interactome, but can also be taken as an advantage as it will yield more information on the functions the bait protein is involved in. This property of AP-MS, in addition to the moderate rate of false positives and false negatives, means that often follow-up experiments for validation are required for the most relevant hits.<sup>7</sup>

One of the biggest goals of biology is to help us understand and treat disease. The study of PPIs has the potential to help in this mission. If we understand the functioning of the cell as being driven by carefully orchestrated protein interactions, then disease represents a state where this homeostasis is perturbed or lost, resulting in a pathological phenotype in the organism<sup>8,9</sup>. This is particularly useful in the case of genetic disease (inherited or *de novo*), where a specific change in the genome sequence results in a modification of the interactome of the cell by either removing a protein from the network or by modifying its sequence (and binding partners). The rapid advances in genomic sciences result in constantly increasing numbers of variants that are associated to disease. Unfortunately, as the complexity of biology requires much more time to find mechanistic explanation for disease variants, the list of unsolved disease variants continues to grow. The study of how PPIs are affected during disease holds the potential to accelerate our ability to unravel disease mechanisms and guide the design of therapeutic strategies<sup>8,9</sup>. Protein-coding genetic variants linked to disease hold particularly strong potential for this approach: since both ‘healthy’ and ‘pathological’ protein variants code

for protein, we can compare their interaction partners them to tease out which ones are specifically involved in pathogenesis. This is one of the most promising applications of the so-called ‘differential interactomics’<sup>9-11</sup>(Figure 1.2).

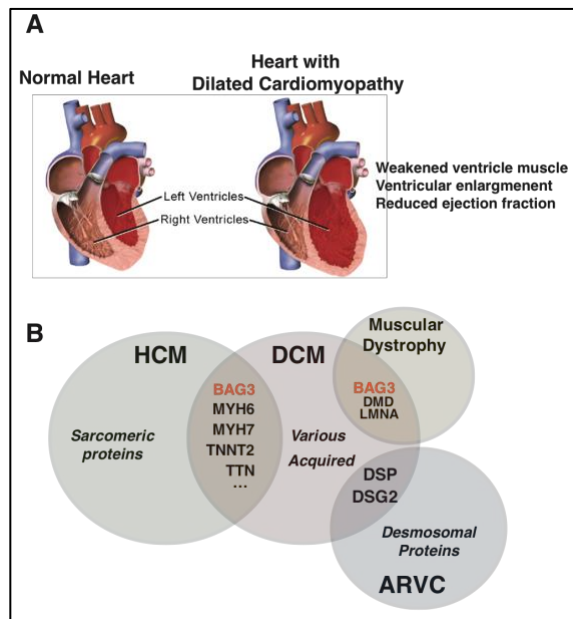


**Figure 1.2** - Protein-protein interactions bridge the gap between genotype and phenotype. Proteins (circles) perform their function by interacting with other proteins in complex networks. Sometimes, genetic variation can cause the loss of the expression of a protein (bottom, left) or the modification of the sequence of a protein, both with a consequent loss (or gain) of interactions. Since proteins are the functional effectors of genes, the phenotype observed when specific variants are present (e.g. variants associated to disease) is determined by how it affects the protein-protein interaction network. A comparison of the interactions of an unmodified (‘healthy’) protein interactome (top) versus those in a modified/‘disease’ state can help us understand the basis of genetic control of phenotype and disease.



## 1.2 BAG3 variants cause heart disease: protein quality control in the heart

Heart disease is one of the main causes of death in the world<sup>12</sup>. Dilated cardiomyopathy (DCM) is a major cause of morbidity and mortality worldwide, and the single most common cause of heart transplant in the United States<sup>12</sup>. In DCM, the heart muscle (mostly left ventricle) weakens and expands, resulting in inefficient blood pump and eventually heart failure<sup>13</sup> (Figure 1.3a). Multiple studies have pointed out to a strong genetic component to the development of DCM<sup>14</sup>. Despite an expanding list of DCM-associated genes (Figure 1.3b), therapies remain elusive. Gaining insight into how genetic variation affects the development of disease could aid in a better understanding of the pathogenesis mechanism and help in the development of therapeutic strategies.



**Figure 1.3 – Dilated cardiomyopathy** has a strong genetic component. A) In dilated cardiomyopathy, the ventricle (usually the left one) becomes enlarged and progressively deficient at pumping blood. This eventually develops into heart failure. (Image modified from Blausen Gallery<sup>15</sup>). B) A Venn diagram highlighting some genes associated with different DCM-related disorder. The BAG3 gene has been associated with DCM but also other myopathies, mostly through different variants (HCM: hypertrophic cardiomyopathy, DCM: dilated cardiomyopathy, ARVC: Arrhythmogenic right ventricular cardiomyopathy).

One of these DCM-associated genes is the BAG3 (Bcl-2 associated athanogene 3). BAG3 is a particularly interesting case for a number of reasons. First, multiple reports, both from genome wide association studies and family analyses, have linked both rare and common variants in the BAG3 gene to the onset of dilated cardiomyopathy<sup>16-22</sup>.

Second, Different variants of the BAG3 gene are associated with distinct clinical presentations:

- Nonsense mutations and missense variants in the C-terminal region of the protein have been associated with DCM<sup>16,17,19,20</sup>. Heterozygous patients develop the disease, suggesting a loss-of-function phenotype for the BAG3 gene. BAG3 function has also been associated with acquired (“takotsubo”) cardiomyopathy<sup>21,22</sup>.
- Transitions on a specific amino acid (P209L, P209Q) causes lethal severe childhood onset myofibrillar myopathy and giant axonal neuropathy<sup>23-27</sup>. The striking phenotypic difference between this variant and the adult-onset DCM associated with loss of function in BAG3 suggests that this may be a gain-of-function variant.
- A genome wide association study<sup>18</sup> identified a coding variant of BAG3 (C151R) as associated to a lower incidence of DCM. This variant was in complete disequilibrium with the major allele of another relatively common variant, P407L, but the data suggests that the C151R variant is the one directly implicated in this protective phenotype. C151R is a

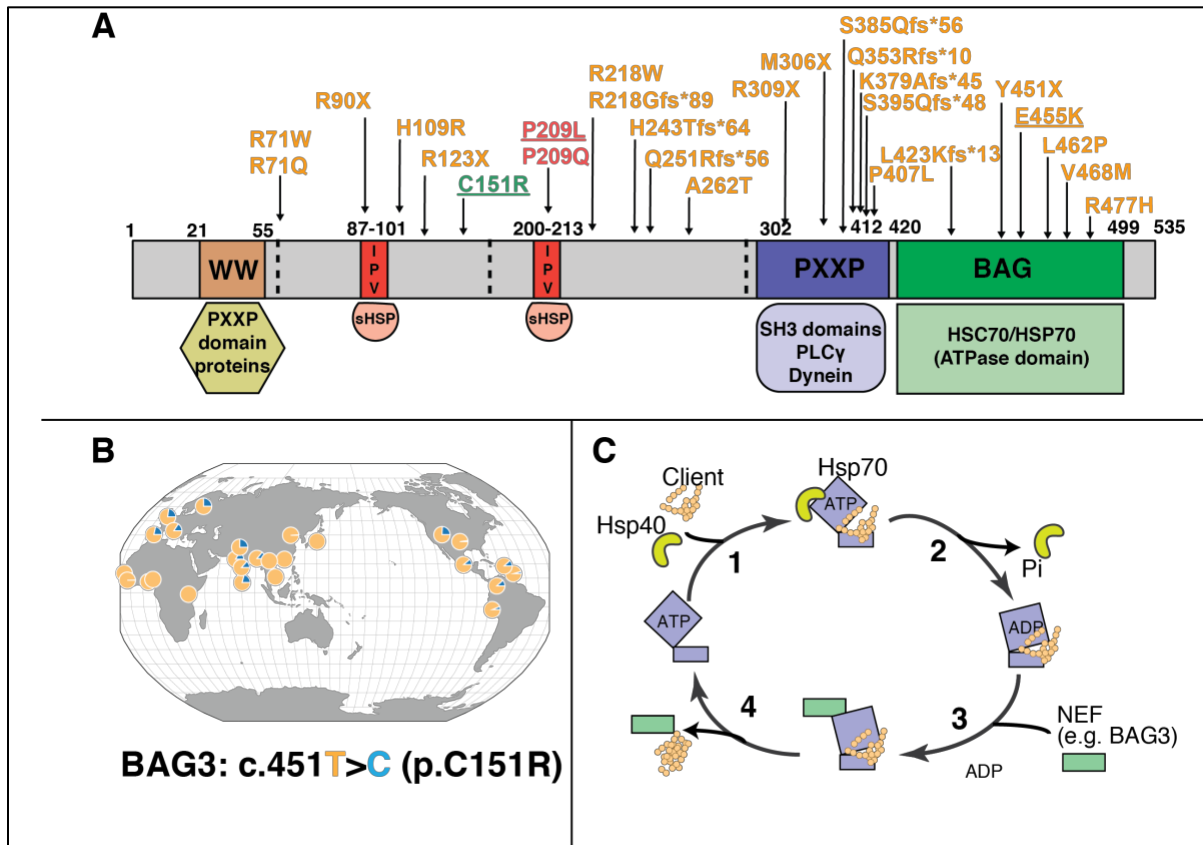
common variant, with an allelic frequency around 25% in some European and Asian populations<sup>28</sup> (Figure 1.4).

This makes BAG3 variants an interesting target for the differential interactomics approach outlined earlier. By comparing the interactions of the protein isoforms, we should be able to narrow down the list of BAG3 interactions to those specifically associated with the development of the disease phenotype.

Finally, BAG3 is ubiquitously expressed, although its expression is significantly higher in heart and skeletal muscle<sup>29</sup>. In muscle, BAG3 localizes in the Z-disk, a protein-dense structure that locates in between adjacent sarcomeres<sup>19,30</sup>. Since loss-of-function is specifically associated to DCM, this points towards BAG3 performing a unique function in the heart muscle. Comparing the interactions of BAG3 in cardiomyocytes with those in other tissues would allow us to understand what this heart-specific role is.

BAG3 is a member of the BAG family of co-chaperones. BAG proteins are known to bind to the HSP70 family of chaperones through their BAG domain (Figure 1.4). The HSP70 family of chaperones is involved in assisted folding of a high portion of the proteome, ensuring their proper function and quality control<sup>31</sup>. Several genes that code for HSP70 family members have been described, of which the most representative are HSPA1A (stress inducible) and HSC70 (constitutively expressed)<sup>31</sup>. HSP70 chaperones bind exposed hydrophobic peptide segments in proteins. Due to the enormous diversity of potential “client” proteins, it has been hypothesized that other co-factors narrow and direct the activity of HSP70, recruiting them to specific proteins to fold<sup>31,32</sup>. The members of the BAG family of proteins act as nucleotide

exchange factor (NEF) that catalyze client protein release from HSP70 and accelerate the chaperone binding cycle, and also work bridging the chaperone with specific substrates<sup>32–34</sup>.



**Figure 1.4** - Variants in the BAG3 co-chaperone are associated with DCM. A) BAG3 protein domain structure, showing most well-known binding partners and location of disease-modifying variants. Green: reduced disease risk; Red: myofibrillar myopathy and giant axon neuropathy; Orange: early-onset dilated cardiomyopathy. Underlined variants were chosen for genome engineering and AP-MS in Chapters 5 and 6. Dashed lines represent exon boundaries. Note that most variants associated with DCM affect the C-terminal region. References for variant-disease associations: <sup>16–20,22,23,25–27,35–39</sup>. B) Geographical distribution of the common variant C151R, associated with a decreased DCM incidence. Map generated using the Geography of Genetic Variants tool<sup>40</sup>. C) The HSP70 chaperone cycle. When HSP70 is bound to ATP, it adopts an open conformation, allowing substrate binding. Then, hydrolysis of ATP (catalyzed by co-chaperones such as the HSP40/DNAJ family) creates a conformational change that closes the lid around the substrate. Nucleotide Exchange Factors (NEF) then catalyze the release of the substrate and the replacement of ADP for ATP, resetting HSP70 for another cycle of binding. It is hypothesized that co-chaperones can present HSP70 with specific client proteins.

In addition to its canonical function as HSP70 co-chaperone, BAG3 is the only member of the BAG family to contain IPV (isoleucine-proline-valine) short motifs that are involved in the interaction with the members of another family of proteins, the small heat shock protein (sHSP) family. This is a family of ATP-independent chaperones (“holdases”) that bind to aggregating substrates preventing their aggregation and potentially aiding in their refolding or clearing. BAG3 is the only known link between sHSPs and HSP70s, suggesting it may coordinate the work of these two families of chaperones, representing an important connector in the chaperone network. In human cells, the BAG3-HSP70-HSPB ternary complex regulates degradation of ubiquitinated proteins via the proteasome and autophagy pathways<sup>32,41,42</sup>. Some studies have described this complex, along with CHIP ubiquitin ligase and p62 ubiquitin binding protein, as a key piece of the “chaperone-assisted selective autophagy” (CASA) that is involved in the quality control of a specific set of muscle proteins<sup>43,44</sup>.

The role of BAG3 in the protein quality control network BAG3 raises the hypothesis that this protein is playing a specifically important role in the proteostasis of the heart muscle. Indeed, evidence is accumulating that BAG3 has a central role in coordinating the cardiac protein quality control apparatus<sup>43–47</sup>. In muscle, strict control of protein folding by molecular chaperones is thought to maintain the proper function of the myocyte<sup>48–50</sup>. Cardiomyocytes maintain constant contractile function throughout a human lifetime, with consequent continuous mechanical and oxidative stress. This can lead to protein damage and misfolding that can impair contractile function and lead to formation of toxic aggregates<sup>49</sup>. Also, the heart has poor regenerative capacity, so cardiomyocytes must compensate for this proteotoxic stress

by strictly regulating protein synthesis and degradation<sup>51</sup>. To this end, heart cells maintain high-levels of constitutive and tissue-specific chaperones<sup>52</sup>. The BAG3 chaperone is located in the Z-disc of cardiomyocytes and is well positioned to bind/fold hundreds of potential “client proteins”<sup>30,45</sup>. This idea is fostered by the observation that many Z-disk-associated genes are mutated in DCM<sup>53–55</sup>.

Besides the BAG domain and IPV motifs that mediate interactions with the chaperone machinery, BAG3 also contains other domains that are involved in other processes in the cell (Figure 1.4). The WW motif near the N-termini allow it to bind proline-rich regions (such as PXXP domains), and mediates an interaction with synaptopodin-2 that has been described as necessary for autophagosome formation in CASA<sup>43,44</sup> and for autophagy induction in a cancer cell line<sup>56</sup>. BAG3 also contains a proline-rich (PXXP) region, a signature binding site for SH3 domain-containing proteins such as phospholipase C $\gamma$  and that is necessary for the binding of dynein and aggresome targeting of substrates by BAG3<sup>42</sup>.

In addition to all these domain-defined interactions, a high-throughput analysis of interactors in the HeLa cell line revealed over 300 putative interactions, including signaling molecules, cytoskeletal components and transcription factors, among others<sup>57</sup>. This modular domain structure of BAG3 to be a ‘scaffold’ protein, integrating functions from different cellular processes (a good review of these can be found elsewhere<sup>58</sup>), presumably connecting them with the protein quality control machinery.

Due to dosage sensitivity of the chaperone network, we predicted that BAG3 interactors need to be studied using native levels of expression of the bait. This means avoiding

overexpression, ideally using endogenously expressed protein. If I wanted to perform differential AP-MS comparing protein isoforms, I would also need to find a way to enable the study of the different protein variants. Two technologies allow us to overcome these issues: stem cell technology and genome engineering.

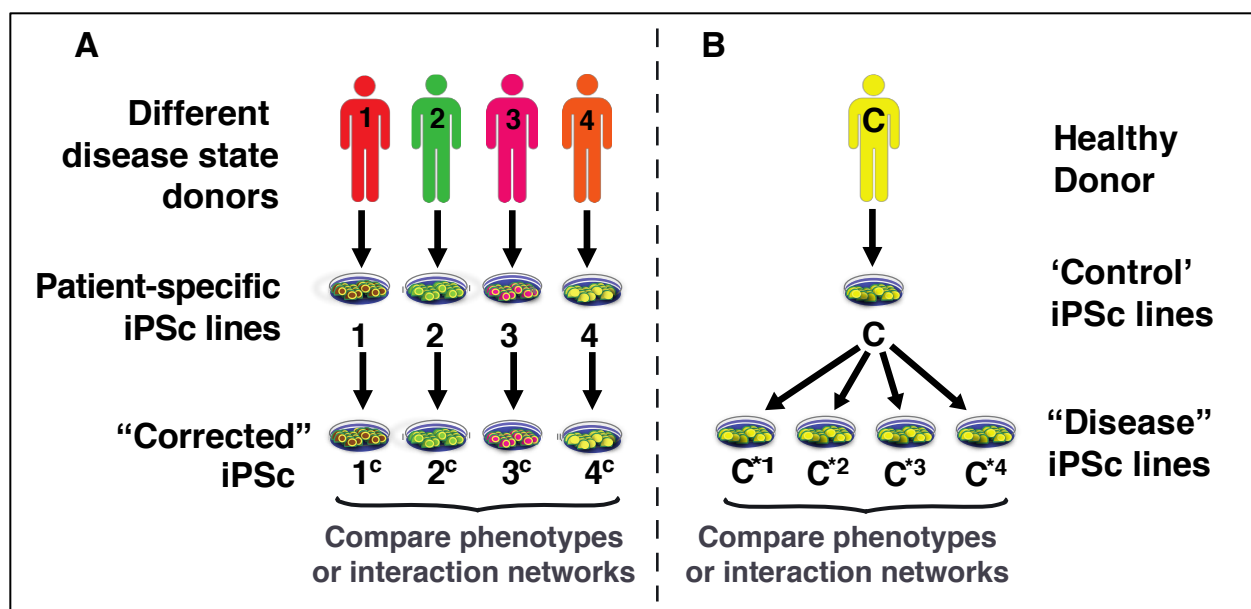
### 1.3 Stem cell technology and genome engineering enable the study of heart disease genetics

The combination of interesting disease genetics and scaffolding character of the BAG3 protein made it an ideal target for the use of the differential AP-MS approach described earlier. Understanding how specific genes cause cardiomyopathy requires the study of these variants in a relevant cell type, in this case heart muscle. Not too long ago, this was seemingly unapproachable: human heart tissue was only obtainable post-mortem, and at very limited amounts. Animal models such as mice allow you to obtain heart tissue, and they also have the advantage that you can model disease in an organismal way. Rat primary cardiomyocytes are also a very common source for the study of heart cells. However, obtaining enough material to perform some experiments (such as AP-MS) is challenging, usually requiring pooling hearts from multiple animals. More importantly, rodent coding gene sequences are not identical to that of human, and some specific variants cannot be modeled accordingly. Additionally, the animal models may not recapitulate the human pathogenesis. In the case of BAG3, mutation or knockout animal models display myopathy phenotypes<sup>45,59-61</sup>. However, BAG3<sup>+/-</sup> mice are reported to be normal<sup>61</sup>, failing to reproduce the heterozygous loss-of-function phenotype

that is seen in patients. This highlights that rodent cells might not be a good model for the study of human genetic variation in general and that of BAG3 cardiomyopathy in particular.

The development of stem cell culture and differentiation has provided scientists with a very important tool for studying heart cells. Stem cells can proliferate almost indefinitely in culture and be differentiated into cardiac muscle cells at any time<sup>62-64</sup>. This represents an excellent source of genetically homogeneous human heart cells, that allow us to perform many experiments that would have not been possible before. The development of induced Pluripotent Stem (iPS) cells<sup>63,65</sup> represented a very important leap forward for the stem cell field (Figure 1.5). These stem cells are obtained from reprogramming adult somatic cells (normally skin fibroblasts) by the expression of a combination of transcription factors. These cells have the advantage of being relatively easy to obtain, they avoid the ethical burdens associated with embryonic stem cell, and also that allow us to work with cells derived from an adult individual, so we can try to correlate the features of the cell in culture with the heart condition observed in the donor<sup>62,63,66,67</sup>.



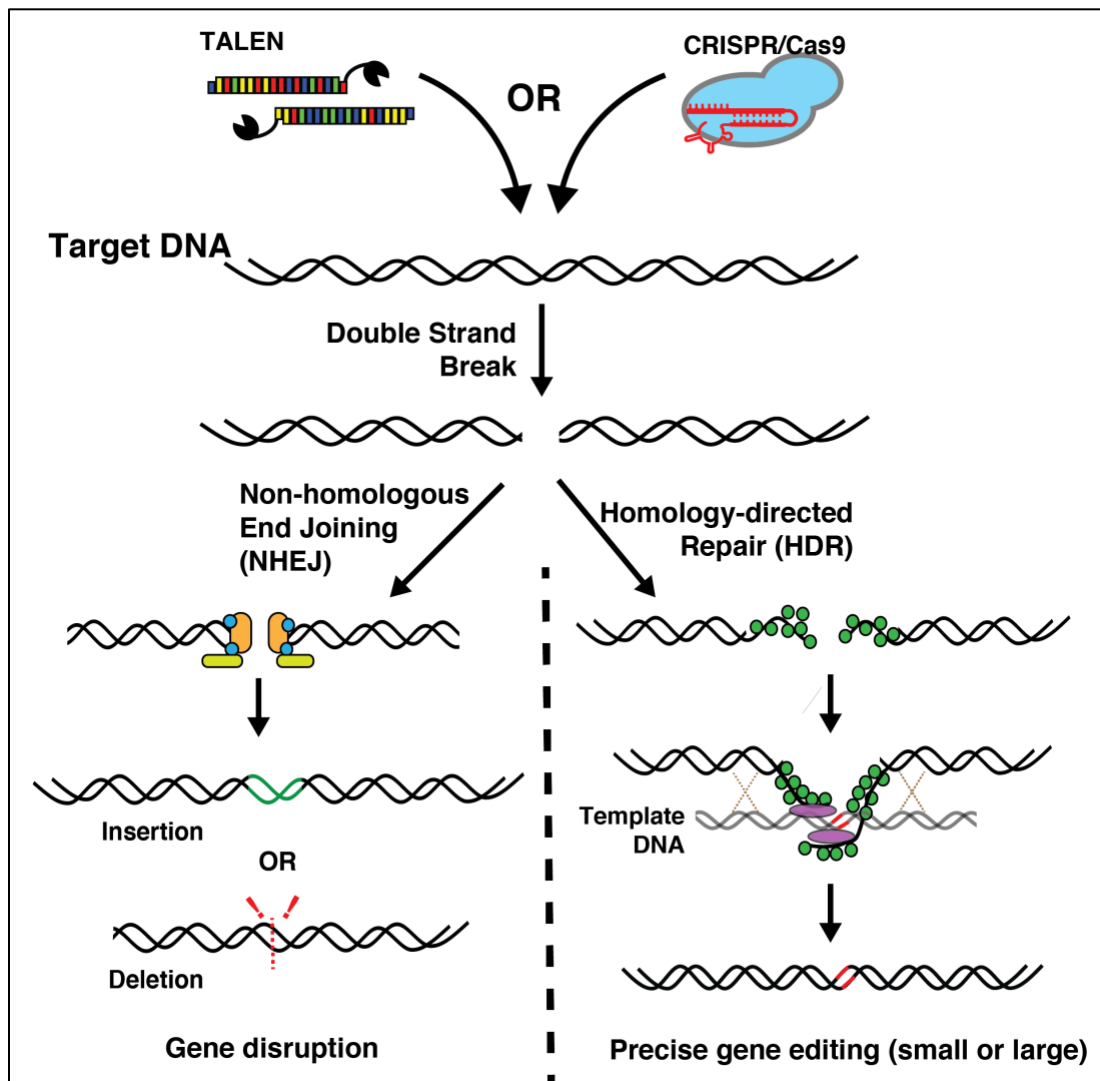


**Figure 1.5** - Two different approaches to disease modeling using iPSCs. A) Generating different iPSC cell lines from multiple patients enables identification of phenotype on-a-plate for each one of the donors. However, the genetic variation between individuals means that it would be difficult to attribute the observed phenotype or change in interactions to a specific genetic variant. Genome engineering can be used to correct specific variants and see if that affects the phenotype. B) Starting with a single donor cell line, genome engineering can be used to introduce specific genetic variants in the cells. This allows us to assign the observed phenotype or interaction changes to the specific variant introduced. This second approach allows us to dissect variability associated to specific variants in a more straightforward way, and it is easier to apply to the study of multiple variants.

Although stem cells allow us to perform cell-demanding experiments such as AP-MS, the comparative study of genetic variants for a protein presents the additional challenge of obtaining cells containing the different gene mutations (for an extended discussion of why it is important to work with endogenously expressed genes and avoid ectopic overexpression, see Chapter 2). With iPSC cells this could be done by reprogramming cell lines from different patients bearing the variants of interest<sup>62,66</sup>. However, individual variability between humans is relatively large (approximately 0.1% of the genome or 3 million basepairs), so if we observed a specific difference in the phenotype or interactome of the cells we would not know if this is

due to the variant of interest or another undesired variant. In addition, getting samples that harbor some rare variants can be really challenging as we would need to find a suitable patient and then obtain consent.

An alternative to using cell lines from different patients is provided by genome engineering technologies. In genome engineering, nucleases are used to create DNA breaks on specific sites of the genome<sup>68</sup>. Then the repair mechanism of the cell can either correct the break by religating the ends (nonhomologous end joining, NHEJ) or by using another strand of DNA as a template (homology-directed repair, or HDR)<sup>68</sup>(Figure 1.6). In the NHEJ case, sometimes the repair will create insertions or deletions that would often result in the disruption of the gene expression. More interesting for us is the case of HDR, where a ‘donor’ DNA template can be provided carrying a specific modification and that will be subsequently copied in the genome of the cell. In the last decade different methods for genome engineering have been described, of which the most representative are zinc-finger endonucleases<sup>69</sup>, TAL effector endonucleases (TALENs)<sup>70</sup> and the clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) system<sup>71,72</sup>. Due to its ease of use and high efficiency, CRISPR/Cas9 has represented the explosion of genome engineering, as it has made it possible for virtually every lab to perform genome engineering<sup>73</sup>.



**Figure 1.6.** – Genome engineering using targeted endonucleases. First, a targeted endonuclease that binds specifically to the target sequence is used to insert a double strand break (showing here TALENs and the CRISPR/Cas9 systems, used in this study). Then, two different pathways can be used by the cell to repair the lesion. (left) In non-homologous end joining, the result will be an insertion or deletion of a nucleotide sequence of variable length (mostly unpredictable). NEHJ is used mostly for gene disruption. (right) In homology-directed repair, a template DNA sequence with homology at both sides of the cut will be used as a reference for the repair. This template DNA can be from a homologous chromosome or from a provided exogenous sequence. In this last case, any modifications introduced in the template will be copied to the endogenous DNA. This can be used for the induction of specific short modifications (e.g. basepair changes) or larger insertions (e.g. gene trapping or protein fusions).

The combination of genome engineering and stem cell technology allows for the generation of *in vitro* models for genetic disease where the genetic variability is controlled for (e.g. <sup>74</sup>). But it is also a very powerful addition to AP-MS technology for the differential interactomics approach, as it allows us to introduce specific variants on the endogenous copy of the gene of interest in iPS cells. Then these cells can be differentiated into the cell type in which we want to study the interactome (e.g. cardiomyocytes). In addition, genome engineering also allows for the knock-in of a sequence coding for an epitope tag fused to the gene, which we would then use in the affinity purification step. This will allow for a better control for nonspecific hits and for the correction for any influence that the modified amino acid variants could have in the binding of antibodies against the bait (this concept is further developed in Chapter 2).

# Chapter 2 - Genome engineering and protein interaction mapping: new opportunities and remaining challenges

## Introduction

Characterizing protein-protein interactions (PPIs) is a crucial step to understanding the cell function and an essential tool to foster our understanding of disease<sup>6,75,76</sup>. Our ability to identify and characterize protein interactions has improved dramatically in the last few decades, going from studies describing a single interaction to high-throughput studies yielding thousands of interactions<sup>6</sup>. Affinity purification combined with mass spectrometry (AP-MS) has rapidly become the preeminent method for the high-throughput characterization of protein-protein interactions<sup>77</sup>. This is mostly because of the ability to study interactions *in vivo* inside the cell, unlike other methods that rely on non-native environments and are prone to false-positives<sup>78,79</sup>. The popularity of AP-MS has been fostered by improvements in mass spectrometry instrumentation, sample preparation automation and data analysis methods. However, most of the studies mapping PPIs rely on the use of exogenous protein overexpression and/or using immortalized cell lines, yielding interactions that can differ significantly from those present in the living organism. The recent progress in genome engineering<sup>73</sup> is now providing us with a set of tools that allow us to overcome these issues and generate PPI interactomes that are relevant to human tissues and human disease. In this

review, we describe how genome engineering can overcome the limitations of many current studies, while critically analyzing these new approaches. We also try to provide a rationale for using genome engineering as an essential tool for deriving higher quality datasets for the study of cell function and human disease.

## 2.1 Limitations of current protein-protein interaction mapping datasets

At the time of writing this manuscript, a number of studies have provided drafts for the human ‘interactome’ using AP-MS<sup>76,80,81</sup>. These represent invaluable resources to understanding some general principles of the functional organization of the cell, providing insights that could not be extracted from individual interactions alone<sup>6</sup>. However, the information obtained from these high throughput studies is limited by two main factors, non-physiologic protein expression levels and cell type used. Here we proceed to describe how these interaction datasets are likely to have false positive interactions that limit their application to human physiology or disease mechanisms where proteins are tightly regulated.

### 2.1.1 1.-Bait protein overexpression

Most AP-MS studies rely on the overexpression of the ‘bait’ protein to levels substantially higher than they are present when expressed from the endogenous copy of the gene. This is usually achieved by transient transfection of a plasmid overexpressing the protein of interest or by the generation of a cell line that stably expresses the protein of interest under a generic strong promoter. The consequences of protein overexpression on its PPIs impose a

limitation to the interpretation of the data. Some reports have estimated ~20% of yeast proteins mislocalize when overexpressed<sup>82,83</sup> or produce some form of growth defect<sup>84</sup>. This mislocalization has also been studied, though at a smaller scale, in human cell lines<sup>85,86</sup>. Currently there are no studies comprehensively comparing protein-protein interactions of overexpressed proteins versus endogenously expressed proteins. However, it is widely assumed that overexpression will have a dramatic impact on the PPIs, even for those proteins that don't change their location dramatically. One reason is that protein overexpression can alter the entire protein network, affecting protein complex assembly and stoichiometry, especially on regulatory and signaling proteins and complexes<sup>86,87</sup>. In addition, there is evidence that protein overexpression can increase protein promiscuity through mass action, promoting interactions that would not happen at normal levels of expression<sup>88,89</sup>. In fact, there is a correlation between dose sensitivity and degree of interaction of a protein, as well as with its intrinsic protein disorder and linear motif content<sup>88</sup>. An additional and especially dramatic consequence of gene dosage imbalance is the misfolding and aggregation of the protein of interest, which can also influence other proteins being aggregated with it<sup>90,91</sup>.

The issue of protein expression is especially important when we think of applying PPI information to study disease. Many diseases are associated with overexpression or misregulation of gene products<sup>92-94</sup>, which suggests that protein network homeostasis is dosage sensitive. Although using protein overexpression can help mimicking the disease state<sup>94</sup>, the lack of information on the native function of the proteins involved makes it very difficult to infer the effect of such change in expression in the system.

A noteworthy effort to map a PPI network of proteins expressed at near endogenous levels was performed recently<sup>80</sup> using cell lines that stably express the gene of interest in a large cassette comprising the surrounding genomic region of the gene. Although this method allows for the capturing of the endogenous promoter and some regulatory regions, most of the genomic context of the gene is not captured which may result in incomplete mimicking of the transcriptional regulation of the gene. In addition, this study used HEK293 cells, a cancer immortalized cell line (see below).

The most straightforward way to study the PPIs for genes expressed and regulated at endogenous levels and without increasing the gene dose is to work with the endogenous copy of the gene. This can be achieved by using antibodies against the protein to be studied, or using homologous recombination to insert an epitope tag at the gene loci. The advantages and drawbacks of these approaches are discussed in more detail later in the text and in Figure 2.1.

### **2.1.2 2.-Use of immortalized cell lines**

A more prevalent feature in the high throughput AP-MS interaction datasets is the use of immortalized cell lines (mostly HEK293 and HeLa). These cell lines have the advantage of being easy to grow in big batches and relatively easy to transfect. They are also more amenable to gene targeting for the insertion of expression cassettes in safe-harbor locus or epitope tags in endogenous genes. For these reasons they spearheaded the PPI mapping in human cells. However, these advantages come at the cost of the cells being substantially different from what would be expected of a cell in a live organism. Their gene expression profile is optimized for unlimited growth capability<sup>95,96</sup>. In addition, they contain a substantial amount of gross



chromosomal abnormalities (amplifications, rearrangements), which results in a modified gene dosage for thousands of proteins. Finally, these cell lines have been described to have lost many of the specific functions of the tissue of origin *in vivo*<sup>97</sup>.

The function of a protein on a specific cell type can only be fully understood if its protein interactions are studied in the cell of interest. Different tissues contain different protein expression profiles<sup>98</sup>. Even proteins that are expressed in multiple tissues are expected to have different functions, based on whether their interactors are expressed or not, and at which level, in the tissue of interest. Given that most human diseases are associated with dysfunction in a specific cell type or tissue, it seems clear that a full understanding of the role of specific proteins in disease will only be achieved by analyzing their interactors in the cell types that are affected in disease. Evidence of this was provided in a study using yeast two hybrid screening<sup>99</sup>, that described how protein variants associated with disease disrupted interactions with protein partners that are expressed in disease-relevant tissues.

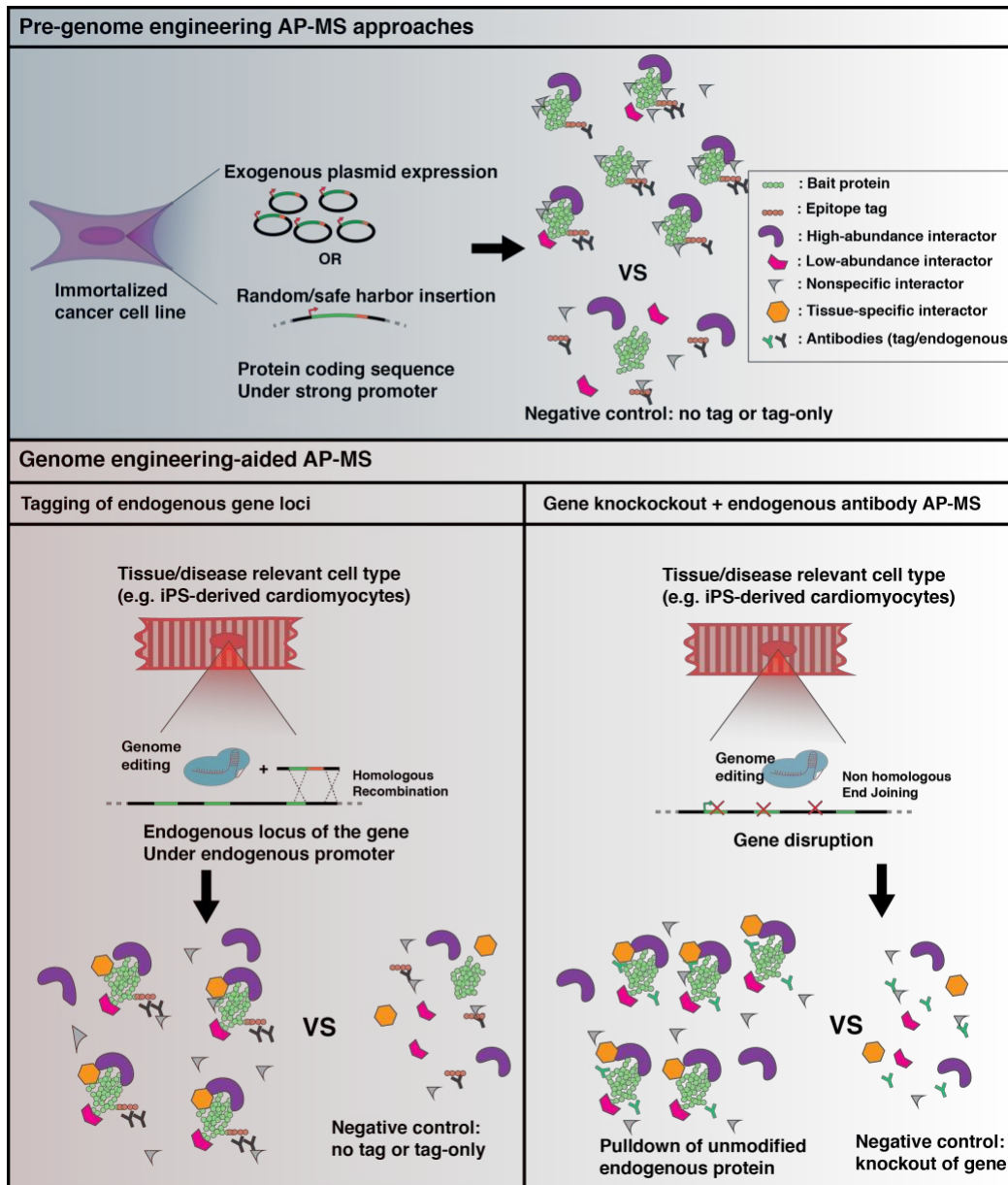
The most obvious alternative to the use of a generic immortalized cell line is the use of primary cells. However, these are generally very difficult or impossible to acquire in an amount that allows for proteomic studies. This is especially true for neurodegenerative and cardiac disease. An alternative to primary tissue is the use of stem cells, which can be expanded indefinitely and can be directed to differentiate into multiple cell types of interest, including but not limited to cardiac myocytes and neurons (multiple types)<sup>62,63</sup>. The main obstacle towards the implementation of these stem cell-derived systems for the study of protein interactions has been the cost of cell production and the inability to efficiently modify the

genome of the cells to enable AP-MS on differentiated tissue. However, stem cell culture and differentiation is getting increasingly easy, and genome engineering tools have enabled custom genomic modification of stem cells<sup>63,73,100</sup>.

## 2.2 Genome engineering allows us to overcome these issues

The factors described above seem impose important limitations of current high-throughput PPI mapping in human cells, especially when it comes to getting knowledge that's applicable to studying disease mechanisms. However, in the last five years the advances in genome editing technologies (more representatively the CRISPR/Cas9 system<sup>73,101,102</sup>) have improved dramatically our ability to introduce custom modifications in the genome of cells, representing a key tool towards overcoming these issues. The two main ways in which genome engineering can aid the development of better cellular systems for AP-MS are discussed in Fig.1. On one side, genome engineering tools have enabled increased efficiency of homologous recombination in custom genomic regions. This allows for the insertion of sequences coding for epitope tags and other useful protein fusions (fluorophores, biotinylating enzymes) in the endogenous locus of the genes of interest, enabling experiments such as AP-MS without presumably affecting the expression levels and regulation of the gene. The idea of using homologous recombination to target genes has been around for decades<sup>103-105</sup>, but the low efficiency of recombination and the length of the process made it difficult to justify the approach for routine or mid/high-throughput experiments such as AP-MS. This is especially true for human cells, where efficiencies have ranged between 0.1%-1% at best, frequently

aided by viral transduction<sup>106,106–109</sup>, and generally lower for non-immortalized cell lines. With the advent of CRISPR/Cas9 technology multiple studies have been able to knock-in DNA sequences in specific genomic regions with substantially higher efficiencies and in a variety of non-immortalized cell lines<sup>85,100,110–114</sup>, including induced pluripotent stem cells (iPSc) which can then be differentiated into a variety of cell types<sup>63,103</sup>. This has allowed for increased throughput, generating multiple cell lines in a relatively short span of time. Another way genome engineering allows for better PPI studies is by enabling effective, high efficiency knockout of genes even in generally difficult cell lines<sup>115,116</sup>. This is an invaluable tool for validation studies, but it also allows for the easy generation of cell lines that can be used as very informative controls for pulldown experiments using antibodies against the endogenous protein (a similar approach had been used in the past using siRNA technology<sup>117</sup>, and reviewed elsewhere<sup>7</sup>). Using Cas9+gRNA ribonucleoprotein complexes<sup>115,116</sup>, high-efficiency knockout can also be produced in primary cells that are often not amenable for efficient homologous recombination<sup>118,119</sup>, opening the door to an exciting new avenue of PPI studies. A critical analysis of these two approaches (endogenous tagging vs knockout for controls) can be found in Figure 2.1.



**Figure 2.1** - Comparison of the main approaches for protein interaction analysis using AP-MS. Top: most commonly used approach, where proteins are overexpressed on immortalized, cancer cell lines. The over-expression of the protein potentially causes non-native interactions by mass action kinetics, and also may involve the loss of low-abundance interactions (see main text for further discussion). Bottom: approaches enabled by the use of genome engineering. Homologous recombination allows for the study of proteins expressed at endogenous levels of expression, and the use of tissue-relevant cell types allows for the identification of cell specific interactions. Advantages and disadvantages of the two main approaches using genome engineering are also enumerated.

**Table 2.1** - Advantages and disadvantages of the main approaches for protein interaction analysis using AP-MS (described in Figure 2.1).

Epitope tag-based	Endogenous antibody-based
<ul style="list-style-type: none"> <li>+The same antibody can be used on all pull-downs, increasing control for nonspecific interactions.</li> <li>+Antibody binding will not affect interactions. +Antibody binding is the same regardless of protein variant.</li> <li>+Allows for the insertion of other helpful protein fusions (fluorophores<sup>100,110</sup>, biotinylation enzymes<sup>120</sup>...)</li> <li>-Tag can interfere with protein expression, function, stability or interactions.</li> <li>-Homologous recombination is difficult to achieve in some cell types.</li> <li>-Tag may bind proteins, increasing false positives if experimental design does not account for this.</li> </ul>	<ul style="list-style-type: none"> <li>+ Knockouts (for control) are much easier to generate than targeted homologous recombination insertions.</li> <li>+ Very good control for nonspecific binding.</li> <li>+ Potentially allows for primary tissue studies.</li> <li>- Good quality antibodies are not available for a big fraction of the proteome.</li> <li>- Antibody may affect protein-protein interactions.</li> <li>- A different antibody will need to be used for each bait</li> <li>- Different protein variants may have different binding to antibody.</li> <li>- Knockout of gene of interest may be lethal or cause another dramatic phenotype.</li> </ul>

## 2.3 Outlook / Conclusion

A lot of technical and conceptual progress has been made in the last decade towards mapping protein-protein interactions in human cells, giving rise to interactome maps containing thousands of proteins and interactions. Although these studies have been invaluable to understand some essential principles of the functional organization of the cell, the vast majority of APMS studies rely on protein overexpression and the use of immortalized cancer cell lines. These factors compromise the ability recapitulate the interactions that happen in the living organism. The recent explosion of the genome engineering field presents us with a

number of invaluable tools that can help us overcome these issues, allowing for the study of the interaction of endogenously expressed proteins in multiple cell types (virtually any cell type if we consider iPSc-derived cells). The democratizing nature of the CRISPR/Cas9 technology allows any laboratory to generate cell lines that would have been very difficult to justify making for APMS studies just five years ago. We do not claim that genome engineering will be the solution to all the current limitations of the APMS approach, and certainly genome engineering application brings a set of potential issues on its own (we try to summarize these factors in Table 1). However, we should not let these limitations prevent us from using proper expression and cell models that can will undoubtedly yield higher quality, more informative interaction datasets.

Genome engineering is not only transformative in our ability to generate cell line for better PPI mapping, but it also provides us with a plethora of tools to be used for the functional validation of the data. Another very exciting avenue that has been opened by genome engineering is the induction of discrete modifications in the genome of cells to allow for the study of PPIs for different protein variants<sup>121–123</sup>. When combined with quantitative mass spectrometry<sup>124</sup>, this can boost our understanding of how specific protein-coding gene variants cause disease by allowing us to narrow down our list of interactions to those that are relevant in this process<sup>8,9</sup>.

Protein-protein interactions hold the power to elucidate the function of virtually every gene, and by doing so help us unravel the molecular mechanisms of cell function and disease<sup>8</sup>. To be able to fulfill this promise, we need to exit the ‘cataloging’ era of PPI studies, where

getting big numbers of interactions was the priority, and transition into a search for higher quality data, more representative of the primary tissue. A lot of progress needs to be done in multiple fronts before we achieve this and the journey is bound to be hard, but genome engineering presents itself as a key travel companion.

**Table 2.2** - Limitations of genome engineering and AP-MS technologies for the mapping of Protein-protein interactions. Marked with an asterisk, those factors that the authors of this review identify as predominant issues for which no clear solutions are on sight.

<b>Genome engineering related limitations</b>	
<b>Issue</b>	<b>Prospect</b>
Modifications in unintended genomic sites (off-target) <sup>125,126</sup> . Non intended modifications can also happen in the form of undesired insertion or deletions in the unedited allele of a target gene.	Multiple methods in that reduce such effect can be used <sup>127</sup> . Most notably, ribonucleoprotein complex delivery <sup>115,116,126</sup> and alternative variants of Cas9 <sup>128,129</sup> .
(if using epitope tag insertion) Epitope tag insertion can interfere with bait protein expression, stability or interactions. It may also have an effect on the expression of other genes (*).	Currently, testing different locations for the insertion and running quality control in the form of immunoblot/immunostaining remains the only options.
Making lines still very time consuming. Mid/high throughput datasets challenging to achieve.	There is a steady increase in our ability to make lines by homologous recombination, with some studies reporting up to a few dozen <sup>100,111</sup> .
<b>AP-MS related limitations</b>	
<b>Issue</b>	<b>Prospects</b>

<p>Discrimination between <i>bona fide</i> interactors and nonspecific interactors (mostly coming from antibody/affinity matrix)<sup>77,130</sup>.</p>	<p>Careful experimental design with good controls(Figure 2.1) and rigorous statistical analysis greatly address this problem<sup>77,131,132</sup>.</p>
<p>Affinity purification captures only stable interactions and complexes. Most transient or weak interactions are expected to be undetected. (*)</p>	<p>Some approaches have been developed to improve weak PPI detection, most notably proximity-based biotinylation<sup>120</sup> and crosslinking-affinity purification<sup>133</sup>.</p>
<p>Cell lysis and disruption might promote protein interactions that wouldn't happen <i>in vivo</i>.</p>	<p>See above.</p>
<p>Mass spectrometry data is very prone to missing values, especially for less abundant proteins. It is difficult to discriminate protein.</p>	<p>Mass spectrometry instrumentation and acquisition methods are constantly improving. Targeted and data-independent acquisition proteomics provide quantitative data without missing values<sup>124</sup>. Top-down proteomics is still in its infancy but can help solve isoform identification problems<sup>134</sup>.</p>



# Chapter 3 - A BAG3 Chaperone Complex Maintains Cardiomyocyte Function During Proteotoxic Stress

*Luke M. Judge<sup>1,2</sup>, Juan A. Perez-Bermejo<sup>2,3</sup>, Annie Truong<sup>2</sup>, Alexandre J. S. Ribeiro<sup>2,4\*\*</sup>, Jennie C. Yoo<sup>2</sup>, Christina L. Jensen<sup>2</sup>, Mohammad A. Mandegar<sup>2\*</sup>, Nathaniel Huebsch<sup>2</sup>, Robyn M. Kaake<sup>2</sup>, Po-Lin So<sup>2</sup>, Deepak Srivastava<sup>1,2</sup>, Beth L. Pruitt<sup>4,5</sup>, Nevan J. Krogan<sup>2,3</sup> and Bruce R. Conklin<sup>2,6\*</sup>*

<sup>1</sup>Department of Pediatrics, University of California, San Francisco, San Francisco, California 94143, USA

<sup>2</sup>Gladstone Institutes, San Francisco, California 94158, USA

<sup>3</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California 94143, USA

<sup>4</sup>Stanford Cardiovascular Institute and Mechanical Engineering Department, Stanford University, Stanford, California 94305, USA

<sup>5</sup>Bioengineering and Molecular and Cellular Physiology Departments, Stanford University, Stanford, California 94305, USA

<sup>6</sup>Department of Medicine, University of California, San Francisco, San Francisco, California 94143, USA

Corresponding Author: Bruce R. Conklin, Gladstone Institute of Cardiovascular Disease, 1650 Owens St. San Francisco CA 94158

## 3.1 Abstract

Molecular chaperones regulate quality control in the human proteome, pathways that have been implicated in many diseases, including heart failure. Mutations in the BAG3 gene, which encodes a co-chaperone protein, have been associated with heart failure due to both inherited and sporadic dilated cardiomyopathy. Familial BAG3 mutations are autosomal dominant and frequently cause truncation of the coding sequence, suggesting a heterozygous loss-of-function mechanism. However, heterozygous knockout of the murine BAG3 gene did not cause a detectable phenotype. To model BAG3 cardiomyopathy in a human system, we generated an isogenic series of human induced pluripotent stem cells (iPSCs) with loss-of-function mutations in BAG3. Heterozygous BAG3 mutations reduced protein expression, disrupted myofibril structure, and compromised contractile function in iPSC-derived cardiomyocytes (iPS-CMs). BAG3 deficient iPS-CM were particularly sensitive to further myofibril disruption and contractile dysfunction upon exposure to proteasome inhibitors known to cause cardiotoxicity. We performed affinity tagging of the endogenous BAG3 protein and mass spectrometry proteomics to further define the cardio-protective chaperone complex that BAG3 coordinates in the human heart. Our results establish a model for evaluating protein quality control pathways in human cardiomyocytes and their potential as therapeutic targets and susceptibility factors for cardiac drug toxicity.

## 3.2 Introduction

Cardiomyocytes must maintain constant contractile function throughout a human lifetime. As a result, the cells undergo continuous mechanical and oxidative stress, which leads to protein damage and misfolding that can impair contractile function and lead to formation and aggregation of toxic peptides<sup>49</sup>. Additionally, the heart has minimal regenerative capacity,

so cardiomyocytes must compensate for this proteotoxic stress by strictly regulating new protein synthesis and degrading damaged protein components<sup>51</sup>. To this end, cardiac tissue maintains high-levels of constitutive and tissue-specific chaperones in which rare mutations can cause severe disease<sup>23,135</sup>.

Protein quality control pathways are increasingly recognized for their importance in both inherited and sporadic cardiac disease, and as potential therapeutic targets<sup>52</sup>. The co-chaperone BAG3 is particularly interesting, because variants of BAG3 can be both pathologic and protective. For example, the rare P209L missense mutation causes severe childhood onset myofibrillar myopathy, a lethal disease that affects both skeletal and cardiac muscle, possibly through a toxic gain-of-function in the BAG3 protein<sup>23,59,60</sup>. Additionally, a variety of heterozygous mutations in BAG3, many of which are nonsense or frameshift mutations consistent with loss-of-function, cause autosomal dominant familial dilated cardiomyopathy<sup>16-18,20</sup>. BAG3 has also been associated with acquired forms of cardiomyopathy, such as stress (“Takotsubo”) cardiomyopathy<sup>19,21,22</sup>. Alternatively, a common coding polymorphism in BAG3 was associated with a significantly lower odds ratio for developing idiopathic dilated cardiomyopathy, suggesting that this polymorphism could be protective<sup>18</sup>. In animal models, BAG3 mutations or deficiency generally result in myopathy phenotypes<sup>17,59,61</sup>. However, in mice, knockout of the BAG3 gene produced inconsistent findings: one group reported a primary myopathy phenotype and the other reported systemic pathology without a clear muscle phenotype<sup>61,136</sup>. Interestingly, neither group detected any phenotype in heterozygous null mice. Because BAG3-related cardiac disease in humans is most frequently associated with heterozygous loss-of-function mutations, mice may not be an ideal model for the disease.

BAG3 is a stress-response gene induced by heat shock factor 1<sup>137</sup>. Although BAG3 is widely expressed in many tissues, it is most highly expressed in skeletal and cardiac muscle, where it localizes with the sarcomeric Z-disk<sup>61</sup>. As a co-chaperone, the BAG3 protein modulates ATP turnover and the protein folding activity of HSC70/HSP70<sup>32,33</sup>. It also physically and functionally interacts with a separate family of chaperones, the small heat shock proteins (HSPB genes), and acts as a scaffold that links the functions of HSP70 and HSPB families<sup>138–140</sup>. In human cells, the BAG3-HSP70-HSPB complex regulates degradation of ubiquitinated proteins via the proteasome and autophagy pathways<sup>41,43,141</sup>. However, these experimental studies used overexpression and/or transient knockdown in immortalized cell lines with uncertain physiologic relevance. Given the cardiac-specific pathology caused by human BAG3 mutations and the unique expression pattern of the protein in the target tissue, disease modeling will be most informative if performed in cardiomyocytes. Furthermore, to avoid spurious interactions produced by less precise methods, manipulation of the endogenous gene is the preferred method. To evaluate the role of BAG3 in the physiological and pathological functions of human cardiomyocytes, we generated an isogenic series of cardiomyocytes derived from genetically engineered human induced pluripotent stem cells (iPSCs). In this way, we aimed to more accurately recapitulate aspects of human cardiomyopathy and mechanisms that protect against cardiac stress.

### 3.3 Results

We detected BAG3 protein in the cytoplasm of iPSCs, which increased ~10-fold in iPSC-derived cardiomyocytes (iPS-CMs), where it was enriched at the sarcomeric Z-disk and

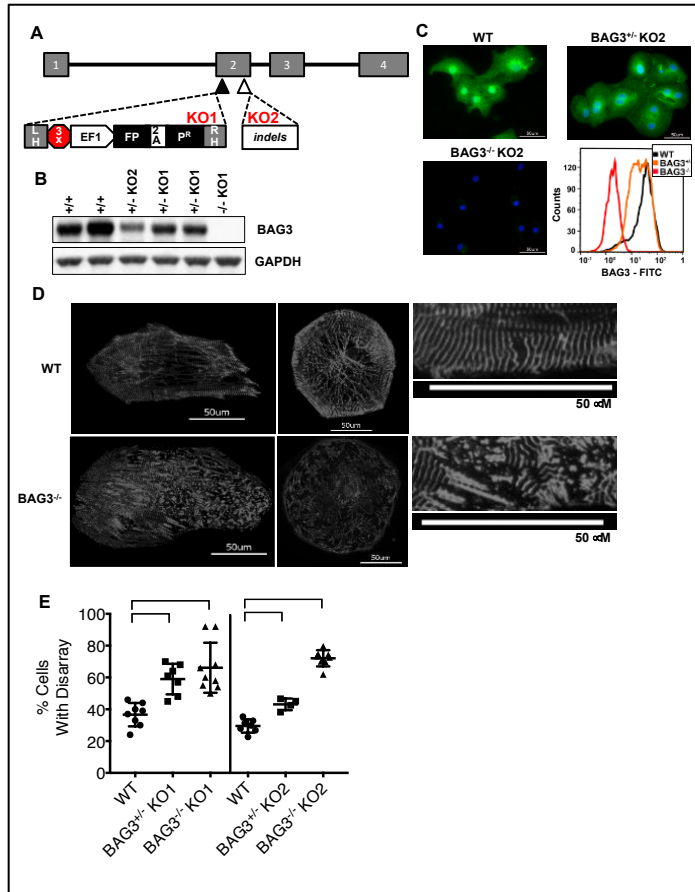
the perinuclear region (Figure 3.1 and Supplementary Figure 3.1). We first studied the effects of BAG3 deficiency in cardiomyocytes by generating loss-of-function mutations via engineering the endogenous BAG3 locus<sup>121</sup>. The strategy was designed to mimic effects of early nonsense mutations such as R90X and R123X that were reported in dilated cardiomyopathy patients<sup>17</sup>. To maximize the insights we might gain into molecular mechanisms and minimize the chances of common off-target editing, we used multiple genome-editing tools to create mutations at two different sites in the BAG3 gene (Figure 3.1). We isolated both heterozygous and homozygous clones with loss-of-function mutations using TALEN- and CRISPR-based strategies targeting two different parts of the second exon (Figure 3.1 and Supplementary Figure 3.2). We verified the genotype of each engineered mutant line by sequencing. All iPSC lines had a normal karyotype, expressed pluripotency markers, and could efficiently differentiate into cardiomyocytes (Supplementary Figs. 3 and 4).

As expected, BAG3 loss-of-function mutations decreased BAG3 protein expression (Figure 3.1 and Supplementary Figure 3.1). Mutant cardiomyocytes were maintained for >30 days of differentiation without displaying obvious abnormalities in cell viability, morphology, or beating. This result is consistent with the murine BAG3 knockout model that exhibits normal cardiac morphology at birth. Based on human clinical reports and animal studies, we hypothesized that BAG3 mutant cells would develop disrupted myofilament structure over time, particularly at the sarcomeric Z-disk<sup>52,59,61</sup>. To examine myofilament structure, we replated cardiomyocytes onto glass coverslips and cultured them for variable intervals of time before fixing and staining them for the sarcomeric Z-disk protein  $\alpha$ -actinin (ACTN2).

Cardiomyocytes continued to beat after being plated on glass coverslips, but after  $\geq 7$  days, we observed a dramatic disruption of the Z-disk structure in BAG3<sup>-/-</sup> cardiomyocytes (Figure 3.1). To quantify this phenotype, we blindly scored the extent of sarcomeric disarray in individual fixed cells on a five-point scale (Supplementary Figure 3.5)<sup>142</sup>. We observed an increase in the proportion of BAG3-mutant cells with significantly disordered myofilaments compared to wild-type cells, and BAG3<sup>-/-</sup> cells displayed a trend toward more severe phenotype than BAG3<sup>+/-</sup> cells (Figure 3.1 and Supplementary Figure 3.5). These results were consistent between lines generated by the different genome-targeting strategies, indicating that the phenotype was caused by the loss of BAG3 expression and not off-target effects.

Cardiomyocytes cultured on standard tissue culture surfaces frequently display a different morphologic shape and myofibril arrangement than that seen in normal tissue. Furthermore, culture on glass surfaces, which are extremely stiff, exposes the cells to an artificial mechanical environment. To evaluate cardiomyocyte contractile function under more physiologically relevant conditions, we cultured cardiomyocytes on micro-patterned surfaces with constrained geometry and on polyacrylamide substrates of different stiffness. We then measured contractile function under those varying mechanical conditions. Cardiomyocytes were seeded onto rectangular patterns with a 7:1 aspect ratio to promote mature sarcomeric organization and function, with substrate stiffness mimicking physiologic (10 kPa) or pathologically increased (35 kPa) myocardial stiffness. Contraction power was calculated for individual cells by measuring the displacement of fluorescent beads embedded in the substrate as previously described<sup>143</sup>. When cultured on substrates with both physiologic and increased

stiffness, BAG3<sup>+/-</sup> and BAG3<sup>-/-</sup> cardiomyocytes generated less contraction power than wild-type cardiomyocytes (Figure 3.2 and Supplementary Figure 3.6).



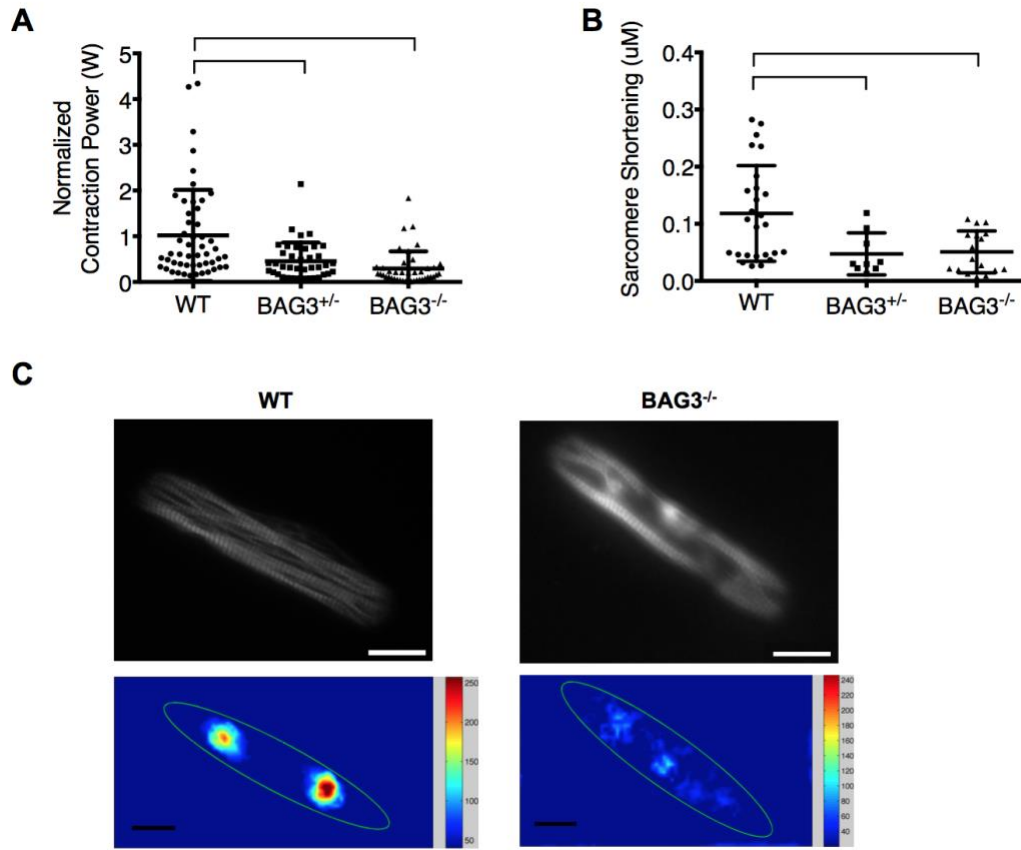
**Figure 3.1 - Genome engineering an isogenic series of BAG3 mutations leads to sarcomeric disarray in human iPS-derived cardiomyocytes.** (A) Schematic of the BAG3 gene with four exons in the predominant coding isoform. Because an alternatively spliced isoform excludes the first exon, the second exon was targeted for knockout. BAG3 knockout lines were generated using TALEN-induced (black triangle) targeted integration of the knockout vector in exon 2 (KO1). Vector included flanking left and right homology arms (LH and RH, respectively) with a terminator sequence in three reading frames followed by the mCherry fluorescent protein (FP) and a puromycin selection cassette (P<sup>r</sup>) driven by EF1 $\alpha$  promoter. Transient expression of Cas9 with a guide RNA targeted downstream in exon 2 (open triangle) induced small insertions/deletions (indels) by non-homologous end joining, resulting in frameshift and nonsense mutations (KO2). (B) Western blot for BAG3 protein in iPS-CMs. (C) Immunofluorescent staining and flow cytometry with antibody targeting BAG3 in iPS-CMs. (D) Examples of pathology seen in BAG3<sup>-/-</sup> iPS-CMs compared to wild-type (WT) controls. Cells were plated on glass coverslips, fixed, and stained with antibody to ACTN2 to label Z-disks. All scale bars are 50  $\mu$ m. (E) Quantification of sarcomeric disarray from blinded scoring on a five-point scale, with disarray defined as the percentage of cells scored as class 3–5. Individual replicates are plotted with mean and s.d. Brackets indicate significant difference by one-way ANOVA with Bonferroni's test for multiple comparisons,  $p < 0.001$ . For KO1 analysis 7-9 independent cultures per line from three separate differentiations were scored. For KO2 analysis 4-8 independent cultures per line from two separate differentiations were scored. From each culture >50 cells were scored.

(A) Schematic of the BAG3 gene with four exons in the predominant coding isoform. Because an alternatively spliced isoform excludes the first exon, the second exon was targeted for knockout. BAG3 knockout lines were generated using TALEN-induced (black triangle) targeted integration of the knockout vector in exon 2 (KO1). Vector included flanking left and right homology arms (LH and RH, respectively) with a terminator sequence in three reading frames followed by the mCherry fluorescent protein (FP) and a puromycin selection cassette (P<sup>r</sup>) driven by EF1 $\alpha$  promoter. Transient expression of Cas9 with a guide RNA targeted downstream in exon 2 (open triangle) induced small insertions/deletions (indels) by non-homologous end joining, resulting in frameshift and nonsense mutations (KO2). (B) Western blot for BAG3 protein in iPS-CMs. (C) Immunofluorescent staining and flow cytometry with antibody targeting BAG3 in iPS-CMs. (D) Examples of pathology seen in BAG3<sup>-/-</sup> iPS-CMs compared to wild-type (WT) controls. Cells were plated on glass coverslips, fixed, and stained with antibody to ACTN2 to label Z-disks. All scale bars are 50  $\mu$ m. (E) Quantification of sarcomeric disarray from blinded scoring on a five-point scale, with disarray defined as the percentage of cells scored as class 3–5. Individual replicates are plotted with mean and s.d. Brackets indicate significant difference by one-way ANOVA with Bonferroni's test for multiple comparisons,  $p < 0.001$ . For KO1 analysis 7-9 independent cultures per line from three separate differentiations were scored. For KO2 analysis 4-8 independent cultures per line from two separate differentiations were scored. From each culture >50 cells were scored.

To evaluate myofibril structure and measure sarcomere shortening, we also labeled patterned cells with Lifeact<sup>143</sup>. With this approach, both BAG3<sup>+/-</sup> and BAG3<sup>-/-</sup> cardiomyocytes displayed defects in sarcomere shortening compared to wild-type cardiomyocytes (Figure 3.2 and Supplementary Figure 3.6). BAG3 mutant cardiomyocytes cultured on substrate with physiological stiffness also displayed disrupted myofibrils (Figure 3.2), supporting our initial findings. Thus, our data confirm that partial loss of BAG3 function compromises contractile performance in human cardiomyocytes, localized to disruption of myofibril structure and activity, supporting a dilated cardiomyopathy phenotype.

BAG3 participates in targeting ubiquitinated proteins for degradation via the proteasome or autophagy pathways<sup>141,144</sup>. Thus, we hypothesized that cardiomyocytes with defects in BAG3 co-chaperone activity would be further compromised by proteasome inhibition. Proteasome inhibitors are an important class of compounds recently developed for cancer therapy, although treatment can be complicated by cardiac toxicity, which has been observed in animals and humans<sup>145,146</sup>. To measure the effects of proteasome inhibitors on iPS-CM contractile motion, we used an automated video microscopy system<sup>147</sup>. With this approach, we observed a dose-dependent decrease in iPS-CM contractility after a single exposure to the two FDA-approved proteasome inhibitors, bortezomib and carfilzomib (Supplementary Figure 3.7). These effects occurred at concentrations well within the range of reported plasma concentrations after intravenous infusion in patients<sup>148,149</sup>. Contractility continued to decrease even after the proteasome inhibitors were removed, but it recovered after several days (except at the highest doses). Notably, bortezomib had more potent and longer-lasting effects on contractility than carfilzomib (Supplementary Figure 3.7).

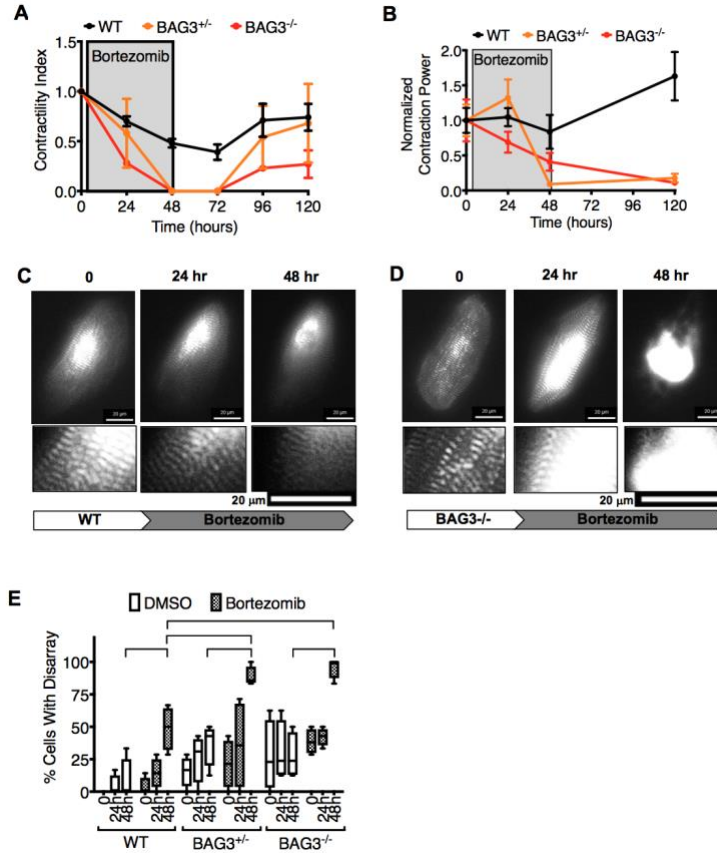




**Figure 3.2 -** BAG3 mutations produce contractile deficits in iPS-CMs cultured on micro-patterned substrates. Day >30 purified iPS-CMs were cultured on micro-patterned polyacrylamide hydrogel substrates with a mechanical stiffness of 10 kPa. BAG3 KO1 mutant lines were used. **(A)** Contraction power was calculated from the measured force and contraction velocity determined by traction force microscopy from the movement of fluorescent beads in the substrate. Results were normalized to wild type and individual replicates plotted with mean and s.d. Brackets indicate significant difference by one-way ANOVA with Bonferroni's test for multiple comparisons,  $p < 0.001$ . **(B)** Sarcomere shortening was measured in Lifact-labeled myofibrils. Results were normalized to wild type and individual replicates plotted with mean and s.d. Brackets indicate significant difference by one-way ANOVA with Bonferroni's test for multiple comparisons,  $p < 0.05$ . Measurements were obtained from three independent device cultures, prepared from two separate differentiation batches. For force measurements 40-51 total cells were analyzed per line. For sarcomere shortening 9-26 total cells were analyzed per line. **(C)** Representative images of patterned Lifact labeled cells, and associated heat maps for surface traction stress (scale in Pa). Scale bars are 20  $\mu\text{m}$ .

With this same assay, we compared the effects of bortezomib on wild-type and BAG3-mutant cardiomyocytes. Exposure to 0.1  $\mu\text{M}$  bortezomib strikingly decreased contractility of BAG3-mutant iPS-CMs compared to wild-type controls (Figure 3.3). To confirm this result, we treated cardiomyocytes cultured on micro-patterned substrates with physiologic stiffness with the same concentration and duration of bortezomib exposure, followed by serial measurements of contraction power. We found that compared to wild-type cardiomyocytes, BAG3-mutant cardiomyocytes had a greater decrease in contraction power during exposure to bortezomib, and failed to recover contractile activity (Figure 3.3).

To determine whether the decrease in contractility caused by bortezomib was a result of disruption of myofilament structure, we expressed a fluorescent-tagged ACTN2 to label Z-disks in live cells<sup>150</sup>. Cells expressing the transgene were viable and continued to contract normally. We used time-lapse fluorescence microscopy to image cells with unambiguously labeled myofibrils at baseline, which revealed defects in myofilament structure within 48 hours of bortezomib treatment (Figs. 3c–e). We used the aforementioned scoring system to quantify the degree of myofibrillar disarray in these treated samples. As expected, BAG3<sup>+/-</sup> and BAG3<sup>-/-</sup> cardiomyocytes displayed more severely disrupted myofilament structure after treatment with bortezomib than wild-type cardiomyocytes (Figure 3.3e). These results support that BAG3 mutants failed to compensate for proteotoxic stress induced by proteasome impairment.

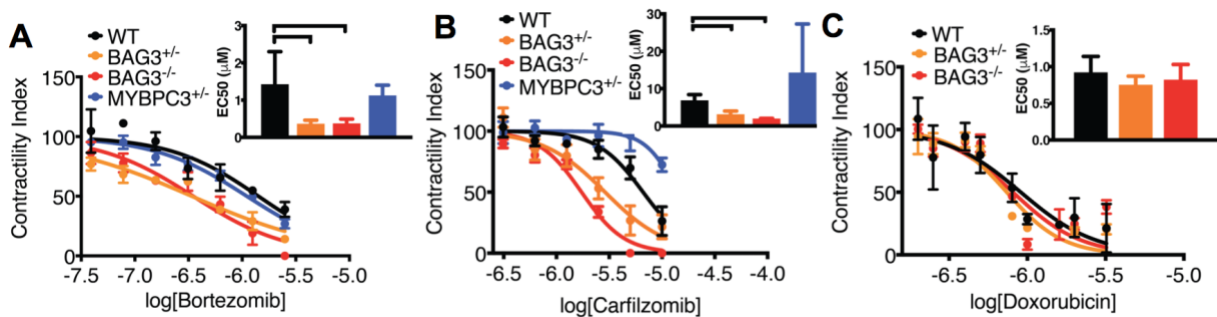


**Figure 3.3** - BAG3 is required to prevent severe cardiotoxicity from the proteasome inhibitor bortezomib. **(A)** Automated video microscopy system (Celloly Pulse) was used to serially measure iPS-CM contractility before and every 24 hr after exposure to bortezomib (0.1  $\mu$ M). WT and BAG3-mutant (KO1) iPS-CMs were exposed to drug for 48 h and then allowed to recover in RPMI/B27 media for 3 days. Contractility index represents the contraction peak height at each time point normalized to the baseline value for each well. Mean and s.e.m. are plotted from 4-8 independent replicates. **(B)** WT and BAG3-mutant iPS-CMs cultured on micropatterned substrates mimicking physiologic stiffness (10 kPa) were exposed to bortezomib (0.1  $\mu$ M) for 48 h and then allowed to recover in RPMI/B27 media for 3 days. Contraction power was measured at each time point and normalized to the baseline value for each population. Plotted are the mean and s.e.m. of measurements from 9-24 cells at each time point. **(C-D)** iPS-CMs were transfected with plasmid expressing fluorescent fusion protein ACTN2-mKate2 to label Z-disks. Individual cells were imaged at baseline and by time-lapse microscopy every 24 hr after treatment with DMSO (0.01%) or bortezomib (0.1  $\mu$ M). Scale bars are 20  $\mu$ m. **(E)** Individual cells were scored at each time point using the five-point scoring system. The percentage of cells scored in class 3-5 was determined from four separate cultures for each condition. Plotted as mean with boxes representing interquartile range and whiskers showing min-max. Brackets represent significant difference by two-way ANOVA with Bonferroni's test for multiple comparisons,  $p < 0.05$ .

To ensure that the interaction between BAG3 mutations and proteasome inhibitors represented a specific biological mechanism rather than a general phenotype of myopathic disease, we performed several pharmacological experiments. For example, disparate genetic models of dilated cardiomyopathy in human iPS-CMs revealed a defective chronotropic response to adrenergic stimulation, along with an exacerbated disease phenotype after repetitive dosing of norepinephrine<sup>151,152</sup>. However, BAG3-mutant lines displayed a positive chronotropic response to stimulation of both  $\alpha$ - and  $\beta$ -adrenergic receptors by phenylephrine and isoproterenol, respectively (Supplementary Figure 3.8). Treatment with norepinephrine produced a similar chronotropic response in all cell lines over repeated dosing, without loss in contractility (Supplementary Figure 3.8). We also did not observe a significant increase in myofibrillar disarray after chronic adrenergic stimulation (data not shown).

Next, we performed dose-response assays to examine the effects of bortezomib and carfilzomib on contractility and viability using wild-type, BAG3-mutant, and MYPBC3-mutant iPS-CMs (Figure 3.4). MYPBC3 is commonly mutated in cardiomyopathies<sup>14</sup>, and based on affinity-purification mass spectrometry, it likely interacts with the BAG3 chaperone complex (see below). The EC<sub>50</sub> for the inhibition of contractility by bortezomib and carfilzomib was significantly lower in both BAG3<sup>+/-</sup> and BAG3<sup>-/-</sup> cells compared to wild-type cells, but not MYPBC3<sup>+/-</sup> cells (Figure 3.4). These effects on contractility were not simply due to cell death, since the effects of bortezomib on cell viability was only significantly different in the BAG3<sup>-/-</sup> cells, and carfilzomib had minimal effects on viability at doses that suppressed contractility (Supplementary Figure 3.9). In contrast, carfilzomib minimally affected viability in wild-type and MYPBC3<sup>+/-</sup> cells, even at the high doses that impaired contractility. To

further confirm specificity of drug toxicity, we repeated these assays with doxorubicin, a cardiotoxic chemotherapy agent. By measuring doxorubicin toxicity by decreases in both contractility and viability, we found no evidence of increased toxicity in BAG3-mutant cardiomyocytes (Figure 3.4, Supplementary Figure 3.9). Finally, for each drug, the EC<sub>50</sub> in wild-type cells was very close to the reported peak plasma concentration after intravenous infusion in patients, confirming that the assay was sensitive at clinically relevant doses<sup>148,149,153</sup>.

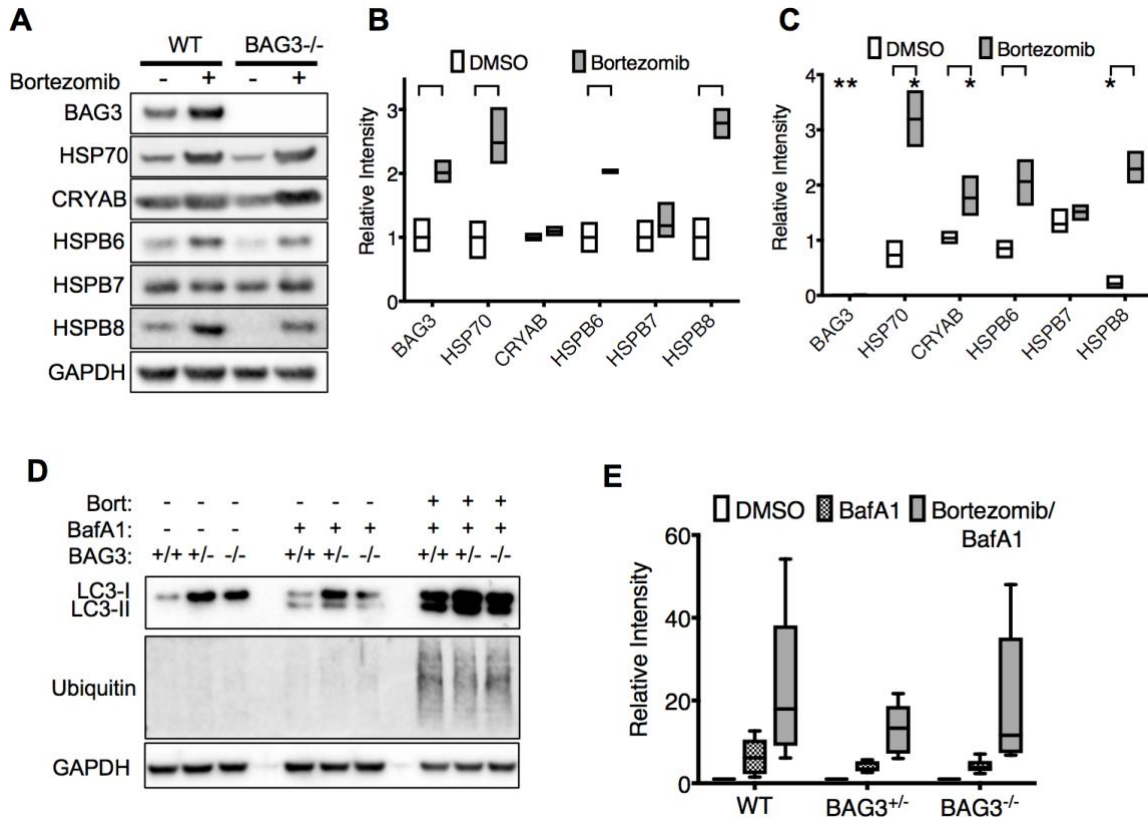


**Figure 3.4** - Genetic and pharmacologic specificity of the interaction between BAG3 and proteasome inhibitors. Wild type, BAG3-mutant (KO2), and MYBPC3 mutant day >30 iPS-CM were treated with (A) bortezomib, (B) carfilzomib, or (C) doxorubicin at the doses indicated. Contraction peak height was measured prior to and 48 hours after drug exposure, and results normalized to the baseline for each independent sample (contractility index). Mean and s.e.m. are plotted for triplicate samples at each dose. Inset graphs represent the calculated EC<sub>50</sub> and 95% confidence interval from each corresponding dose response curve using non-linear regression analysis. Brackets indicate significant differences by one-way ANOVA with Bonferroni's test for multiple comparisons,  $p < 0.05$ .

We hypothesized that the genetic knockout of BAG3 would change the expression of specific BAG3-interacting chaperones that participate in a coordinated stress response. We found that genetically knocking out BAG3 led specifically to a loss of HSPB8 protein levels, while other chaperones were unaffected (Figs. 5a–c). Consistent with findings in immortalized cell lines, BAG3 and HSP70 protein levels were significantly induced by treatment with

bortezomib (Figs. 5a,b)<sup>141,153</sup>. Similar findings occurred with MG132 and carfilzomib (data not shown). Surprisingly, of the cardiac-enriched, small HSP proteins (HSPB5-8), only HSPB6 and HSPB8 were induced by proteasome inhibition in wild-type cells (Figs. 5a,b). On the other hand, CRYAB (HSPB5) protein accumulated specifically in BAG3<sup>-/-</sup> cells after bortezomib treatment (Figs. 5a,c).

Our results implicate HSPB8 as a particularly important partner for BAG3 function in cardiomyocytes. Previous studies reported that BAG3 and HSPB8 are required to activate autophagy to degrade proteotoxic peptides<sup>41,141,144</sup>. Therefore, we hypothesized that loss of BAG3 expression, combined with proteasome inhibition, would impair the ability of cardiomyocytes to regulate autophagy that controls compensatory degradation of ubiquitinated proteins. To measure autophagy flux, we used a Western blot assay to monitor LC3 protein levels with and without bafilomycin A1, which inhibits autophagic degradation. The rapidly degraded LC3-II was increased four–six-fold with bafilomycin A1 treatment, consistent with active autophagy flux in cardiomyocytes. As expected, additional treatment with bortezomib dramatically increased LC3-II levels, consistent with a compensatory induction in autophagy (Figs. 5d,e). However, neither heterozygous nor homozygous knockout of BAG3 significantly changed the levels of LC3-II in any condition. Furthermore, the accumulation of ubiquitinated proteins after bortezomib treatment was the same in wild-type and BAG3-deficient cardiomyocytes (Figure 3.5d). These findings indicate that BAG3 is not required for the regulation of bulk autophagy flux in response to proteasome inhibition in human cardiomyocytes; however, it does not rule out a role for targeting specific clients to this pathway.

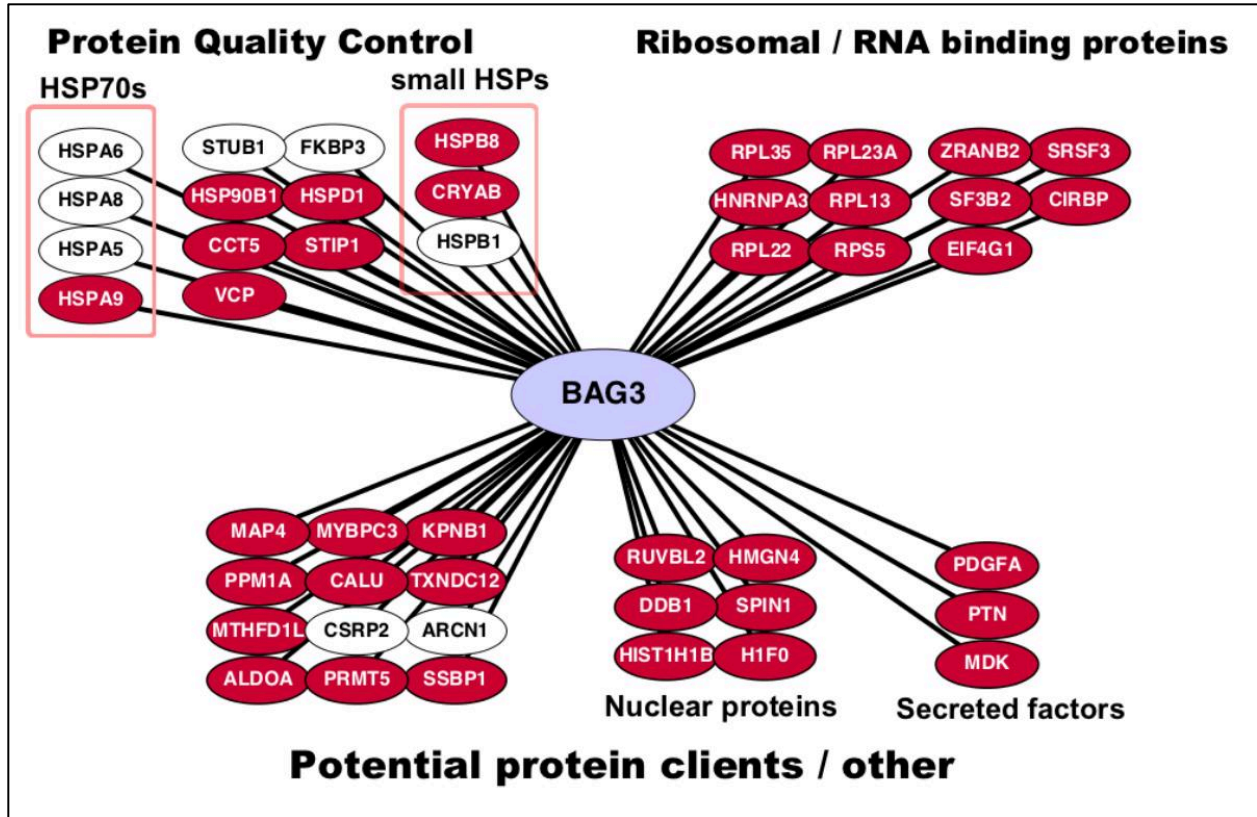


**Figure 3.5** - Bortezomib induces a cardiac chaperone stress-response and increases autophagy flux, the latter of which does not require BAG3. (A) iPS-CMs were treated for 20 hr with vehicle control or 1  $\mu$ M bortezomib, followed by protein extraction and Western blot analysis as shown. (B–C) Band intensities were quantified relative to GAPDH loading control and normalized to vehicle-treated WT control. WT cardiomyocytes treated with vehicle and bortezomib are compared in (B), with BAG3<sup>-/-</sup> cardiomyocytes treated with vehicle and bortezomib compared in (C). Mean and range of triplicate samples are plotted with brackets indicating significant differences from corresponding vehicle-treated samples, \* indicating BAG3<sup>-/-</sup> samples with significant difference from corresponding wild type samples using two-way ANOVA with Bonferroni’s test for multiple comparisons,  $p < 0.05$ . (D) iPS-CMs were treated with DMSO, 100 nM bafilomycin A1 (BafA1) for 6 h, or 1  $\mu$ M bortezomib for 14 hr followed by 100 nM bafilomycin A1 for 6 hr. Total protein extracts were prepared and Western blot performed using antibody to LC3A/B. Representative blot from five biological replicates is shown. (E) LC3-II band intensities were quantified relative to GAPDH loading control and normalized to vehicle-treated sample for each cell line. Plotted as mean of five experiments with boxes representing interquartile range and whiskers showing min-max. There was no significant difference between cell lines using two-way ANOVA and alpha 0.05.

To identify novel and cardiac-specific components and/or clients of the BAG3 chaperone complex, we performed immunoprecipitation followed by mass spectrometry (AP-MS)<sup>154</sup>. We used genome engineering to knock-in a C-terminal 3xFLAG affinity tag into the endogenous BAG3 locus (Supplementary Figure 3.10). We then performed AP-MS on both iPS-CMs and undifferentiated iPSCs. In both cases we also performed AP-MS on cells lacking a 3xFLAG tag in order to exclude non-specific interactions. We identified a total of 46 high-confidence interactions with the endogenous BAG3-FLAG in iPS-CMs (Figure 3.6, Supplementary Table). We found that chaperones and related proteins comprised the single largest group of interactors that we identified, supporting our hypothesis that BAG3 regulates protein quality control pathways in the heart. We did not observe any proteins directly associated with autophagy in our interaction network, which is consistent with the lack of effect of BAG3 in regulating LC3 flux. As expected, the primary overlap between interactors identified from undifferentiated iPSCs and iPS-CMs occurred among ubiquitously expressed chaperones and their associated proteins. However, additional chaperone-related proteins were identified only from cardiomyocytes with the endogenous BAG3 tag, including HSPB8 and CRYAB. We were interested to find that other small HSPs expressed in cardiomyocytes were not present, including HSPB6 and HSPB7. Interestingly, several ribosomal and other RNA-binding proteins were also identified as part of the cardiac-specific BAG3 complex. Although unpredicted, these results support a recently described role for the BAG3-HSP70-HSPB8 complex in processing stress granules and defective ribosomal products, which may represent an underappreciated aspect of protein quality control<sup>155</sup>. The remaining identified



interactors were a variety of cytoplasmic, sarcomeric, nuclear, and secreted proteins that could be targets of the BAG3 chaperone complex, or they suggest a link to other cellular functions.



**Figure 3.6** - Network representing cardiac-specific BAG3 protein interactions. All 46 interactors identified by AP-MS in iPSC-CMs with endogenous BAG3-3xFLAG tag were manually grouped by functional categories. White circles indicate interactors that were also identified from undifferentiated iPSCs, red were only identified from iPSC-CM.

### 3.4 Discussion

Our study shows that disease phenotypes consistent with dilated cardiomyopathy can be recapitulated by *de novo* engineering a series of isogenic BAG3 mutations in an iPSC line from a healthy donor. Previous iPSC studies of genetic cardiomyopathy and drug toxicity have taken an opposite approach that relies on patient-derived samples, with some using genome

engineering to “correct” mutations to create an isogenic control<sup>156–158</sup>. Although powerful, this approach often requires the comparison of cell lines in a mutational series with different genetic backgrounds. Our approach, however, allows the direct comparison of different mutations in a single isogenic background, which will be crucial for further dissecting mechanisms that correlate genotype and phenotype, as seen with BAG3. Indeed, missense mutations in conserved domains of the BAG3 protein cause different, although severe, clinical phenotypes<sup>16,52</sup>. For simplicity, we first modeled the genetic loss of BAG3 function, but future studies will characterize the functional effects of missense mutations in key protein-protein interaction domains. Most importantly, we discovered that heterozygous BAG3 mutations that lead to a ~50% reduction in the BAG3 protein produce a phenotype in human cardiomyocytes that recapitulates human disease. Surprisingly, in several assays, both partial and complete loss of BAG3 function similarly affected cardiomyocyte physiology. The importance of strictly regulating BAG3 expression levels is further highlighted by the observation that several stimuli that lead to cellular stress induce BAG3 expression<sup>58</sup>. This effect likely explains why BAG3 mutations have been associated with both inherited and sporadic cardiomyopathy, the latter of which may result from mutations that subtly affect BAG3 function or expression and manifest as a failure to compensate for an acquired stress. These factors could include infection, ischemia, medication, or exposure to other toxins. Conversely, manipulating the pathways regulated by BAG3 may also protect against cardiac injury.

We studied the effect of proteasome inhibitors on BAG3-mutant cardiomyocytes for several reasons. First, proteasome inhibitors are useful probes to induce proteotoxic stress,

because they inhibit degradation of ubiquitinated proteins, leading to accumulation of potentially aggregation-prone and toxic peptides. Second, proteasome inhibition induces BAG3 and related chaperones, which may facilitate alternative degradation pathways<sup>141,153</sup>. There is limited evidence that BAG3 may itself directly regulate the proteasome<sup>57</sup>. Finally, in a clinical setting, proteasome inhibitors cause cardiac toxicity in an unpredictable subset of patients<sup>12/27/2017 11:33:00 PM</sup>. We demonstrate that even partial loss of BAG3 function in human cardiomyocytes leads to severe decompensation in contractile function and myofilament integrity after a single exposure to bortezomib or carfilzomib, confirming that, importantly, BAG3 co-chaperone function protects against proteotoxic stress. This finding has potential clinical implications, because patients with genetic variants in BAG3 or functionally related genes could be predisposed to severe cardiotoxicity from this class of medication. In addition, human iPS-CM models such as ours could be useful for developing safer versions of these and other drugs.

Our data support a model in which BAG3 maintains myofibril function and protects the heart from injury by interacting with multiple families of chaperone proteins. Rigorous biochemical studies demonstrate that BAG3 acts as a co-chaperone by linking the HSP70 and HSPB families and modulating their function, while cellular experiments support the activity of a BAG3-HSP70-HSPB ternary complex<sup>41,155</sup>. Using AP-MS, we identified specific proteins in the HSP70 and small HSP families that interact with BAG3 in cardiomyocytes, which are also co-regulated by proteotoxic stress. For example, our data indicate that CRYAB and HSPB8 are the most relevant small heat shock proteins interacting with BAG3 in cardiomyocytes. We further demonstrated that an endogenous epitope tag allowed us to study

protein interactions at physiologic levels of expression in the cell type of interest. These data highlight the tissue-specific nature of its role in cardiac biology. With this approach, we identified several novel BAG3-interacting proteins. These include VCP, mutations in which cause inclusion body myopathy, and MYPBC3, which is commonly mutated in hypertrophic and dilated cardiomyopathy<sup>14,159</sup>. Interestingly, MYPBC3 was the only sarcomeric protein that was co-purified with the BAG3 complex, suggesting that it could be an important client protein. Various proteins identified likely represent additional clients of the BAG3 chaperone complex, while others suggest additional functions for BAG3 in regulating the proteome, including at the level of translation by the ribosome. Future genetic and functional studies of these pathways may reveal mechanisms of heart failure, genetic predisposition to drug toxicity, and novel therapeutic strategies.

## 3.5 Materials and Methods

### 3.5.1 Human iPSC culture

Control human iPSC cells were generated from dermal fibroblasts from a healthy male subject (WTc) using the episomal reprogramming method<sup>160</sup>. The subject has a normal ECG and no evidence of cardiac disease with 8 years follow-up. Complete exome sequence was obtained for WTc and is available on our laboratory website. Exons 1–4 of the BAG3 gene were also amplified by PCR and sequenced by the Sanger method, with no BAG3-coding variants identified. Cells were maintained in mTeSR1 (STEMCELL Technologies) or E8 (ThermoFisher Scientific) media on growth-factor-reduced Matrigel (8 µg/ml, BD Biosciences) and passaged every 3–4 days using Accutase (STEMCELL Technologies). ROCK

inhibitor Y-27632 (10  $\mu$ M, Selleckchem) was added to the media for 24 h after each passage. Cells were cryopreserved as a suspension in Fetal Bovine Serum (FBS, Hyclone) containing 10% DMSO and 10  $\mu$ M ROCK inhibitor. Cells were used between passage 25–80 for all experiments. Cultures were monitored for mycoplasma on a quarterly basis using a commercial kit (Stratagene).

### 3.5.2 TALEN and CRISPR construct design

ZiFiT (<http://zifit.partners.org/ZiFiT/ChoiceMenu.aspx>) was used to design BAG3 and MYPBC3 TALENs. TALENs were constructed using the Voytas laboratory Golden Gate assembly system and backbone vector MR015, as described<sup>121</sup>. All TALENs used the NN repeat variable di-residues (RVDs). Target sequences for BAG3 TALENs were: left TALEN 5'-CCTCTGCCAATGGCCG-3' and right TALEN 5'-AGCAGGGCGGCAGCCT-3'. Target sequences for MYBPC3 TALENs were: left TALEN 5'-TGACGTCTCTCAGGATGC-3' and right TALEN 5'-TCCTAAAGCTACCTGGC-3'. For designing CRISPR guide RNA targeting BAG3, we used the Zhang Lab website ([crispr.mit.edu](http://crispr.mit.edu)). The guide sequence was cloned into a gRNA expression vector using the protocol described by the Zhang Lab CRISPR resources on Addgene (<sup>72</sup>). pX330-U6-Chimeric\_BB\_CBh-hSpCas9 was a gift from Feng Zhang (Addgene plasmid #42230). A modified version in which the Cas9 expression cassette was removed (pEN102) was a gift from Elphege Nora and Benoit Bruneau. Sequence of gRNA targeting exon 2 for generating KO2 was 5'-ACCTGTCCACACTGTTTATC-3'. Sequence of gRNA targeting exon 4 for generating C-terminal FLAG was 5'-GCAGAGGCTACGGTGCTGCT-3'.

### 3.5.3 Transfection of iPSCs for gene targeting

Cells were passaged at near confluent density and  $1-2 \times 10^6$  cells were transfected using Amaxa Nucleofector 2b system and Human Stem Cell Solution I, executing program A-023 (Lonza). For TALEN mutagenesis, we transfected WTc with left and right TALEN plasmids and donor plasmid. For FLAG tagging, we transfected WTc with pX330 and donor plasmid. For BAG3 KO2 we transfected the transgenic CRISPRn line (derived from WTc)(51) with pEN102. In the latter case, doxycycline ( $1 \mu\text{g/ml}$ ) was added to the media 24 h before transfection and for 48 h after to transiently induce Cas9 expression. After transfection, cells were immediately plated onto Matrigel-coated plates and cultured for 24 h with media containing ROCK inhibitor.

### 3.5.4 Isolation of modified iPSC clones

Cells transfected with TALEN or Cas9 plasmids along with BAG3-knockout vector, MYPBC3-knockout vector, or FLAG-tag vector were cultured in selective media containing  $0.5 \mu\text{g/ml}$  puromycin and  $10 \mu\text{g/ml}$  ROCK inhibitor, starting 48 h after transfection. Selective media was exchanged daily until stable mCherry-positive colonies remained (5–7 days). After polyclonal enrichment and confirmation of genotype, clonal populations were derived with 1–2 rounds of manual clone picking. For the FLAG-tag line, the pooled population was nucleofected with a plasmid expressing Cre recombinase, and mCherry negative clones were selected. For the CRISPR BAG3 KO2 lines, clones were picked directly without enrichment. For sub-cloning,  $2 \times 10^4$  cells were plated on Matrigel-coated 10-cm tissue culture dishes. After 4–6 days, individual colonies were manually picked and expanded in culture, then divided into

one portion for cryopreservation and another portion for genomic DNA extraction (Qiagen DNAeasy Blood & Tissue kit) and genotyping. Several genotyping strategies were used depending upon the mutation. To genotype clones with integration of BAG3 KO, MYPBC3 KO, or FLAG tag we used PCR assays to distinguish the endogenous and modified alleles. For each predicted integration event we designed primers to amplify across the integration junction at both the 5' and 3' end of the targeting vector, as well as a separate primer pair to amplify the unmodified allele. Sequences for all primers used for genotyping are shown in the supplementary table. For selected clones, we sequenced each PCR product to confirm integrity of the integration junctions and to rule out mutations in the untargeted alleles. For off-target analysis of TALEN-mediated mutant clones we used the TALENoffer software<sup>161</sup>. We chose the top 10 off-target loci as predicted by TALENoffer and designed flanking primers to amplify each region by PCR. The PCR products from all 10 loci for BAG3<sup>+/-</sup> and BAG3<sup>-/-</sup> lines were sequenced and did not demonstrate any TALEN-induced mutations. To genotype clones with Cas9-mediated frameshift mutations (KO2), we PCR-amplified exon 2 and sequenced the products to identify clones with small insertions or deletions. We verified the precise sequence of individual mutant alleles by cloning the PCR products into pCR2.1 (TOPO TA cloning kit, ThermoFisher Scientific) and sequencing multiple plasmid clones with M13 forward and reverse primers. We performed a droplet digital PCR assay to determine the copy number of FLAG and puromycin insertions (see supplementary table) as compared to a reference gene RPP30 (Bio-Rad).

### **3.5.5 Karyotyping**

All karyotyping was performed by Cell Line Genetics (Madison, WI).

### 3.5.6 Pluripotency Staining for iPSCs

Performed as described previously<sup>121</sup>.

### 3.5.7 iPS-CM Differentiation and Culture

We used modifications of the WNT modulation method for directed differentiation of iPSCs into cardiomyocytes<sup>162</sup>. Freshly passaged iPSCs were seeded on Matrigel-coated 12-well plates, with initial seeding density optimized for each line ( $6.25 \times 10^3 - 2.5 \times 10^4$  cells/cm<sup>2</sup>). Approximately 72 h after plating (day 0), media was changed to RPMI1640 with B27 supplement (without insulin) containing CHIR99021 (Tocris). The optimal concentration and exposure time for CHIR was optimized for each line, either 12  $\mu$ M for 24 h or 6  $\mu$ M for 48 h. For high-dose CHIR concentrations, the media was changed to RPMI/B27 (without insulin) on day 1, and then RPMI/B27 (without insulin) containing 5  $\mu$ M IWP2 (Tocris) on day 3. For low-dose CHIR concentrations, media was changed to RPMI/B27 (without insulin) containing 5  $\mu$ M IWP2 on day 2. For both protocols, the media containing IWP2 was left for 48 h and then exchanged changed to RPMI/B27 (without insulin) only. After another 48 h, media was changed to RPMI/B27 containing insulin. Fresh RPMI/B27 was exchanged every 3–4 days thereafter. For purification of iPS-CMs, we used a metabolic selection protocol with glucose-free DMEM containing lactate<sup>163</sup>. Cells were replated on day 15–18 of differentiation, then on day 20–22, media was exchanged for DMEM (without glucose, with sodium pyruvate, ThermoFisher Scientific) supplemented with Glutamax, Non-Essential Amino Acids, and buffered lactate (4 mM). Stock-buffered lactate solution was prepared by dissolving Sodium L-lactate powder (Sigma) at 1M concentration in 1M HEPES solution. Lactate media was



exchanged 2–3 times, with total exposure of 48 h for each treatment. After final lactate treatment (day 24–28) media was changed to RPMI/B27, which was exchanged every 3–4 days thereafter. After day 30, iPS-CMs were either re-plated directly for experiments or cryopreserved. For any given experiment using multiple cells lines, all samples were either directly re-plated in parallel at the same differentiation stage, or thawed and then re-plated in parallel.

### **3.5.8 Immunofluorescent staining of iPS-CMs**

Glass coverslips were placed in 12-well plates and coated with Matrigel. Lactate-purified iPS-CMs were prepared as described above and plated at a density of  $2 \times 10^4$  cells/coverslip. Cells were maintained in culture for 7–14 days and then fixed in 4% paraformaldehyde for 15 min at room temperature. Fixed coverslips were washed repeatedly with PBS containing 0.1% Triton X-100 (PBS-T), then blocked in a solution of 5% bovine serum albumin (BSA, Sigma) in PBS-T at room temperature for 1 h. Primary antibodies were diluted in 5% BSA solution and incubated overnight at 4°C. Coverslips were washed again repeatedly with PBS-T and then incubated with secondary antibodies diluted in 5% BSA solution for 1–2 hours at room temperature. Coverslips were washed a final time and then mounted onto glass slides with VECTASHIELD HardSet with DAPI (Vector Laboratories). Primary antibodies included mouse monoclonal anti-sarcomeric actinin (clone EA-53, Sigma) and rabbit anti-BAG3 (Protein Tech 10599-1-AP). Secondary antibodies included Alexa Fluor 488 goat anti-mouse IgG, Alexa Fluor 594 goat anti-mouse IgG, Alexa Fluor 488 goat anti-

rabbit IgG, and Alexa Fluor 594 goat anti-rabbit IgG (Molecular Probes). All images were taken on a BZ-9000 microscope (Keyence).

### **3.5.9 Flow cytometry**

Flow cytometry was performed as previously described<sup>164</sup> using primary rabbit anti-BAG3 antibody (Protein Tech 10599-1-AP) and mouse monoclonal antibody to cardiac troponin-T (clone 13-11, ThermoFisher).

### **3.5.10 Sarcomere scoring in fixed iPS-CMs**

Day >30 iPS-CMs were cultured on glass coverslips for 10–14 days (initial experiments were performed with a 14-day interval, later experiments with a 10-day interval gave equivalent results). Coverslips were then fixed in 4% paraformaldehyde, stained with anti- $\alpha$ -actinin antibody, and mounted on slides as described above. The slides were then labeled with an alphanumeric code to blind scientists performing the experiment to the identity of each sample. Between 20–24 images were taken with a 40X objective for each slide, with a total of 50–150 cells imaged from each slide. Each cell was then assigned a score between 1 and 5 corresponding to the relative degree of disordered myofilaments as follows: 1 = all myofilaments were continuous and well-ordered with majority in parallel; 2 = all myofilaments were continuous and well-ordered but not in parallel; 3 = significantly disordered myofilaments with fragmentation, disintegration of Z-disk structures, or actinin aggregates making up <50% of the total area; 4 = significantly disordered myofilaments with fragmentation, disintegration of Z-disk structures, or actinin aggregates making up  $\geq$ 50% of the total area; 5 = positive actinin staining, but no identifiable myofilament structure. After all

images were scored, the code was broken and each sample assigned to the corresponding cell line. The percentage of cells in each category was calculated for each slide, and the mean and standard deviation was calculated for replicate samples from each cell line. For analysis of KO1 lines, samples from three independent experiments were combined (WT,  $n = 8$ ; BAG3<sup>+/-</sup>,  $n = 7$ , BAG3<sup>-/-</sup>,  $n = 9$ ). For KO2 lines, samples from two independent experiments were combined (WT,  $n = 7$ ; BAG3<sup>+/-</sup>,  $n = 4$ , BAG3<sup>-/-</sup>,  $n = 8$ ).

### 3.5.11 Analysis of micro-patterned iPS-CMs

Complete methodology has been previously described<sup>165</sup>. In brief, day >30 iPS-CMs were thawed on fibronectin-coated wells and recovered for 5 days before seeding on micropatterned polyacrylamide substrates.  $1 \times 10^5$  singularized hiPSC-CMs in RPMI-1640 medium plus B27 supplement (50X), penicillin (25  $\mu\text{g}/\text{mL}$ ), and streptomycin (50  $\mu\text{g}/\text{mL}$ ) (all from Life Technologies) with 5  $\mu\text{M}$  Y27623 ROCK inhibitor were seeded and cultured on the top of hydrogel devices. Cell culture medium without ROCK inhibitor was changed after two days of culture. After 3 days, 20-30 fps videos of beating single cells and of the movement of microbeads embedded in the hydrogel substrate underneath each cell were acquired in a Zeiss Axiovert 200 M with a Zeiss AxioCam MRm camera, using a 40x objective. Power contractile output of single hiPSC-CMs was calculated by multiplying the force it generates by the velocity of microbeads during contractions. Cell contractile force was calculated by submitting frames within videos of fluorescent microbeads to algorithms of cross-correlation and traction force microscopy developed in MATLAB. Cross-correlation also generated information on the displacement of each frame and time, which were used to calculate microbead velocity. For analysis of baseline contraction power, measurements from two independent differentiation

batches were combined for 10kPa substrate (WT,  $n = 51$ ; BAG3<sup>+/-</sup>,  $n = 40$ , BAG3<sup>-/-</sup>,  $n = 42$ ) as well as 35-kPa substrate (WT,  $n = 40$ ; BAG3<sup>+/-</sup>,  $n = 11$ , BAG3<sup>-/-</sup>,  $n = 47$ ).

Micropatterned hiPSC-CMs were transfected with RFP-LifeAct (Ibidi) to label sarcomeres and sarcomere length was calculated from the distance between Z-lines in beating cells. For this purpose, we acquired videos of fluorescently-labeled micropatterned beating cells. Sarcomere shortening of each cell resulted from subtracting average sarcomere length of the cell in its contracted state from the average sarcomere length of the cell in its relaxed state. For analysis of sarcomere shortening, measurements from cells seeded on 10-kPa substrate from two independent differentiation batches were combined (WT,  $n = 26$ ; BAG3<sup>+/-</sup>,  $n = 9$ , BAG3<sup>-/-</sup>,  $n = 17$ ). Measurements from cells seeded on 35-kPa substrate were from a single differentiation batch (WT,  $n = 8$ ; BAG3<sup>-/-</sup>,  $n = 7$ ).

### **3.5.12 Live sarcomere labeling and scoring in iPS-CMs**

Day >30 iPS-CMs were dissociated and counted as described above. Then  $0.5\text{--}1.5 \times 10^6$  cells were resuspended in 100  $\mu\text{l}$  Human Stem Cell Solution I with 2  $\mu\text{g}$  of  $\alpha$ -actinin-mKate2 reporter plasmid and transfected with Amaxa Nucleofector 2b system, executing program G-009 (Lonza). After transfection, cells were immediately plated onto Matrigel-coated 48-well plates and cultured for 24 h with media containing ROCK inhibitor. The cells recovered in culture for 7 days before starting experiments. Imaging was performed on a Zeiss AxioObserver microscope with an automated stage. On day 0 of the experiment, 5–6 individual 40X fields-of-interest containing cells with well-labeled sarcomeres were acquired from each well, and the X/Y/Z coordinates were saved for each field. Fresh media containing

either 0.01% DMSO or 0.1  $\mu$ M bortezomib was added to each well, with 4 replicates per cell line treated with each condition. The saved coordinates were imaged again at 24 and 48 h after addition of drug. After all images were acquired, each cell was assigned a score for degree of myofilament disarray as described above. The percentage of cells in each category was calculated for each well, and the mean and standard deviation was calculated from four replicate wells for each condition.

### **3.5.13 Drug treatment and contractility assays in iPS-CMs**

Day >30 iPS-CMs were dissociated and counted as described above. Cells were plated in 48-well or 96-well plates at a density of  $4 \times 10^4$  or  $2 \times 10^4$  cells/well respectively. Cells were maintained in culture for 7–10 days before beginning experiments. Bortezomib (Cell Signaling Technology), carfilizomib (Selleckchem), and doxorubicin (Cell Signaling Technology) were dissolved in DMSO at 10 mM concentration, filtered through a 0.22  $\mu$ M syringe filter, and stored in aliquots at  $-20^\circ\text{C}$ . Contractile motion was captured by automated video microscopy and analyzed using the Cellogy Pulse system. Fresh media was exchanged, and the cells equilibrated in the incubator for 2–4 hours before obtaining baseline data. Drugs or vehicle were diluted to a 3X final concentration in complete media and added to the existing media in each well for the corresponding final concentration. Data were acquired on the Cellogy Pulse system every 24 h, and the basal media was exchanged 48 h after addition of drugs and again every 48 h after. Contractility index was calculated by dividing the contraction peak height at each time point from the baseline value for each well. Normalized beat rate was calculated in a similar fashion.

For testing adrenergic agonists, 10 mM norepinephrine bitartrate (Sigma) was prepared in water with 100 mM citric acid and 10 mM phenylephrine hydrochloride (Sigma) and 10 mM isoproterenol hydrochloride (Sigma) were prepared in water. All solutions were filtered through a 0.22  $\mu$ M syringe filter. Norepinephrine and phenylephrine were aliquoted and stored at -20°C, and isoproterenol was prepared fresh. Baseline data were obtained as described and then again 1 h after addition of drug. For repetitive norepinephrine treatment, fresh media was exchanged each morning with baseline data obtained 4 h later. Additional media containing 3X norepinephrine or vehicle was then added with data acquisition 1 h later. This process was repeated for 5 consecutive days.

#### **3.5.14 Resazurin viability assay in iPS-CMs**

Cells used for contractility assays recovered for 5 days after removal of drug. Then, 7 days from the beginning of the experiment, media was exchanged for fresh RPMI/B27 with 10% PrestoBlue reagent (ThermoFisher) and incubated for 1 h at 37°C. Fluorescence intensity was measured with 560 nM excitation and 590 nM emission on a SpectraMax i3 plate reader (Molecular Devices). Background measurement from wells with media only was subtracted from all samples. For each cell line, the fluorescence intensity was normalized to control wells with no drug added.

#### **3.5.15 Western blot analysis**

Cells were dissociated with accutase (iPSCs) or 0.25% trypsin/EDTA (iPS-CMs), and cell pellets were washed once with PBS. Cells were lysed in RIPA buffer containing Complete protease inhibitors (Roche), and the protein concentration was measured using a Bradford-

based assay (Bio-Rad). Samples were prepared at equal concentration in NuPAGE LDS sample buffer, loaded on NuPAGE 4–12% Bis-Tris gels, and ran in NuPAGE MES running buffer (ThermoFisher Scientific). Proteins were transferred to nitrocellulose membranes using the iBlot transfer system (ThermoFisher Scientific). Membranes were blocked in Odyssey blocking buffer (LI-COR) for 1 h at room temperature. Primary antibodies were diluted in blocking buffer with 0.1% Tween 20. Membranes were incubated in primary antibody solution at 4°C overnight, then washed with three aliquots of PBS with 0.1% Tween 20 (PBS-T). Fluorescent-conjugated secondary antibodies were diluted in blocking buffer with 0.1% Tween 20 and 0.01% SDS. Membranes were incubated in secondary antibody solution for 1 h at room temperature, then washed with three aliquots of PBS-T. After the final washing step, membranes were imaged immediately with an Odyssey Fc fluorescent imaging system (LI-COR). Individual band intensities were measured using Image Studio software (LI-COR), and the intensity of each band-of-interest was normalized to the loading control (GAPDH) for the corresponding sample. Primary antibodies used included mouse anti-HSP70/HSP72 (clone C92F3A-5, Enzo Life Sciences), rabbit anti-BAG3 (Protein Tech 10599-1-AP), mouse anti-alphaB crystallin (clone 1B6.1-3G4, Enzo Life Sciences), rabbit anti-Hsp20 (HSPB6, Abcam ab13492), mouse anti-cvHSP (HSPB7, Abcam ab57093), rabbit anti-Hsp22 (HSPB8, Abcam ab96837), mouse anti-FLAG (Sigma, F3165), and rabbit anti-GAPDH (Abcam ab9485). Secondary antibodies used included IRDye 680LT donkey anti-mouse and IRDye 800CW donkey anti-rabbit (LI-COR).

### 3.5.16 Affinity Purification – Mass Spectrometry of BAG3 interacting partners

For AP-MS, two replicates each of BAG3<sup>3xFLAG/3xFLAG</sup> undifferentiated iPSCs and differentiated iPS-CMs were used. Two replicates of WTc iPSCs and iPS-CMs (not expressing BAG3-3xFLAG) were used as controls for non-specific interactions. For iPSC pulldowns, 4 x 15cm dishes at ~80% confluence were used. For iPS-CM pulldowns, 25–30x10<sup>6</sup> lacatate-purified cells at day 30 of differentiation were used. Cells were harvested in PBS by scraping and resuspended in 2 volumes of lysis buffer (0.1% NP-40, 300 mM NaCl, 25% Glycerol, 2 mM MgCl<sub>2</sub>, 0.5 mM EDTA, 0.5 mM EGTA, 1 mM PMSF, 1 mM DDT in 50 mM HEPES-NaOH pH 8.0, supplemented with complete protease inhibitor cocktail (Roche) and Benzonase Nuclease (50 U/ml, Sigma-Aldrich). Cells were lysed by four freeze-thaw cycles, followed by incubation for 20 min at 4°C. After removing the insoluble fraction, protein extracts were diluted three-fold to reduce salt content and incubated with 30 [1 anti-FLAG M2 Magnetic Beads slurry (Sigma-Aldrich) for 2 h. Beads were rinsed in wash buffer (3x washes in 0.01% NP-40, 1 mM PMSF, 0.5% EDTA, 0.5% EGTA in 50 mM HEPES, pH 8.0; 1x wash on buffer without NP-40). FLAG-enriched proteins were reduced (5 mM TCEP), alkylated (15 mM iodoacetamide), and digested with 1%(w/v) trypsin overnight. Resulting peptides were desalted by UltraMicroSpin columns (The Nest Group) and dried.

Peptides were resuspended in 0.1% formic acid and analyzed by liquid chromatography tandem mass spectrometry (LC MS/MS) with an Easy-nLC 1000 (Thermo Fisher, San Jose, CA) coupled to an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA). Online LC separation was carried out using a 75 μm x 25 cm fused silica IntregraFrit



capillary column (New Objective, Woburn, MA) packed in-house with 1.9  $\mu\text{m}$  Reprisil-Pur C18 AQ reverse-phase resin (Dr. Maisch-GmbH). Peptides were eluted at a flowrate of 300 nL/min using a linear gradient of 5–30% B in 45 min, and 30–95% B for 25 min (mobile phase buffer A: 100 H<sub>2</sub>O/0.1% formic acid; mobile phase buffer B: 100%ACN/0.1% formic acid). Survey scans of peptide precursors from 400 to 1600 m/z were performed at 120K resolution in the Orbitrap, with an AGC target of  $2 \times 10^5$ , and a maximum injection time of 100 ms. Tandem MS (MS<sup>2</sup>) was performed by isolation with the quadrupole, HCD fragmentation with normalized collision energy of 30%, and rapid scan MS analysis in the ion trap. The MS<sup>2</sup> ion count target was set to  $10^4$  and the max injection time was 35 ms. Precursors with charge state 2–7 were sampled for MS<sup>2</sup> and dynamically excluded for 20 s (tolerance of 10 ppm). Monoisotopic precursor selection was turned on, and the instrument was run in top speed mode with 3-s cycles.

For protein identification and quantification, MaxQuant software v1.5.3.30 was used<sup>166</sup>. Tandem mass spectrometry (MS/MS) spectra were searched against the November 2016 release of the UniProt complete human proteome sequence database, modified to include the FLAG peptide sequence. MaxQuant was run on default parameters, allowing for 2 maximum missed cleavages, with a first search peptide tolerance of 20 ppm and a main search peptide tolerance of 4.5 ppm. Methionine oxidation and N-terminal acetylation were set as variable modifications, and carbamidomethylation of cysteines as fixed modification. The ‘match between runs’ setting was activated (window of 7 min) to improve peptide identification.

For the analysis of data, proteins with two or fewer peptides identified were discarded, as were typical common contaminant proteins (downloaded from <http://maxquant.org/>). Since the two sets were run separately, protein intensity ratios (BAG3<sup>3xFLAG</sup>/<sup>3xFLAG</sup>/ WTc) were calculated separately for each cell type and for every replicate. Intensity values for missing proteins in each sample were inferred as 10% of the lowest intensity value for that sample. Z-scores were calculated and p-values were corrected for multiple testing using the Benjamini-Hockberg procedure<sup>167</sup>. We selected as interactors those proteins with a significant p-value ( $p < 0.05$ ) across the replicates. A list of significant hits can be found in the Supplementary Table).

### **3.5.17 Software and statistical analysis**

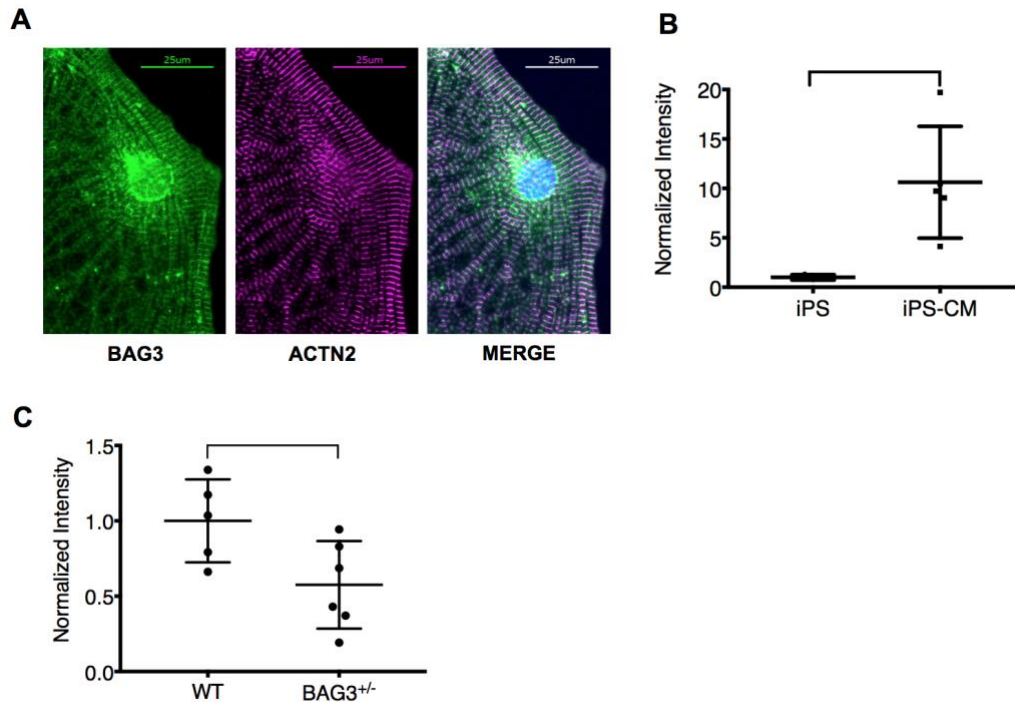
Data analysis and graphing was performed with GraphPad Prism v6.0h software. Statistical tests were performed in Prism using an unpaired two tailed *t*-test with equal variance, or ANOVA with Bonferroni's test for multiple comparisons. A p-value less than 0.05 was considered significant. EC<sub>50</sub> and LD<sub>50</sub> values with 95% confidence intervals were calculated from dose-response curves in Prism using log(inhibitor) vs. normalized response with variable slope. Analysis of BAG3-interacting partners was done using R<sup>168</sup> v3.3.0, and the visualization figure was created using Cytoscape v3.4.0<sup>169</sup>. Generation of Venn diagrams was done using BioVenn<sup>170</sup>.

### 3.5.18 Study approval

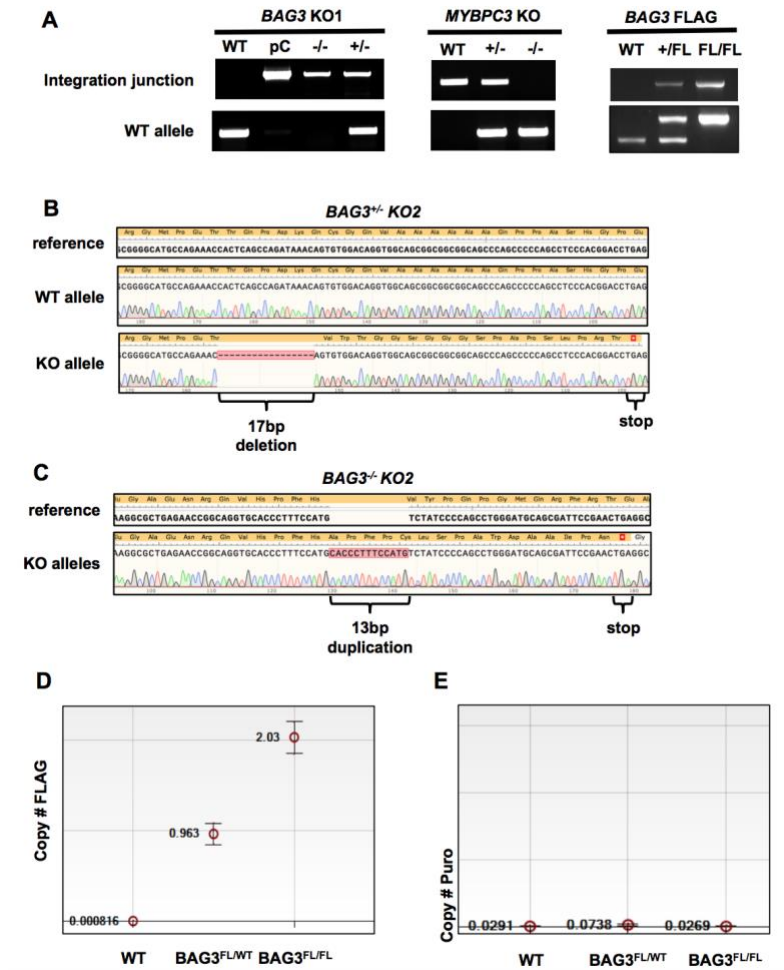
Derivation and use of human iPSCs was approved by the UCSF Committee on Human Research, San Francisco, CA (study #10-02521). All subjects provided informed consent prior to participation.

**Author Contributions:** L.M.J. and B.R.C. conceived the project. L.M.J. designed the majority of experiments and performed them with assistance from A.T., C.L.J., and J.Y. A.T. and J.P.B. generated the endogenous BAG3-3xFLAG iPSC line. J.P.B. performed affinity-purification and mass-spectrometry analysis with assistance from R.K. A.J.S.R generated the micro-patterned devices and performed all associated assays. J.Y. and M.A.M. generated the MYPBC3<sup>+/-</sup> iPSC line. N.H. assisted with design and construction of  $\alpha$ -actinin fluorescent reporter plasmid. B.L.P., D.S., N.J.K., and P.L.S. contributed to valuable conceptual discussions and experimental planning. L.M.J. and B.R.C. composed the manuscript with assistance from all authors. Correspondence and requests for materials should be addressed to [bconklin@gladstone.ucsf.edu](mailto:bconklin@gladstone.ucsf.edu).

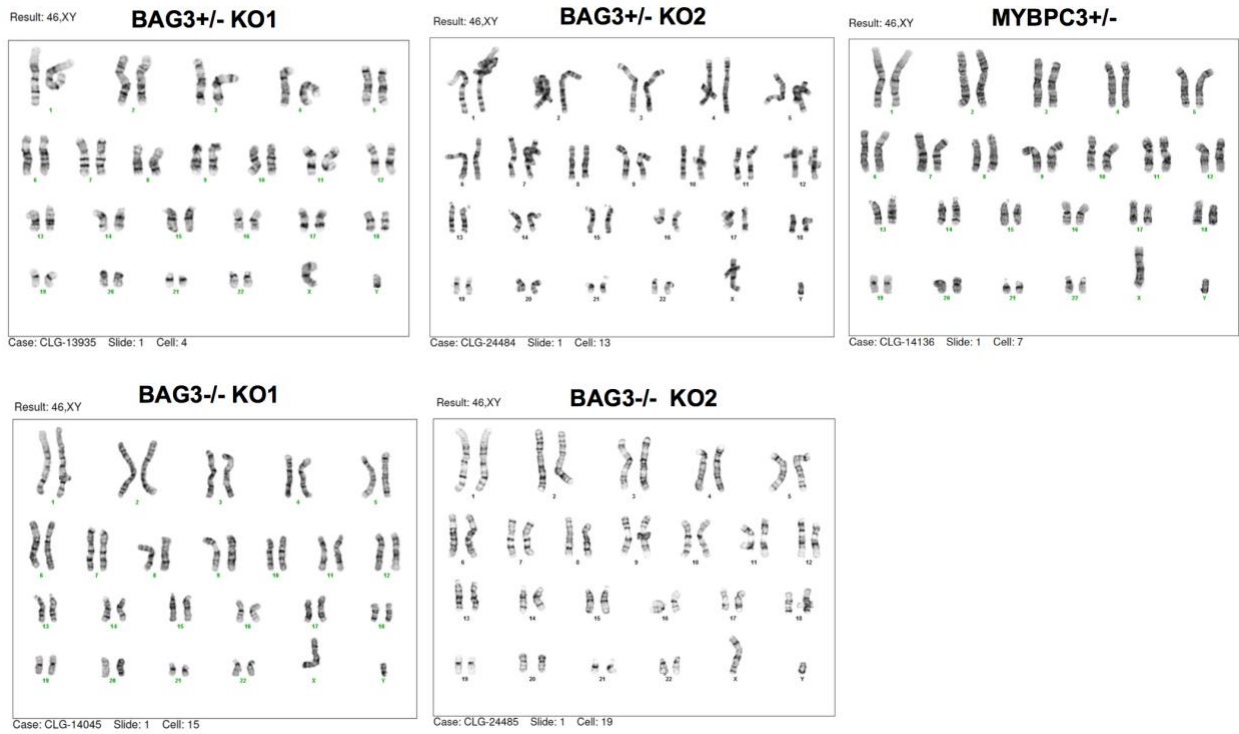
## 3.6 Supplementary Figures



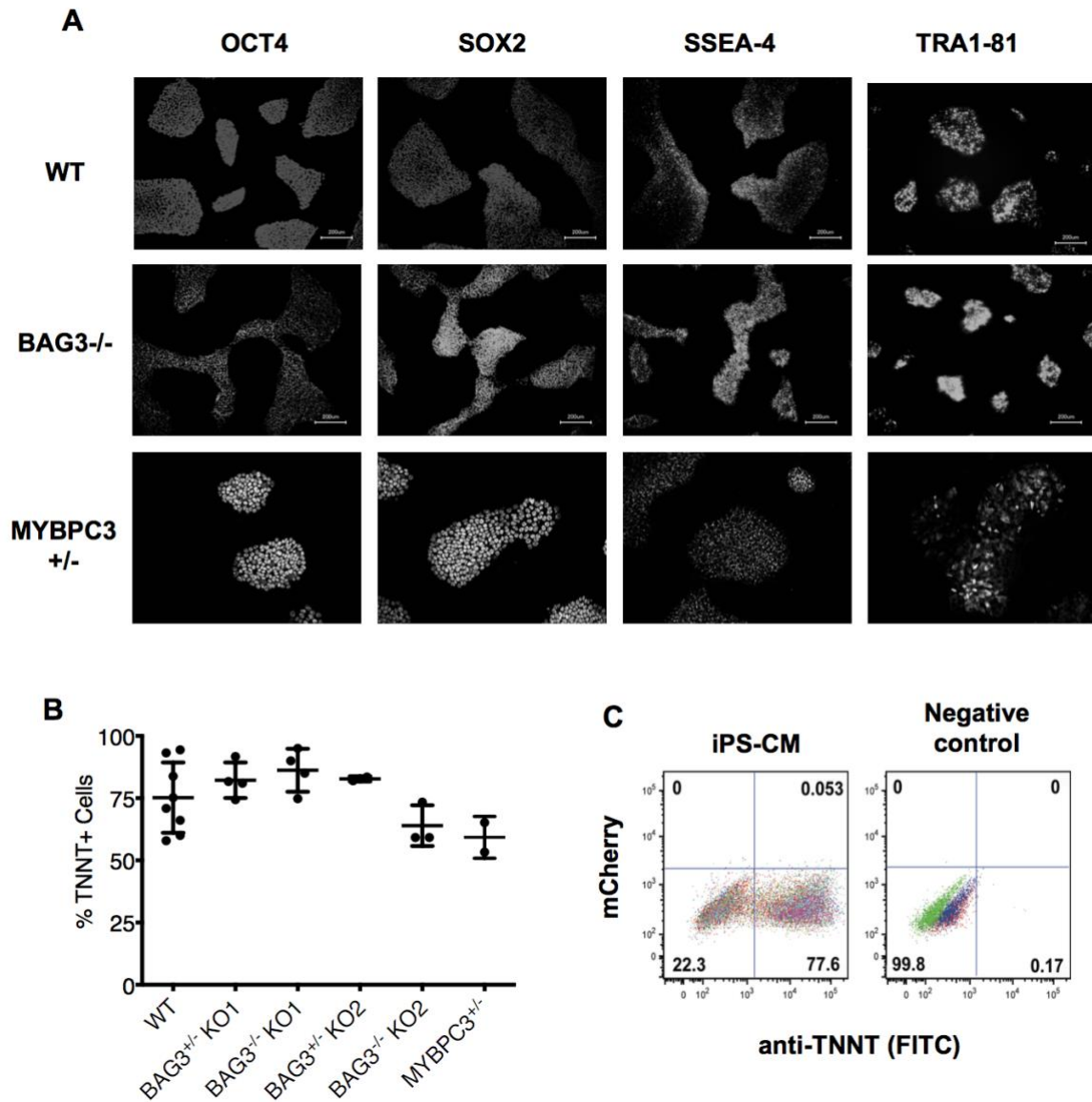
**Supplementary Figure 3.1** - Expression of BAG3 protein in human iPS and iPS-CMs. (A) Cardiomyocytes plated on glass surfaces were fixed and stained with antibodies to BAG3 (green) and sarcomeric  $\alpha$ -actinin (ACTN2, magenta). Nucleus stained with DAPI (blue). Scale bars are 25  $\mu$ m. (B) Total protein extracted from undifferentiated iPS (n = 3) and Day >30 iPS-CM (n = 4) was analyzed by Western Blot with BAG3 antibody. Band intensities were calculated relative to GAPDH loading control and normalized to iPS samples. (C) Total protein extracted from Day >30 wild type (n = 5) and BAG3<sup>+/-</sup> (n = 6) iPS-CM was analyzed by Western Blot with BAG3 antibody. Band intensities were calculated relative to GAPDH loading control and normalized to wild type samples. Graphs plotted as individual values with mean and s.d. Brackets indicate significant difference by two-tailed students-t-test,  $p < 0.05$ .



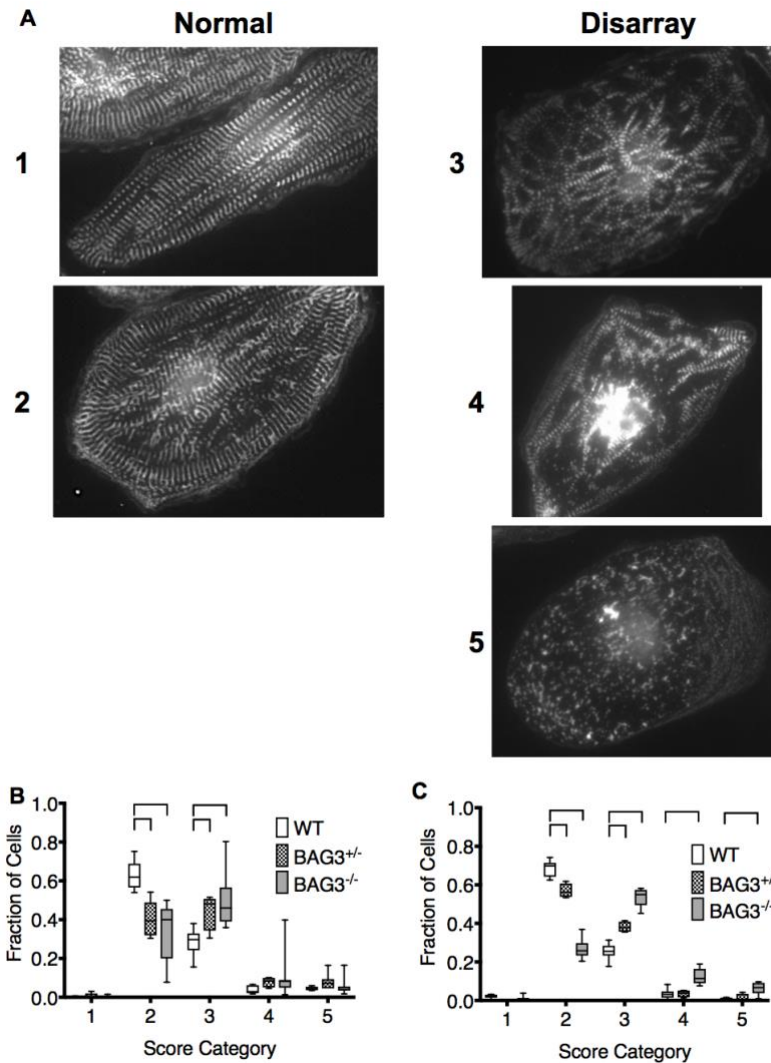
**Supplementary Figure 3.2** - Genotyping iPSC lines with targeted mutations. (A) PCR assays were designed to detect the integration of the knockout vector at the BAG3 and MYBPC3 loci, and the C-terminal FLAG tag at BAG3 locus. Primers were designed to detect the integration junction of the transgene and the endogenous locus. A separate primer pair was used to detect the wild type allele. pC represents a positive control plasmid for integration of the BAG3-knockout vector (KO1). (B-C) Both alleles from heterozygous and homozygous KO2 lines were amplified by PCR and cloned into TOPO-TA vector and sequenced. Both wild type and mutant alleles are shown from BAG3<sup>+/-</sup> KO2 line demonstrating 17-bp deletion leading to a frameshift with 18 aberrant amino acids followed by a stop codon in the mutant allele. Alleles from BAG3<sup>-/-</sup> KO2 were indistinguishable, with homozygous duplication of 13 bp leading to a frameshift with 16 aberrant amino acids followed by a stop codon. (D) Copy number assay for 3xFLAG sequence relative to RPP30 reference allele, using gDNA from the final iPSC BAG3-3xFLAG clones. (E) Copy number assay for puromycin resistance sequence relative to RPP30 reference allele, using gDNA from the final iPSC BAG3-3xFLAG clones.



**Supplementary Figure 3.3** - Karyotype of mutant iPSC lines. All isogenic derived lines demonstrated normal 46 X,Y karyotype consistent with parental line.

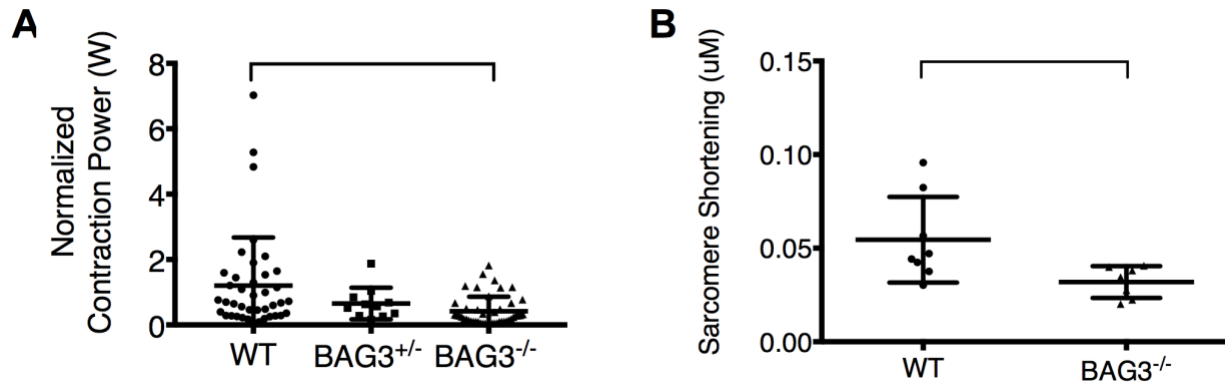


**Supplementary Figure 3.4 - Mutant iPSC lines express pluripotency factors and efficiently differentiate into cardiomyocytes.** (A) Immunofluorescent staining of iPSC colonies with antibodies to OCT4, SOX2, SSEA-4, and TRA1-81 in WT, BAG3<sup>-/-</sup> (KO1), and MYBPC3<sup>+/-</sup> lines. (B) iPS-CMs were collected at day 15 of differentiation (before lactate purification) and assessed by flow cytometry with antibody staining for cardiac troponin-T (TNNT). Individual values graphed with mean and s.d. (C) Representative flow cytometry plots for troponin-T from a typical BAG3<sup>-/-</sup> (KO1) iPS-CM differentiation along with a negative control sample from a failed differentiation. The percentage of cells in each quadrant is indicated.

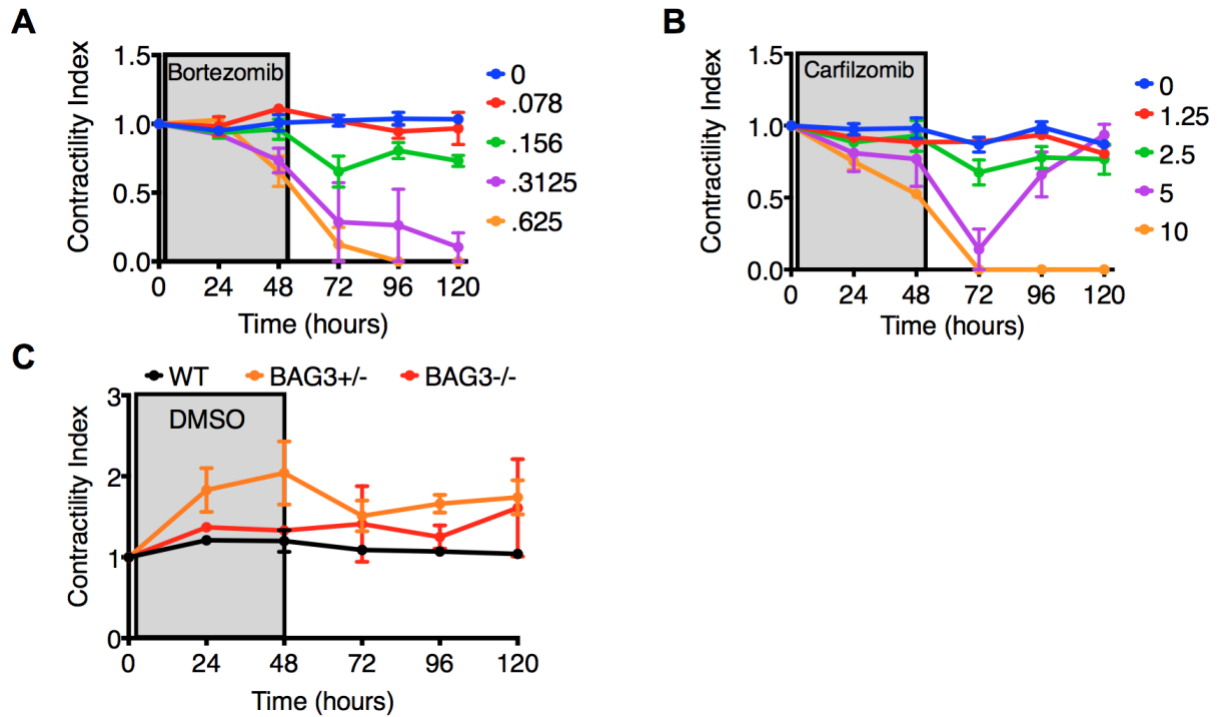


**Supplementary Figure 3.5** - Quantification of myofibrillar disarray in BAG3-mutant cardiomyocytes. **(A)** Example images for each category used to score myofibrillar integrity. Class 1 and 2 represented uniformly intact myofibrils with the distinguishing factor being that most myofibrils were aligned in parallel for class 1 (rarely seen). Class 3, 4, and 5 cells represent myofibrillar fragmentation, disintegration, or aggregation in a progressively increasing proportion of cell volume (3 < 50%, 4 ≥ 50%, 5 = no visible myofibrils). **(B,C)** Distribution of scores in wild type and BAG3-mutant cell lines from two different targeting strategies, **(B)** KO1, **(C)** KO2. Plotted as mean of biological replicates with box representing interquartile range and whiskers min-max. Brackets indicate significant difference from wild type by two-way ANOVA with Bonferroni's test for multiple comparisons,  $p < 0.05$ .

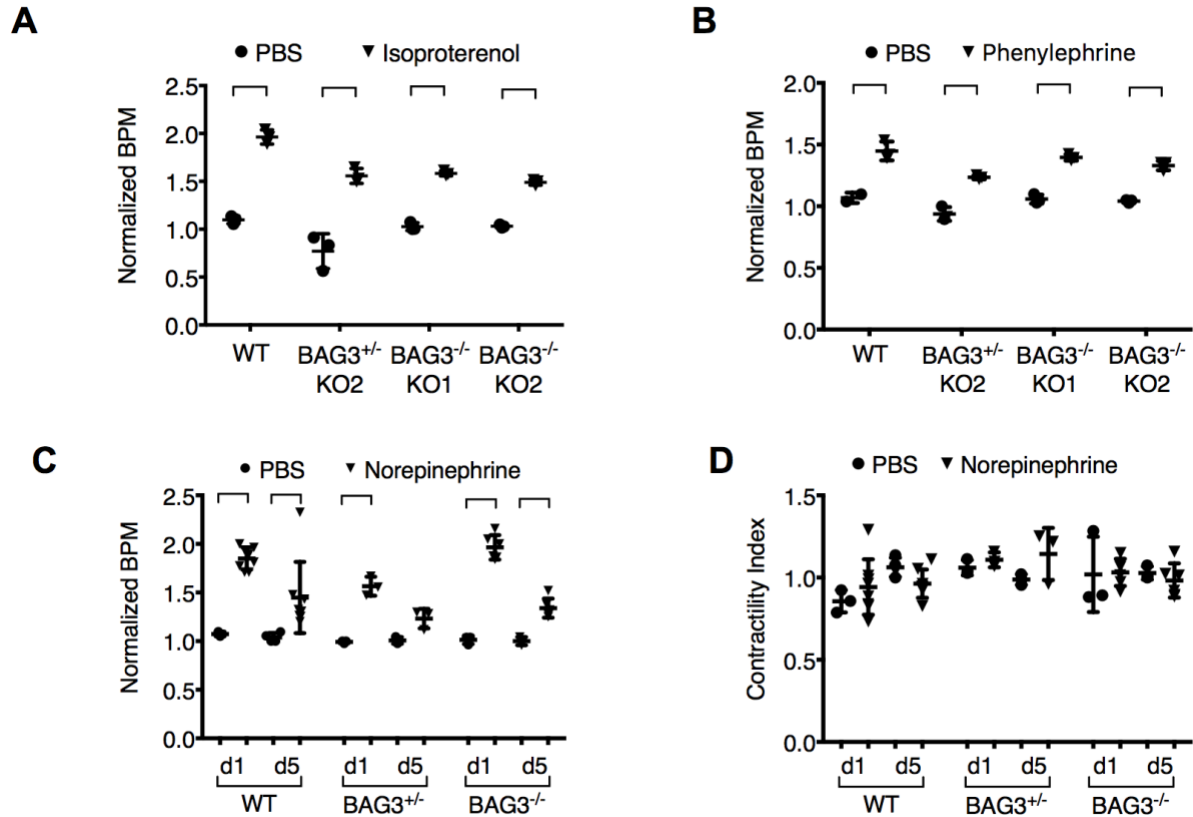




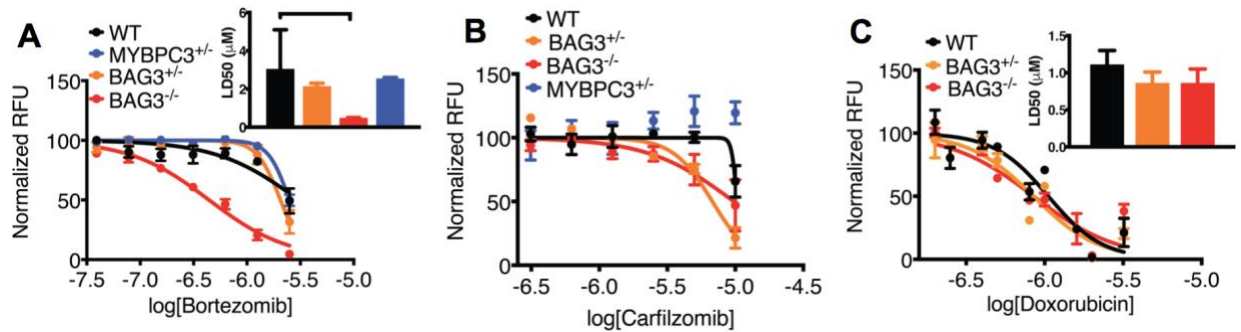
**Supplementary Figure 3.6** - BAG3 mutations produce contractile deficits in iPS-CMs cultured on micro-patterned substrates with increased stiffness. Day >30 purified iPS-CMs were cultured on micro-patterned polyacrylamide hydrogel substrates with a mechanical stiffness of 35 kPa. BAG3 KO1 mutant lines were used. (A) Contraction power was calculated from the measured force and contraction velocity determined by traction force microscopy from the movement of fluorescent beads in the substrate. Results were normalized to wild type and individual replicates plotted with mean and s.d. Brackets indicate significant difference by one-way ANOVA with Bonferroni's test for multiple comparisons,  $p < 0.001$ . (B) Sarcomere shortening was measured in Lifeact-labeled myofibrils. Results were normalized to wild type and individual replicates plotted with mean and s.d. Brackets indicate significant difference by two-tailed students t-test,  $p < 0.05$ . Measurements were obtained from three independent device cultures, with 11-47 cells analyzed per line for force measurement and 7-8 cells analyzed per line for sarcomere shortening measurement.



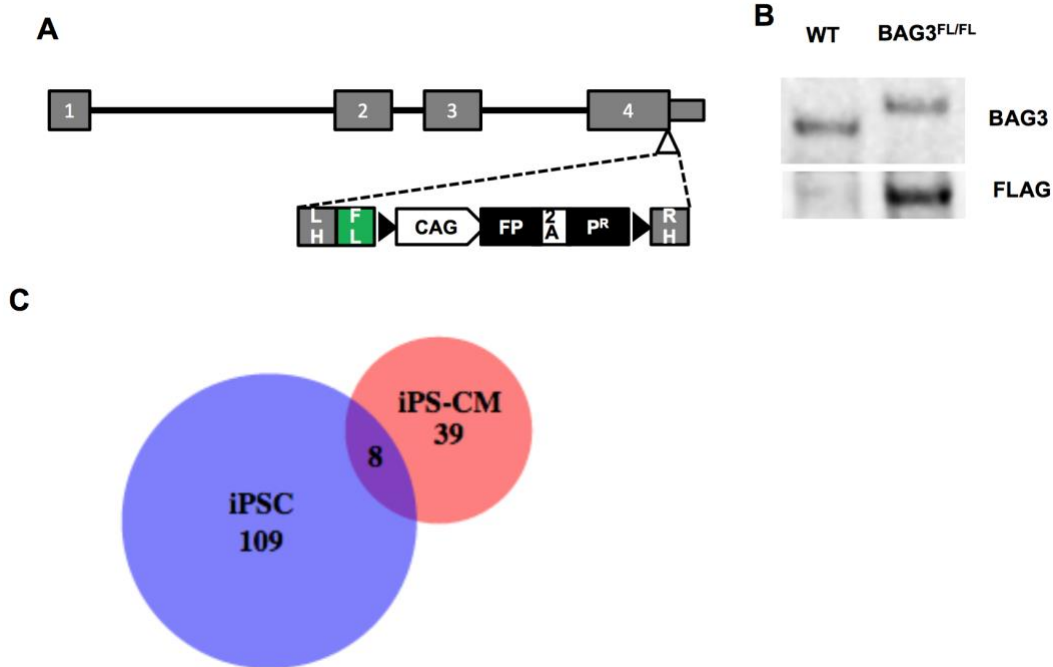
**Supplementary Figure 3.7** - Effect of proteasome inhibitors on WT iPS-CM contractility. (A–B) Contractility was measured before and after varying doses of bortezomib or carfilzomib using the Cellogly Pulse system. Contractility index represents the contraction peak height at each time point normalized to the baseline value for each well. Measurements were obtained every 24 h for 5 days, with cells exposed to drug for the first 48 h. Shown are mean and s.e.m. of triplicate wells. (C) Contractility was measured for wild type and BAG3 mutant (KO1) iPS-CM during and after exposure to DMSO vehicle control.



**Supplementary Figure 3.8** - BAG3 mutant iPS-CMs have a positive chronotropic response to adrenergic stimulation without loss of contractility. (A–D) Lactate-purified > day 30 iPS-CMs were treated with PBS, 10  $\mu$ M norepinephrine, 1  $\mu$ M isoproterenol, or 10  $\mu$ M phenylephrine. Contractile motion was measured on a Cellogy Pulse system at baseline and 1 h after addition of drug. Norepinephrine was dosed daily for 5 consecutive days; results from days 1 and 5 are shown. Individual replicates are graphed (3-8) with mean and s.d. of normalized beat rate or contractility index. Brackets indicate significant difference from PBS control by two-way ANOVA with Bonferroni's test for multiple comparisons,  $p < 0.05$ .



**Supplementary Figure 3.9** - Dose-response assay for cardiomyocyte viability after chemotherapy drug exposure. Wild type, BAG3-mutant (KO2), and MYPBC3 mutant day >30 iPS-CM were treated with **(A)** bortezomib, **(B)** carfilzomib, or **(C)** doxorubicin at the doses indicated for 48 hours. Cells were allowed to recover for 5 days in basal media, after which viability was measured with a PrestoBlue resazurin assay. Relative fluorescent units were normalized to wells treated with vehicle only. Mean and s.e.m. are plotted for triplicate samples at each dose. Inset graphs represent the calculated LD50 and 95% confidence interval from each corresponding dose response curve using non-linear regression analysis. LD50 values were unable to be calculated for wild type and MYPBC3<sup>+/-</sup> lines at the doses of carfilzomib tested. Brackets indicate significant differences by one-way ANOVA with Bonferroni's test for multiple comparisons,  $p < 0.05$ .



**Supplementary Figure 3.10** - Genome engineering a 3X-FLAG affinity tag in the endogenous BAG3 locus of WTc human iPSC. (A) Schematic of the BAG3 gene with four exons in the predominant coding isoform. CRISPR-Cas9 and gRNA was designed to target near the BAG3 stop codon (open triangle). The targeting vector included flanking left and right homology arms (LH and RH, respectively) with the 3x-FLAG tag (FL) followed by selection cassette flanked by loxP sites (black triangles). The selection cassette contained mCherry (FP) and puromycin resistance genes (P<sup>R</sup>) driven by CAG promoter. (B) Western blot for BAG3 protein in iPS-CMs, introduction of 3xFLAG tag was visualized by a larger size band and reactivity to anti-FLAG antibody. (C) Venn diagram demonstrating the overlap in interactors identified from iPSC versus iPS-CM.

**Supplementary Table 3.1** – Genotyping primers used in this study.

BAG3KO 5' forward	TGTTAGGACTTCTGTTGCGC
BAG3KO 5' reverse	CTAGGCACCGGTTCAATTGC
BAG3KO 3' forward	TAGTTGCCAGCCATCTGTTG
BAG3KO 3' reverse	ATGTGTTGTCCCCACTACCC
BAG3 exon 2 forward	CAGGAGACTCCATCCTCTGC
BAG3 exon 2 reverse	CTGGCCTCTCCTTACCTCAG
BAG3FLAG 5' forward	TGCCATTAAGAATACCATCTACAGAG
BAG3FLAG5 5' reverse	GTGGTCCTTGTAGTCGCC
BAG3FLAG 3' forward	CTACAAGGACGACGACGA
BAG3FLAG 3' reverse	CACCACCTGGCTGATTTG
BAG3 exon 4 forward	GTCTATGAACTCCAGCCCAG
BAG3 exon 4 reverse	CTACAAAAGACAGTGCACAAC
MYPBC3KO 5' forward	TACGTGTGACAATCCTGTGC
MYPBC3KO 5' reverse'	GGGCGTACTTGGCATATGAT
MYPBC3KO 3' forward	TGCATGACCCGCAAGCCCGG
MYPBC3KO 3' reverse	CCTTTCATTCCTCAGAATGG
Primers and Probes for ddPCR copy number assays	
FLAG forward	GGTCAAGTCCAGGTCTATGAAC
FLAG reverse	TTGTCGTCGTCGTCCTT
FLAG probe	ATCCGACTACAAGGAC
Puro forward	GTCACCGAGCTGCAAGAA
Puro reverse	CACCTTGCCGATGTCGAG
Puro probe	CTCTTCCTCACGCGCGTCGG

# Chapter 4 - genome engineering of the endogenous copy of the BAG3 gene on induced pluripotent stem cells

## 4.1 Introduction

The advent of site-specific nucleases has revolutionized our ability to modify mammalian cells. The combination of these advances in the genome engineering with the induced pluripotent (iPS) cell-field allows for the creation of better models to study human genetic disease. In this chapter, I will describe the use genome engineering technologies (CRISPR/Cas9 and TALEN) on iPS cells to introduce specific single nucleotide variants and a fusion to the 3xFLAG affinity purification tag in the endogenous copy of the BAG3 gene. The cell lines produced in this chapter then enabled me to study the differences in the interactome of different variants of the BAG3 protein (see Chapters 5 and 6) expressed on an isogenic background (the WTc cell line). The insertion of the 3xFLAG cassette allowed us to perform affinity purification on the BAG3 protein expressed at physiological levels of expression, while the single nucleotide variants inserted allowed us to study the impact that certain disease-related gene variants have in the interaction profile of BAG3 and in the whole cell homeostasis.

## 4.2 Materials and Methods

### 4.2.1 iPS cell culture

For this study, the WTc iPS cell line was used. The WTc line was generated using an episomal reprogramming method<sup>160</sup> on fibroblasts from a healthy male donor. The subject has a normal ECG and no evidence of cardiac disease. The rest of the cell lines used in this study were derived from the WTc line. For routine maintenance, cells were cultured on Growth Factor Reduced Matrigel (8ug/ml, BD Biosciences) and fed with mTesr1 medium (STEMCELL Technologies) every day. Whenever the cells reached 70-90% confluence, they were passaged using Accutase Cell Detachment Solution (STEMCELL Technologies), and seeding them on media supplemented with the ROCK1 inhibitor Y-27632 (10uM, Selleckchem) for 24 hours.

### 4.2.2 TALEN and CRISPR targeting sequence design and cloning

ZiFiT (<http://zifit.partners.org/ZiFiT/ChoiceMenu.aspx>) was used to design TALENs. TALENs were constructed using the Voytas laboratory Golden Gate assembly system and backbone vector MR015, as previously described<sup>121</sup>. All TALENs used the NN repeat variable di-residues (RVDs).

For designing CRISPR guide RNA sequences, we used the CRISPR DESIGN website ([crispr.mit.edu](http://crispr.mit.edu)). A window size of 250bp around the intended modification site was used, and three candidate gRNAs were selected per loci. Final RNA sequences were chosen by their ability to produce homologous recombination. The guide sequence was cloned into a guide RNA (gRNA) expression vector using the protocol described by the Zhang lab CRISPR



resources on Addgene ([www.addgene.org/crispr/zhang](http://www.addgene.org/crispr/zhang)) (53). pX330-U6-Chimeric\_BB\_CBh-hSpCas9 (Addgene plasmid 42230) was a gift from Feng Zhang (Broad Institute, Cambridge, Massachusetts, USA).

Final TALEN and CRISPR target sequences used in this study can be found in Table 4.1.

### **4.2.3 Single nucleotide genome editing, sib-selection**

For the generation of iPS cell lines bearing single nucleotide changes in the BAG3 gene, we used a method described by Miyaoka et al<sup>121</sup>. For each cell line to be generated, 2 million WTc iPS cells were nucleofected with a pair of pX335 expressing the TALEN proteins (3ug each plasmid) and a 60bp oligo donor DNA (6ug). Donor DNA oligos were designed to insert to target the point mutation in the middle of the sequence, and purchased from Integrated DNA Technologies. For nucleofection, we used the Human Stem Cell Nucleofactor Kit-1 and a Nucleofactor 2b Device (both by Lonza Group) on the setting A-023. We split the nucleofected cells in multiple wells of a 96 well plate. Once they reach passaging confluence we harvested them, freezing half and extracting DNA from the other half (see ‘on-plate freezing, thawing and genomic DNA extraction of iPS cell populations’ below). After genotyping the polyclonal cell populations (see ‘droplet digital PCR for detection and quantification of genome editing events’ below), we thawed the wells that presented higher editing, counted the cells and seeded them at low confluence (500 cells per well) in multiple wells of a Matrigel-coated 96-well plate.

This process (sib-selection) was repeated 2-3 times, selecting the populations that contained higher genome editing. Then single colonies were cloned and sequenced (see ‘iPS cell cloning, genotyping and sequencing’ below). We performed an extra round of cloning for each cell line to ensure homogeneity of genotype.

#### **4.2.4 Generating the TetOn-BAG3<sup>3xFLAG</sup>**

For the insertion of the cassette in the AAVS1 locus, we used the TALEN pair described by Hockemeyer et al<sup>171</sup>. Briefly, WTc iPS cells were nucleofected using an Amaxa nucleofector 2B and the Nucleofector Kit C (both from Lonza). We used 2 million cells, and 0.5ug of each AAVS1 TALEN pair (Table 4.3) and 1ug of the donor plasmid (see Figure 4.2). Positive cells were then selected on 1mg/ml G418, before picking single clonal populations (see 4.2.8). This was done for cassettes expressing the cDNAs for BAG3<sup>3xFLAG</sup>, BAG3<sup>P209L-3xFLAG</sup>, and BAG3<sup>E455K-3xFLAG</sup>.

#### **4.2.5 Differentiation of iPS into iPS-CM**

We used a protocol primarily based on the modulation of the Wnt pathway<sup>162</sup>, as described in Chapter 3. All iPS-CM used in the experiments described here were previously purified using a lactate metabolic switch protocol, as also described on Chapter 3.

#### **4.2.6 Insertion of the 3xFLAG sequence in the endogenous BAG3 locus**

WTc iPS cells were nucleofected using a 4D-Nucleofector system and Primary Cell Solution P3 (Lonza). We simultaneously nucleofected 6ug of a donor plasmid (expressing the mCherry fluorophore and Puromycin resistance) and the pX330 plasmid expressing the Cas9-3xNLS protein and a guide RNA. Two days after nucleofection, cells were treated with

mTesr1 supplemented with 10ug/ml of ROCK1 inhibitor and 0.5ug/ml Puromycin (Sigma-Aldrich). Media was exchanged daily until stable mCherry-positive colonies remained (5–7 days). After PCR genotyping to confirm successful editing in this polyclonal population, the cells were nucleofected with a plasmid expressing Cre recombinase. Cells were directly seeded on a 10cm dish and mCherry-negative clones were isolated and genotyped by PCR (as described in ‘iPS cell colony cloning, genotyping and sequencing’ below).

For the generation of the BAG3<sup>E455K-3xFLAG/wt</sup> cell line, we introduced the corresponding single nucleotide mutation in the 3xFLAG donor plasmid using QuikChange Site-Directed Mutagenesis Kit (Agilent). Then this plasmid was used for the process described above.

#### **4.2.7 On-plate freezing, thawing and genomic DNA extraction of iPS cell populations**

To freeze cells on 96 well plates, the wells were added 75ul of freezing media (90% Fetal Bovine Serum (Hyclone; Thermo Scientific), 10% Dimethyl sulfoxide (Sigma)) and covered with 75ul of mineral oil. Plates were sealed with parafilm and stored in a styrofoam box at -80C until thawing.

For DNA extraction from 96 well plates, the cell solution was added to 50ul of lysis buffer (10mM Tris pH7.5, 10mM EDTA, 10mM NaCl, 0.5% N-lauroylsarcosine, supplemented with 1mg/ml of proteinase K). Plates were then incubated overnight at 55C in a sealed, moist environment. The morning after we added 100ul/500ul of ice-cold precipitation solution (75mM NaCl in ethanol) and let sit for 2 hours. We then washed the precipitated DNA with 70% ethanol and resuspended in water.

For thawing, we put the plates in a mammalian cell incubator for approximately 10 minutes. Then we transferred the selected wells into a tube containing 0.5mL of mTsr1 media with ROCK1 inhibitor and seeded them. For 24 well plates, the process followed was the same but using 5x the volume.

#### **4.2.8 iPS cell colony cloning**

To isolate homogeneous populations of cells, we seeded polyclonal populations in Matrigel-coated 10cm dishes at a density of 20,000 cells per plate. Cells were grown in mTsr1 supplemented with ROCK1 inhibitor for 1-2 days, and then in mTsr1 only for 3-4 more days, until they formed separate round colonies. We then picked colonies put them on separate wells of a 24 well plate, letting them grow on mTsr1 supplemented with ROCK1 inhibitor for 5-7 days. Then they were split into two separate 24 well plates, freezing one and using the other for genomic DNA extraction (see 'on-plate freezing, thawing and genomic DNA extraction of iPS cell populations' above).

We performed an additional round of cloning for all the cell lines generated, followed by FLAG and Puromycin copy number variation analysis of the clones to ensure a homogeneous population was obtained.

#### **4.2.9 Cell genotyping**

We performed a droplet digital PCR (ddPCR) assay for quantification of the allelic abundance of the modifications introduced in the cells by genome editing. For each reaction of ddPCR, 50 or 100ng of genomic DNA was mixed with 5uM of each one of two TaqMan MGB detection probe (one with the FAM fluorophore and another one with VIC or HEX),

18uM of forward and reverse amplification primer, and 1x of ddPCR Supermix for Probes (Bio-Rad). We then generated water droplet in oil using a QX100 Droplet Generator (Bio-Rad) and transferred the emulsion into a 96 well plate for thermal cycling. Amplifications were performed on a C1000 Thermal Cycler (Bio-Rad), using the following settings: step 1, 95C for 10 minutes; step 2, 94C for 30 seconds; step 3, annealing temperature (optimized from previous temperature gradient amplification) for 30 seconds; repeat steps 2-3 39 times; then step 4, 98C 10 minutes. We then analyzed the droplet intensity using a QX100 Droplet Reader (Bio-Rad) in the “absolute quantification” setting.

For single nucleotide variant allelic abundance estimation, we designed the primers and probes using the TaqMan MGB Allelic Discrimination option in Primer Express 3.0 software (Life Technologies). Probe pairs were designed to discriminate between introduced and original sequence. A ratio of modified vs original positive droplets was used as a readout.

For copy number variation analysis of the 3xFLAG and puromycin cassettes, we performed a copy number variation analysis. The ratio of positive droplets to a reference gene was used as a readout. The reference gene was RPP30 (PrimePCR™ Probe Assay, BioRad).

For sequencing validation of genotypes, target genomic regions were amplified by PCR with BioMix Red (Bioline) and sequenced. For the 3xFLAG sequence insertion, we ran the PCR product on a gel to separate allele products and sequenced them separately. We also sequenced the homology arm boundaries to ensure no modifications had been introduced.

The sequences for each one of the genotyping reactions are described in Table 4.1.

#### 4.2.10 Karyotyping

All karyotyping of the cells was performed by Cell Line Genetics LLC (Madison, Wisconsin). Only cells with apparently normal karyotype were used in this study.

#### 4.2.11 Phasing assay

For allelic phasing, we performed a ddPCR-based assay combining a TaqMan probe detecting the 3xFLAG cassette and a probe binding the single nucleotide variant we wish to phase. Droplet digital PCR was performed as described above, and the linkage percentage for variants was computed as described by Regan et al<sup>172</sup>. Briefly, If we assume independence between events, then:

$$(N_A * N_{AB})x(N_b * N_{AB}) = N_{ab}$$

$$\frac{(N_A * N_B)}{N_{AB}} = 1 - (N_A + N_B + N_{AB}) = N_{empty}$$

$$(N_E * N_{AB}) = N_A * N_B$$

Where  $N_A$  and  $N_B$  are the number of droplets positive for event A and B, respectively, while  $N_{AB}$  are double-positive droplets due to chance and  $N_E$  are empty droplets. Then, if we consider the case where the two events are physically linked denoted as  $N_{\underline{AB}}$ , and using the Poisson statistics equation for where  $\lambda_x = -\ln(\lambda_{not x})$  (see reference, we obtain:

$$\lambda_{\underline{AB}} = -\ln\left(\frac{N_A + N_B + N_E + \frac{N_A * N_B}{N_E}}{N_{all}}\right)$$

$$\lambda_A = -\ln\left(\frac{N_{not A}}{N_{all}}\right); \text{ and } \lambda_B = -\ln\left(\frac{N_{not B}}{N_{all}}\right)$$

Then, we calculate the percent of linked molecules as:

$$\% \overline{AB} = \frac{\lambda_{\overline{AB}}}{\frac{(\lambda_A + \lambda_B)}{2}} * 100$$

#### 4.2.12 Western Blot

Cells were lysed by adding RIPA buffer to the plate, and incubating on ice for 30 minutes. Then lysates were harvested, clarified by centrifugation for 10 minutes and quantified using a microBCA protein assay kit (Thermo Scientific #23235). A total of 20-30ug total protein extracts were loaded per lane. Samples were prepared in NuPAGE LDS Sample Buffer and NuPAGE Sample Reducing Agent (Thermo Fisher Scientific) following directions by provider. Samples were loaded on NOVEX 4-12% Bis-Tris Protein Gels (1.5mm thick; Thermo Fisher Scientific) and electrophoretic separation was performed on an XCell SureLock mini-cell (Thermo Fisher Scientific) at 150V on NuPAGE MES SDS buffer (Thermo Fisher Scientific). Protein was transferred into PVDF membrane using the using a wet-dry transfer system at 200mA for 2 hours, and blocked with Odyssey Blocking Buffer (PBS) (LiCor) for 1 hour at room temperature. Then the membrane was incubated overnight at 4C with a primary antibody in PBS supplemented with 0.1% Tween (PBS-Tween). The day after, primary antibody was washed three times with PBS-Tween and membranes were incubated with secondary antibodies for 1 hour at room temperature. Membranes were then imaged using an Odyssey FC Imager (Li-Cor), and bands were analyzed and quantified using the ImageStudio Software (Li-Cor). Secondary antibodies were Goat Anti-Mouse IRDye

680LT (dilution 1:20000) and Donkey Anti-Rabbit IgG IRDye (dilution 1:10000) (both by Li-Cor). Primary antibodies and their dilutions used were: Anti-BAG3 rabbit polyclonal (1:2000, ProteinTech, 10599-9-Ap), Monoclonal ANTI-FLAG M2 (1:1000, Sigma-Aldrich, F1804), polyclonal Anti-GAPDH rabbit (1:2000, Abcam, ab9458).

#### **4.2.13 Immunofluorescence staining**

For immunostaining of cells in culture, cells were washed with PBS and fixed with 4% Paraformaldehyde at room temperature for 20 minutes. Then cells were washed three times with PBS supplemented with 0.1% Triton X-100 (PBS-Triton) and incubated with 5% Bovine Albumin Serum in PBS-Triton for 1 hour at room temperature to block and permeabilize. Then cells were incubated overnight at 4C with a solution of primary antibody in PBS-Triton. The morning after, cells were washed three times with PBS-Triton and then incubated in secondary antibody in PBS-Triton for 1 hour at room temperature. Secondary antibodies used were Goat Anti-Rabbit Alexa Fluor 594 and Goat Anti-Mouse IgG Alexa Fluor 488 (both at a 1:500 dilution; both from Thermo Fisher Scientific). For primary antibodies, we used rabbit anti-BAG3 (Protein Tech, 10599-1-A), Monoclonal ANTI-FLAG M2 (1:1000, Sigma-Aldrich, F1804). Cells were imaged using a BZ-9000 microscope (Keyence).



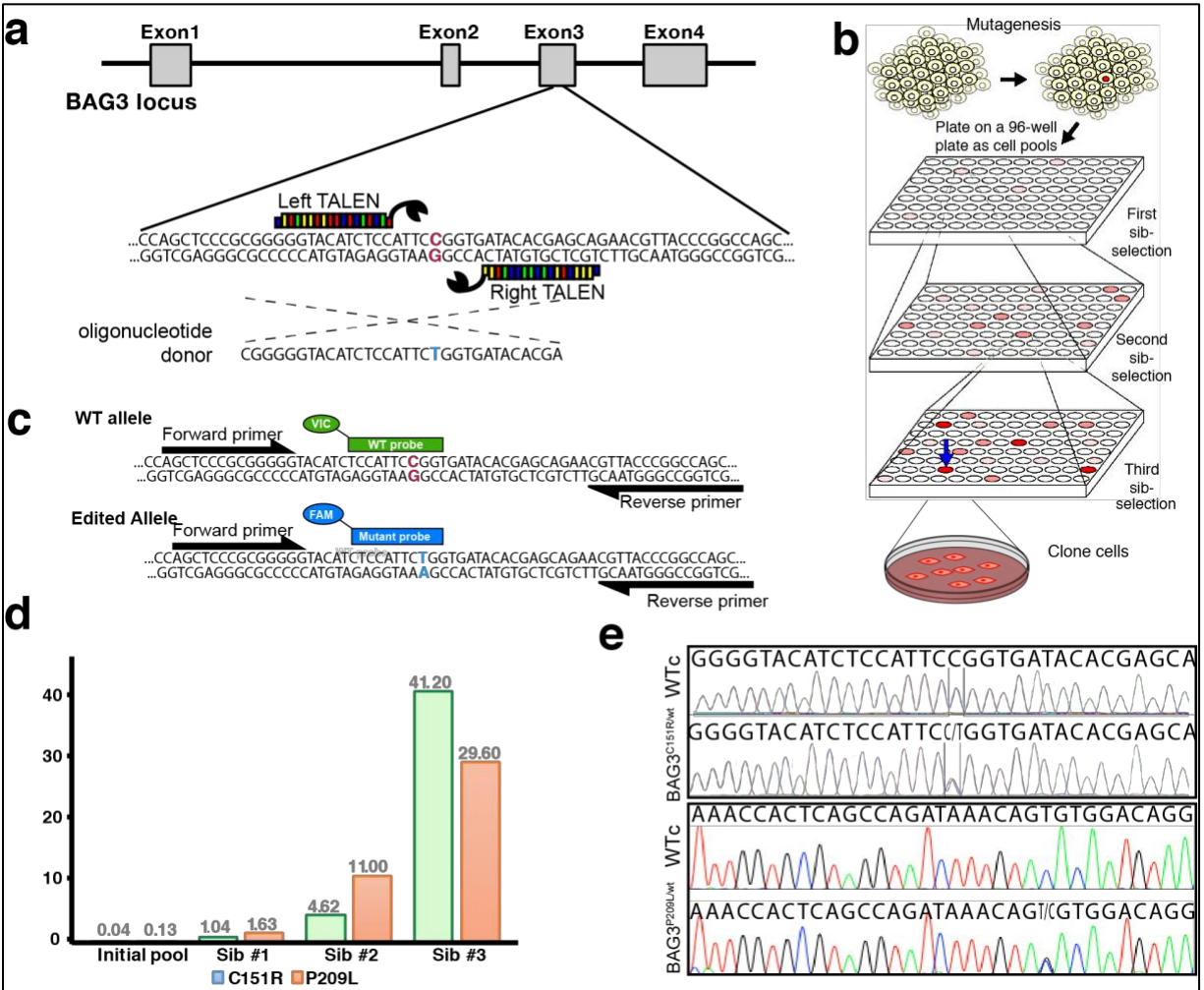
## 4.3 Results and discussion

### 4.3.1 Droplet digital PCR enables identification of low frequency genome editing events and enables scarless single nucleotide genome editing

Many methods aiming to introduce modifications in the cell genome depend on the use of some sort of genetically encoded selection system to enrich the cells bearing the modification. Albeit effective, these methods are mostly unsuitable for the introduction of single nucleotide changes in exonic regions, as the selection system would inevitably co-segregate with the modification and be left as a genetic ‘scar’, potentially influencing the gene function. An alternative is to clone cells directly after introduction of the nuclease enzyme. Unfortunately, human pluripotent stem cells present very low homology directed repair efficiencies (as reviewed in <sup>173</sup>), making this process exceedingly laborious. In addition, increasing the levels of nuclease to improve efficiencies could result in decreased targeting specificity<sup>174</sup>.

One way to detect and enrich for low frequency editing events involves the use of droplet digital PCR (ddPCR) using a set of fluorescently labeled probes that recognize the specific editing event. Since each droplet contains no more than a single molecule of DNA on average, we can get accurate quantification of the amount of alleles that contain the editing in a sample of genomic DNA from a mixed population<sup>175</sup>. Combining this highly sensitive and quantitative method with a sib-selection process (where a heterogeneous population is subdivided and then the subpopulation containing the higher proportion of desired events is further subdivided for another round of quantification) allows us to enrich for a low frequency

genotype until we can clone to obtain a pure population (Fig 4.1). This process has been used successfully in our lab to generate multiple cell lines<sup>121</sup> and was used in this study to generate the cell lines BAG3<sup>P209L/wt</sup> and BAG3<sup>C151R/wt</sup> (Fig 4.1). We used TALEN pairs designed to induce homologous recombination of an exogenous cassette bearing the mutation in the desired sites of BAG3 gene. The cell lines obtained from the selection process contained the desired nucleotide mutations (see Table 4.2) at a heterozygous allelic frequency. We also sequenced the genomic region surrounding the modified nucleotide to ensure no unintended ‘scar’ had been left in the genome. Karyotype of the cell lines was normal (see Fig 4.5) and they were able to differentiate effectively into cardiomyocytes (see chapter 6).



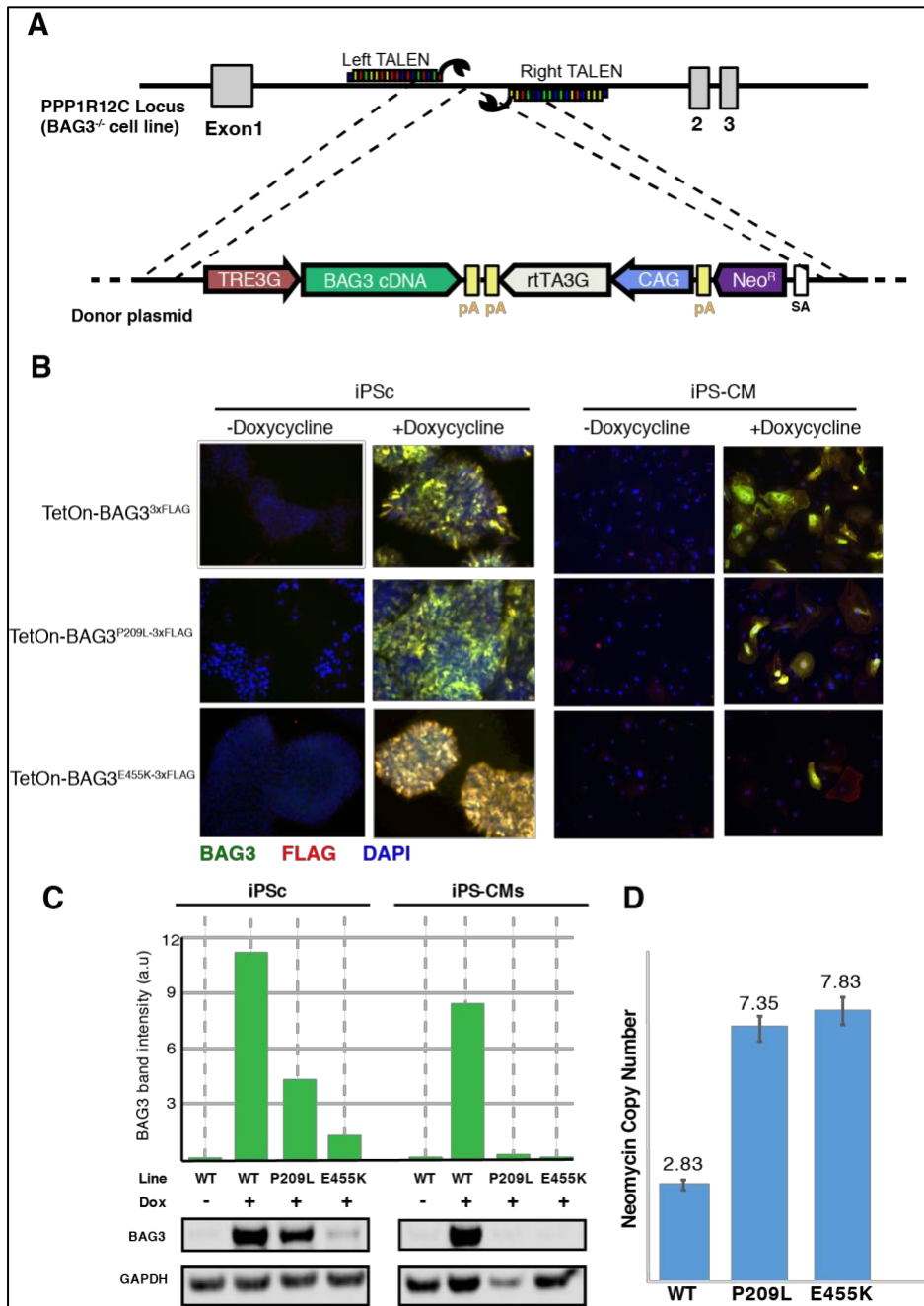
**Figure 4.1** - Scarless introduction of single nucleotide polymorphisms in the BAG3 gene. A) Using a pair of TALEN nucleases we induce a double strand break in the BAG3 gene. A supplied nucleotide donor bearing the modification of interest will be used by the cell for homologous recombination repair. Showing example for P209L variant. Due to very low efficiency of homologous recombination, we performed a ‘sib-selection’ approach (B) to enrich for the cells that bear the modification of interest. At every step, ddPCR (see text) is used to screen for the percentage of correctly edited cells in every well. Two TaqMan probes that can discriminate between single amino acids are used (C). D) Enrichment of edited allelic frequencies, from the initial pool (right after nucleofection of the nucleases and donor) and through the 3 sib-selection steps. After sib step 3, colonies were picked manually. E) Sequencing results from final cell lines used. ddPCR was used to validate that allelic frequency was 50% (not shown).

### 4.3.2 Transgenes inserted in the AAVS1 safe harbor locus express unevenly across cells

Next, we proceed to insert the nucleotide sequence encoding the 3xFLAG affinity tag in the endogenous BAG3 gene, so the product would be a BAG3-3xFLAG fusion protein. Using the ‘scarless’ method previously described proved very difficult, probably due to low efficiency of insertion of long oligonucleotide sequences in this specific site of the iPS cell genome (data not shown). An alternative to editing the endogenous copy of target gene is to introduce a cassette expressing the gene cDNA fused to the affinity purification tag at a ‘safe harbor’ location in the genome, where it does not affect the expression of other genes<sup>171</sup>.

We adopted this strategy and generated three lines containing a cassette expressing 3xFLAG-tagged wild-type BAG3 and two variants (P209L and E455K) under the doxycycline-inducible promoter TetO. The cells lines were produced by the use of a TALEN assay (as in <sup>171</sup>) and successful editing events were selected by antibiotic selection (see Figure 4.2). In spite of cloning single colonies two times sequentially, we found that only a few of the cells expressed BAG3-3xFLAG upon addition of doxycycline (see Figure 4.2). This effect was more pronounced upon differentiation into cardiomyocytes and uneven across the cell lines expressing different variants (Figure 4.2). This inefficient activation of the transgene is likely due to silencing of the inserted cassette. It is not clear whether this silencing is intrinsic of the AAVS1 locus in iPS cells, or due to the specific sequence being inserted there, but silencing of the TetO promoter during differentiation has been reported previously<sup>176</sup>. We were also surprised to observe that that the copy number analysis for the Neomycin selection cassette

suggested multiple off-target insertion events, and that these did not correlate with the observed expression levels (Figure 4.2).



**Figure 4.2** - BAG3<sup>3xFLAG</sup> cassette inserted in the AAVS1 safe harbor gets silenced in iPS-CM. A) Design of construct and strategy. B) Most iPS cells express BAG3 upon induction with doxy-cycline. However, after differentiation into iPS-CM many of them lose expression. (C) Western blot data. In (D), copy number variation of the neomycin selection reveals that the cells got multiple insertions. Interestingly, the BAG3-3xFLAG line contained the least copies detected, and yet was the one that preserved expression the most through differentiation.

The observed inefficient activation of the transgene would increase our cell requirements for future experiments, and make phenotype analysis complicated. In addition,

an uneven expression of the bait protein across cell lines would make comparing results between them very difficult. Because of these reasons, we decided not to use these lines for the differential analysis of the interactome of BAG3 variants. The line expressing the BAG3<sup>wt-3xFLAG</sup> variant will be used as a tool to study the effect of protein overexpression on protein interaction profiles (see Chapter 6).

### **4.3.3 Genome editing tools allow for the insertion of long nucleotide sequences in target sites in the genome and the expression of fusion proteins at physiological levels**

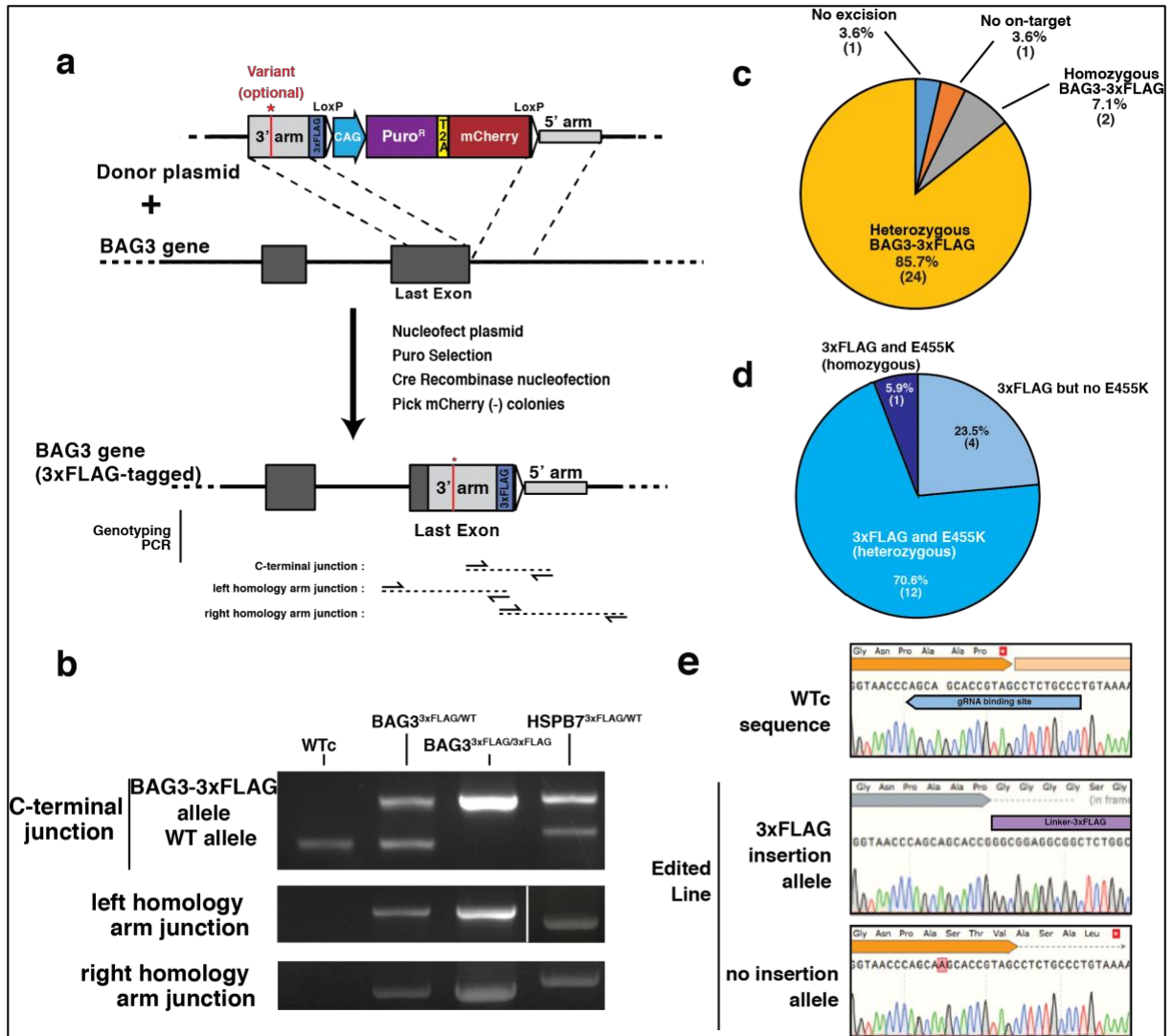
We generated cell lines that bear a nucleotide sequence encoding for the 3xFLAG affinity purification peptide sequence<sup>177</sup> in the 5' end of the endogenous BAG3 gene, right in between the last amino acid and the stop codon.

For the insertion of N- or C- terminal protein fusions, the introduction of a genetic scar is less of an issue as this can go in the untranslated regions of the protein, and as long as it doesn't affect the expression of the protein. In addition, we had failed previously trying to isolate positive clones for 3xFLAG insertion using a scarless editing strategy like we had used for point mutations (data not shown). For these reasons, we decided to use a strategy where a selection cassette would co-segregate with the 3xFLAG, allowing for enrichment of recombination events. The selection cassette is surrounded by recombinase target site, and also contains a fluorophore. This allows for the excision of the whole cassette by the expression of a recombinase enzyme and cloning of single cells losing fluorophore expression, leaving behind only the desired insertion sequence and a small 46bp 'scar'.

The outlined process was used on WTc iPS cells to generate the BAG3<sup>3xFLAG/wt</sup> and BAG3<sup>3xFLAG/3xFLAG</sup> cell lines. It was also used in the BAG3<sup>C151R/wt</sup> and BAG3<sup>P209L/wt</sup> followed by a phasing method to select cells that contained the point mutations in *cis* with the 3xFLAG tag. We also produced a cell line bearing a 3xFLAG in the endogenous HSPB7 locus (HSPB7<sup>3xFLAG/wt</sup>), a gene associated with dilated cardiomyopathy and potentially an interactor of BAG3<sup>18,178</sup>. This served as a validation of this endogenous 3xFLAG tagging approach and also yielded a useful cell line for protein interaction studies.

In all the cases where we used this process, the majority of the colonies picked at the end contained the desired insertion in place, and around 5% of them were homozygous for the insertion (Figure 5.3). The homologous recombination seemed to happen independently in either BAG3 allele (Figure 5.3).

We also found that all the cells that contained successful insertion of the cassette in one of the alleles of the BAG3 gene also contained a single basepair insertion in the other, ‘unedited’ allele. This specific modification was found in all of 71 screened clones. Other studies have reported that CRISPR/Cas9 favors specific NHEJ events at some loci (see, for example, <sup>126</sup>), but this is to our knowledge the first report of a modification that happens in 100% of the cases. The insertion results in a 3 amino acid extension of the BAG3 protein, which is not expected to have major influence in the expression level or function of the protein per se (especially in light of the lack of effect observed for the much longer insertion, of the 3xFLAG peptide (Figure 4.3)).



**Figure 4.3** - Insertion of the 3xFLAG epitope tag fusion sequence in the endogenous BAG3 gene. A) Construct design and strategy. See text for details. Half arrows indicate genotyping primers. Specific variants can be inserted simultaneously to the 3xFLAG by adding them into the homology arms. B) Genotyping PCR show the shift in mass for the FLAG-tagged alleles. C) Summary of the genotyping results for the clones from inserting 3xFLAG in the WTc cell line. No excision: the clone still contained the selection cassette; No on-target: both alleles seemed intact, suggesting selection cassette went to another location in the genome. D) Genotyping summary for the co-segregation of the 3xFLAG and the mutation producing the BAG3<sup>E455K</sup> variant. E) Annotated sequencing results showing the 1bp insertion that was found in all the clones in the allele that did not get the 3xFLAG insertion. The insertion produces a frameshift that results in an extension of 4 amino acids at the end of the protein.



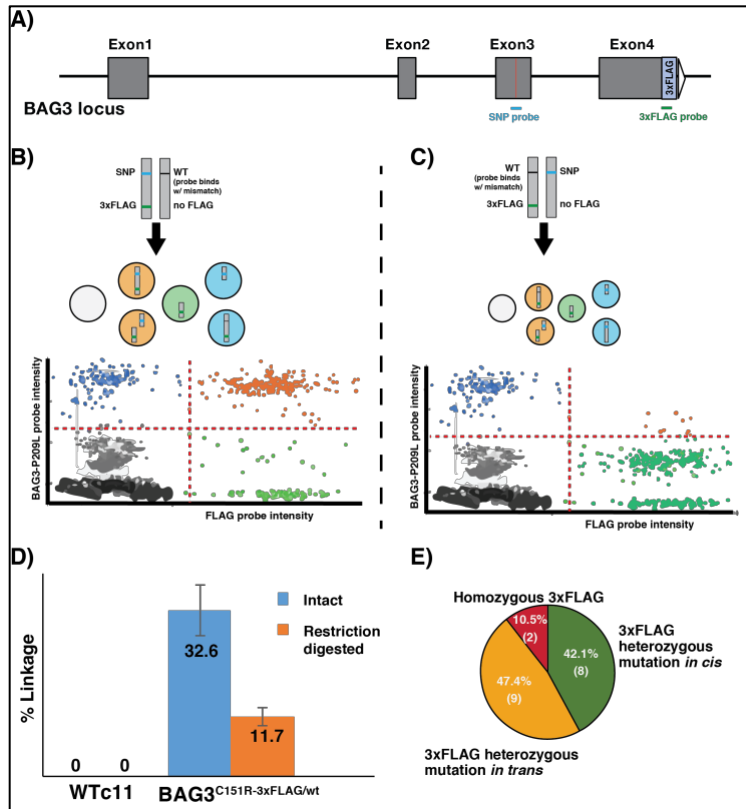
#### 4.3.4 Droplet digital PCR allows for the successful phasing of chromosomal variants

To be able to perform pulldowns on endogenously expressed protein variants, first we need to make sure that the affinity purification tag is in the same loci than the sequence variant we wish to study. Because our edited cell lines contained variants in the BAG3 gene in a heterozygous fashion, we will have to use methods for ‘chromosomal phasing’, that is, elucidating the arrangement of specific alleles as haplotypes. Multiple methods have been described for this, with the most traditional ones being involve long-range PCR, cloning, and/or manual dilution to single-molecule concentrations. In this study, we utilized droplet digital PCR as a rapid method for phasing the variants, as it has been described by others<sup>172</sup>.

By combining one of the allelic discrimination probes and a probe binding to the 3xFLAG sequence, we were able to identify clones that contained the BAG3 point mutation and the 3xFLAG in *cis* (i.e. in the same chromosomal DNA strand), as well as in *trans* (i.e. in separate chromosomes) (Figure 4.4). We observed that annealing temperatures can be optimized to allow for the single nucleotide variant probe to discriminate between alleles (Figure 4.4), probably because the mismatch results in less intensity droplets<sup>172</sup>. This allowed us to obtain allelic ratios in addition to the phasing. We also observed that the 3xFLAG cassette insertion happened independently of the chromosome (Figure 4.4).

Using this procedure, we were able to characterize and isolate cell lines BAG3<sup>P209L</sup>-3xFLAG/wt and BAG3<sup>C151R</sup>-3xFLAG/wt (with the single nucleotide mutations and the affinity tag in *cis*) and BAG3<sup>P209L</sup>/3xFLAG and BAG3<sup>C151R</sup>/3xFLAG (with editing events in *trans*). We concluded that ddPCR is a rapid, effective way to perform allelic phasing and quantification. We

anticipate that the main limitation of this approach is the distance between the loci to be phased, as maintaining the DNA strand intact becomes increasingly difficult for bigger sizes. In spite of this, we were able to comfortably perform phasing at a distance up to 7kb using minimum precautions, and other have reported detecting linkage up to 200kb<sup>172</sup>.

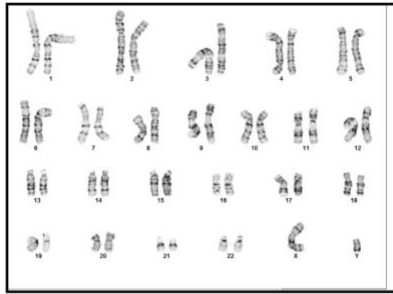


**Figure 4.4** – ddPCR-based assay for phasing of BAG3 variants and the 3xFLAG sequence. Using examples from the BAG3<sup>P209L</sup>-3xFLAG/wt line unless otherwise stated. (A) Schematic representation of the BAG3 gene, with where the genotyping probes bind. (B) The two possible scenarios for a cell line that is double positive for the induced BAG3 variant and the 3xFLAG sequence, along with representative ddPCR result graphs including the different droplet populations that are predicted. (D) Percentage of linkage goes down dramatically when the genomic DNA sample is treated with a restriction enzyme which cuts in between the two target sites (in this case, ECoRI). The non-zero value for the restricted sample is possibly due to DNA entanglement. Shown for BAG3<sup>C151R</sup>-3xFLAG. (E) Results from genotyping 19 clones. The 3xFLAG cassette seemed to integrate with equal probability on either allele.

#### **4.3.5 Homologous recombination replaces kilobase-length regions of the genome and enables for the simultaneous insertion of multiple modifications**

For the generation of the BAG3<sup>E455K-3xFLAG/wt</sup> cell line, we used an approach that allowed us to generate it from WTc cells in a single step. Since the site of the nucleotide responsible for the E455K transition is not too far from the stop codon (363bp), we reasoned we could just put this variant in the homology arm of the vector used for the insertion of the 3xFLAG expression cassette. The process was successful, with 70.4% of the clones containing both events co-segregated (Figure 4.3). As expected, all of the clonal populations isolated contained the E455K mutation and the 3xFLAG expression cassette *in cis* (Figure 4.3).

This suggests that the homology arm that is used as a template by the recombination machinery is also copied into the genome. This points to an application where regions in the length range of a kilobase (Figure 4.3) regions can be modified in multiple points on a single step of genome editing, significantly faster than sequential insertion of multiple variants.



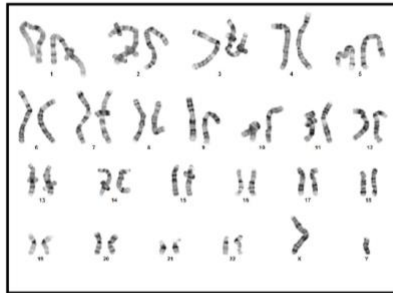
BAG3<sup>3xFLAG/wt</sup>: 46,XY[18/20]



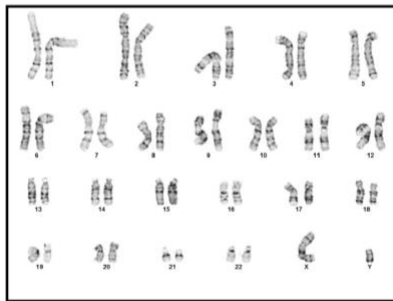
BAG3<sup>3xFLAG/3xFLAG</sup>: 46,XY[20/20]



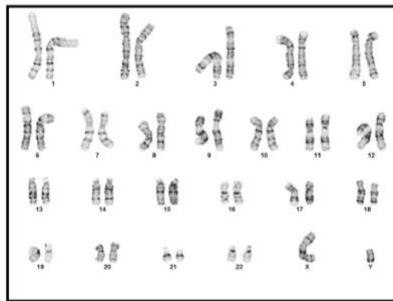
BAG3<sup>C151R-3xFLAG/wt</sup>: 46,XY[18/20]



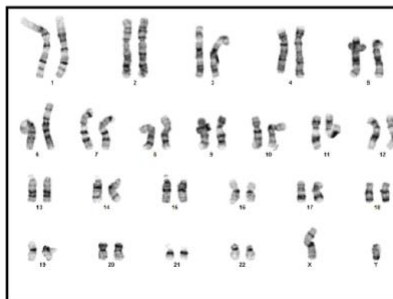
BAG3<sup>P209L-3xFLAG/wt</sup>: 46,XY[19/20]



BAG3<sup>E455K-3xFLAG/wt</sup>: 46,XY[20/20]



TetOn-BAG3<sup>3xFLAG</sup>: 46,XY[18/20]



HSPB7<sup>3xFLAG/wt</sup>: 46,XY[18/20]

**Figure 4.5** – Karyotyping of the cell lines obtained that were used in further experiments. All of them labelled as ‘Apparently NORMAL Human Male Karyotype’. Single cells that were not 46,XY were ‘non-clonal aberrations that are most likely artifacts of culture’.

## 4.4 Conclusion

Induced pluripotent stem cells offer an unprecedented resource to study human genetics and their impact on different cell types. This is especially important to bolster our understanding of tissue-specific human disease (e.g. cardiomyopathies). However, the inevitable genetic differences between cells from individuals makes it very difficult to follow an approach where an observed phenotype is directly linked with a specific genetic variant. In this chapter I have described how we were successful at introducing single nucleotide variant polymorphisms on the genomic copy of the heart disease-related gene BAG3. This allows us to study the effect of these specific variants against a control with an identical genetic background. In addition, we also succeeded at inserting a nucleotide sequence that results the expression of a BAG3-3xFLAG fusion protein, at endogenous levels of expression and following the same regulation pattern. This allows us to perform pulldowns and analyze interactors of the BAG3 protein and its variants by affinity purification-mass spectrometry (APMS). The two strategies followed also fulfill the requirements for the two main scenarios for genome editing for disease modeling purposes: single nucleotide editing with no surrounding genomic 'scar' that could influence protein coding, and long oligonucleotide insertion to enable fusion protein production.

The main caveat of these approaches is the how time consuming it can be to produce a stable cell line. We estimate it takes around 8 weeks to produce a clonal cell line for the scarless editing process, and a bit less (4-5 weeks) to arrive to a final product when using the aid of antibiotic selection cassettes. This makes it difficult to produce high number of variants

simultaneously, such as it would be required for screening. In this regard, the use of a combination strategy (such as what the one followed for BAG3<sup>E455K</sup> production) is very convenient, when distance between target loci allows. It is also important to always screen for potential off-target sites (most importantly in the unedited copy of the target locus). Finally, generating genetically homogeneous cell lines requires the expansion from a single cell (cloning), which may result in drifting of the population and enrichment of faster growing subclones that contain undesired genetic or epigenetic variants<sup>179,180</sup>. Most of these issues can be avoided if using rigorous quality control steps.

Genome editing technologies have opened the gates to a new era of studies aimed to understand human disease genetics. The editing of the endogenous copy of the gene guarantees that studies using the cell lines produced here will be more relevant than simple protein overexpression as the studied protein will reproduce the behavior of a physiologically expressed protein in a patient cell, eliminating any confounding factors from other genetic variants. The usefulness of this approach will be further demonstrated in the following chapters.

**Table 4.1** - Sequences of primers for genotyping reactions (CNV: copy number variation)

Description	Forward Primer	Reverse Primer	TaqMan probe
BAG3 <sup>3xFLAG</sup> upstream arm junction sequencing	TGCCATTAAGAAT ACCATCTACAGAG	GTGGTCCTTGTAGT CGCC	
BAG3 <sup>3xFLAG</sup> downstream arm junction sequencing	CTACAAGGACGAC GACGA	CACCACCTGGCTGA TTTG	

BAG3 <sup>3xFLAG</sup> 3xFLAG region sequencing	GTCTATGAACTCC AGCCCAG	CTACAAAAGACAGT GCACAAC	
3xFLAG (ddPCR)	GGTCAAGTCCAGG TCTATGAAC	TTGTCGTCGTCGTC CTT	ATCCGACTACAAG GAC
Puromycin resistance CNV (ddPCR)	GTCACCGAGCTGC AAGAA	CACCTTGCCGATGT CGAG	CTCTTCCTCACGC GCGTCGG
Neomycin resistance CNV (ddPCR)	CATGGCTGATGCA ATGCG	TCGCTTGGTGGTTCG AATG	CGCTTGATCCGGC TACCTGCC
BAG3 <sup>C151R</sup> ddPCR	CAGCAGCGGCTCC TCAGA	GTGAGCCTGGCCTC TCCTT	CAGATAAACAGTG TGGACA (wt) AGATAAACAGCGT GGACA (C151R)
BAG3 <sup>P209L</sup> ddPCR	ATGCCCTCTACCCT GTGTCTCTT	GGGTAACGTTCTGC TCGTGTATC	TACATCTCCATTCC GG (wt) TACATCTCCATTCT GG (P209L)
BAG3 <sup>E455K</sup> ddPCR	TGAAGGCAAGAAG ACTGACAAAAA	TCCTGGCCTGACGC ACAT	TACCTGATGATCA AAG (E455K)
WT AAV F	CGGTTAATGTGGC TCTGGTT	AGGATCCTCTCTGG CTCCAT	

**Table 4.2.** - Cell lines generated in this study (in **bold**, those used for APMS experiments in Chapter 6)

HGVS name <sup>1</sup>	Short name <sup>2</sup>	Notes
BAG3 c.[451T>C];[=]	BAG3 <sup>C151R/wt</sup>	
BAG3 c.[626C>T];[=]	BAG3 <sup>P209L/wt</sup>	
BAG3 p.[P575_*576ins3xFLAG];[=]	BAG3 <sup>3xFLAG/wt</sup>	C-terminal fusion of the 3xFLAG epitope sequence.

BAG3 p.[P575_*576ins3xFLAG]; [P575_*576ins3xFLAG]	BAG3 <sup>3xFLAG/3xFLAG</sup>	C-terminal fusion of the 3xFLAG epitope sequence. Both gene alleles contain a fusion with the 3xFLAG sequence.
BAG3 p.[C151R; P575_*576ins3xFLAG];[=]	BAG3 <sup>C151R- 3xFLAG/wt</sup>	The amino acid transition and the 3xFLAG fusion are in the same allele ( <i>in cis</i> )
BAG3 p.[P209L; P575_*576ins3xFLAG];[=]	BAG3 <sup>P209L- 3xFLAG/wt</sup>	The amino acid transition and the 3xFLAG fusion are in the same allele ( <i>in cis</i> )
BAG3 p.[E455K; P575_*576ins3xFLAG];[=]	BAG3 <sup>E455K- 3xFLAG/wt</sup>	The amino acid transition and the 3xFLAG fusion are in the same allele ( <i>in cis</i> )
BAG3 p.[C151R];[P575_*576ins3xFLAG]	BAG3 <sup>C151R/3xFLAG</sup>	The amino acid transition and the 3xFLAG fusion are in separate alleles ( <i>in trans</i> ). Not used in this study.
BAG3 p.[P209L];[P575_*576ins3xFLAG]	BAG3 <sup>P209L/3xFLAG</sup>	The amino acid transition and the 3xFLAG fusion are in separate alleles ( <i>in trans</i> ). Not used in this study.
HSPB7 p.[I170_*171ins3xFLAG];[=]	HSPB7 <sup>3xFLAG/wt</sup>	C-terminal fusion of the 3xFLAG epitope sequence.



PPP1R12C p.[=][=][TetOn- BAG3*ins3xFLAG]	TetOn- BAG3 <sup>3xFLAG</sup>	Insertion of a cassette expressing the cDNA of BAG3 in the PPP1R12C (AAVS1) locus
PPP1R12C p.[=][=][P209L_TetOn- BAG3*ins3xFLAG]	TetOn-BAG3 <sup>P209L- 3xFLAG</sup>	Insertion of a cassette expressing the cDNA of BAG3 in the PPP1R12C (AAVS1) locus
PPP1R12C p.[=][=][E455K_TetOn- BAG3*ins3xFLAG]	TetOn- BAG3 <sup>E455K-3xFLAG</sup>	Insertion of a cassette expressing the cDNA of BAG3 in the PPP1R12C (AAVS1) locus

<sup>1</sup>:nomenclature as per guidelines of the Human Genome Variation Society (HGVS)<sup>181</sup>.

<sup>2</sup>:based on protein level changes. Used in the rest of the text.

**Table 4.3** - Sequences of TALEN proteins (for TALEN-based engineering) and gRNAs (for CRISPR/Cas9-mediated editing).

Description	Sequence(s)
gRNA - inserting 3xFLAG cassette at BAG3's 5'UTR	GCAGAGGCTACGGTGCTGCT(GGG)
gRNA - inserting 3xFLAG cassette at HSPB7's 5'UTR	CTGTTGTAATGGGGTTAGCG(AGG)
TALEN - left for BAG3 <sup>C151R</sup>	TCACCTCTGCGGGGCATG
TALEN - right for BAG3 <sup>C151R</sup>	CAGATAAACAGTGTGGA
TALEN – forward target for BAG3 <sup>C151R</sup>	NG HD NI HD HD NG HD NG NN HD NN NN NN NN HD NI NG NN
TALEN – reverse target for BAG3 <sup>C151R</sup>	NG HD HD NI HD NI HD NG NN NG NG NG NI NG HD NG NN
Nucleotide donor for BAG3 <sup>C151R</sup>	TGCCAGAAACCACTCAGCCAGATAAACAGCGT GGACAGGTGGCAGCGGCGGCGGCAGCCC

TALEN forward for BAG3 <sup>P209L</sup>	NN NN NN HD NI NN NG HD NI HD HD NI NN HD NG HD HD
TALEN - reverse for BAG3 <sup>P209L</sup>	HD NG NN HD NG HD NN NG NN NG NI NG HD NI HD HD NN
TALEN - left for BAG3 <sup>P209L</sup>	TGGGCAGTCACCAGCTCC
TALEN - right for BAG3 <sup>P209L</sup>	TCTGCTCGTGTATCACCG
Nucleotide donor for BAG3 <sup>P209L</sup>	ACCAGCTCCCGCGGGGGTACATCTCC ATTCCGGTGATACACGAGCAGAACGTTACCCGGC
TALEN - insertion at AAVS1. Left	CCCCTCCACCCCACAGT
TALEN - insertion at AAVS1. Right	TTTCTGTCACCAATCCT

# Chapter 5 : Affinity Purification - Mass Spectrometry the functional study of two families of proteins and the comparison of disease-relevant BAG3 protein variants in an overexpression system

## 5.1 Introduction

The function of genes is in most cases determined by the interactions of the proteins they code for. Proteins do not perform their function isolated, but rather in combination with other protein partners that allow them to deploy the complex functions the cell requires <sup>1</sup>. Because of this reason, the comprehensive analysis of interaction partners allows us to understand the functions a protein is involved in, and its overall role in the cell<sup>6</sup>.

Beyond the basic understanding of how proteins function, a very important use of functional information on proteins is to study how genetic variation (loss of function, specific variants...) in a gene cause disease<sup>1,9</sup>. A common approach to understand how a specific protein product is linked to disease is to just knockout the gene of interest and look for a phenotype.

However, this becomes challenging since (a) sometimes a phenotype is not easily found,

especially when working with cell culture models, and (b) in many cases it is protein coding variants, and not loss of expression, that is associated to the development of disease (this is especially important for ‘gain-of-function’ variants<sup>9,182,183</sup>). The comprehensive analysis of protein interaction partners addresses these two issues. On the one side, it provides unbiased information on the protein function. Secondly, it enables studies where we can study the effect of protein coding variants by comparing the interactors (functions) of a protein and its variant (for example, a disease-associated variant)<sup>9,10,99</sup>. This comparative analysis can also aid in narrowing down which of the interacting partners of a protein are relevant for disease. This is particularly important for ‘scaffolding’ proteins that contain multiple domains and that interact with many different proteins and pathways. These proteins often work as ‘hubs’ and have been described to concentrate a higher amount of disease associated variants<sup>8,99</sup>.

In this chapter I will describe our use of affinity purification followed by mass spectrometry (APMS) to catalog the functions of a number of proteins of interest. Using protein overexpression in an immortalized cell line system, we studied the interactors for the DCM-related protein BAG3 and some of its disease variants. We also studied all the members of the BAG family of co-chaperones (including different isoforms of the BAG1 protein) and those of the members of the small heat shock protein (sHSP) family. I will also describe our experience testing different methods to increase our sensitivity at detecting protein interaction partners. In addition to novel insights obtained about the function of these proteins, these studies served as a proof-of-concept and tool development platform for the study of BAG3 variant protein interactions in the heart (as described in chapter 4).

## 5.2 Materials and Methods

### 5.2.1 Gene expression data analysis

Genome-wide gene expression data for normal human tissue and cell lines was downloaded from the Human Protein Atlas<sup>184</sup> project website ([www.proteinatlas.org](http://www.proteinatlas.org)). Data was analyzed and plotted using R software for statistical computing. Values were normalized by gene, across all tissues, and some tissues were collapsed into categories (as described in the Protein Atlas information page). Muscle tissues were left uncollapsed as they are more relevant to the purpose of this overall thesis project.

### 5.2.2 HEK293T cell culture and transfection

Immortalized Human Embryonic Kidney cells (HEK293T) were cultured in DMEM (high glucose, L-glutamine, sodium pyruvate; Gibco) supplemented with 10% Hyclone Fetal Bovine Serum (GE Life Sciences), GlutaMax (Gibco) and 1-00U/ml of Penicilin/Streptomycin (Gibco). Cells were harvested using 0.05% Trypsin solution (Gibco) and passaged at a 1:15 split ratio whenever they reached 80-90% confluence levels. For introduction and expression of DNA constructs, cells were transfected using the PolyJet In Vitro DNA Transfection (SignaGen) following provider's instructions. Cells were at a 40-60% confluence level at the time of transfection.

### 5.2.3 Affinity purification of 3xFLAG protein fusions

For each reaction of affinity purification, 5-10x10<sup>6</sup> cells were seeded on a 15cm dish. Approximately 2 hours later, or whenever they reach the appropriate confluence, cells were transfected with 10ug of plasmid as described above. Two days later, cells were harvested using

Phosphate Buffer saline (PBS) supplemented with 10mM EDTA and transferred to a 15mL conical tube. Cells were centrifuged 5 minutes at 1000 rpm and resuspended in PBS for washing. After two washes, the cell pellet was resuspended in 1mL of lysis buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.5mM EDTA, supplemented on the day with 0.5% NP-40 and 1x cOmplete protease inhibitor cocktail (Roche)). Cells were then lysed by rotating at 4C for 30 minutes, and lysates were cleared by centrifuging 20 minutes at 3000xg. For binding, the lysates were then added conditioned FLAG M2 Magnetic Bead (Sigma-Aldrich; 30ul of bead slurry per reaction) and allowed to rotate at 4C for 2 hours. After this, beads were collected using a magnetic stand and washed 3 times using wash buffer (50mM Tris-HCl, pH 7.5, NaCl 150mM, EDTA 0.5mM, supplemented with 0.05% NP-40) and once using wash buffer without NP-40. Then, beads were resuspended in 30ul of elution buffer (1% Rapigest SF Sulfactant (Waters), 5mg/ml 3xFLAG peptide (synthesized by Elim Biopharmaceuticals Inc), in wash buffer without additives) and incubated at room temperature for 30 minutes with constant shaking. Eluted protein solutions were then collected and a fraction was saved for western blot and the rest was submitted for desalting and mass spectrometry.

#### **5.2.4 Mass Spectrometry analysis of protein samples**

Eluted protein samples were reduced with DTT (2.5mM) at 50C for 30 minutes, and then iodoacetamide (2.5mM) was added and incubated for 40 minutes at room temperature in the dark. After that, 0.5ug of sequencing grade trypsin (Promega) was added to the sample and incubated overnight at 37C. The morning after peptides were desalted and concentrated on ZipTip C18 pipette tips (Millipore) following according to the manufacturer's protocol. Peptides were eluted on 0.1% Formic Acid.

Digested peptide mixtures were analyzed by LC-MS/MS on a Thermo Scientific Velos Pro ion trap mass spectrometry system equipped with a Proxeon Easy nLC 1000 ultra high pressure liquid chromatography and autosampler system. Samples were injected onto a pre-column (2 cm x 100  $\mu$ m I.D. packed with ReproSil Pur C18 AQ 5  $\mu$ m particles) in 0.1% formic acid and then separated with a two-hour gradient from 5% to 30% ACN in 0.1% formic acid on an analytical column (10 cm x 75  $\mu$ m I.D. packed with ReproSil Pur C18 AQ 3  $\mu$ m particles). The mass spectrometer collected data in a data-dependent fashion, collecting one full scan followed by 20 collision-induced dissociation MS/MS scans of the 20 most intense peaks from the full scan. Dynamic exclusion was enabled for 30 seconds with a repeat count of 1.

Raw data files were converted into peak lists using UCSF's in-house software PAVA<sup>185</sup>. Spectra were then searched using Protein Prospector 5.10.1 ([http://http://prospector.ucsf.edu/](http://prospector.ucsf.edu/)) using the SwissProt database of human proteins (April 2012). One missing cleavage was allowed. Fragment mass tolerance was set as 0.8 Da and parent mass tolerance as 1 Da. As for modifications, carbamidomethylation of cysteines was set as constant and acetylation of protein N-termini and methionine oxidation and methionine loss at N-termini were set as variable. The results from Protein Prospector were further filtered as follows: minimum Protein Score of 22.0, minimum Peptide Score of 15.0, maximum Protein E-Value of 0.01 and maximum Peptide E-Value of 0.05.

### 5.2.5 Analysis of APMS data

All APMS experiments were checked for bait counts, total spectral counts and total number of unique proteins identified. Samples with low or no bait expression or presenting obvious deviations from the other replicates were discarded. In the end 5 replicates were used per bait for downstream analysis. For controls, cells transfected with empty vector (not expressing any protein fused to the 3xFLAG epitope; 4 replicates) as well as cells transfected with a plasmid expressing the 3xFLAG peptide (4 replicates) were used. As an additional step, for the removal of sample carryover (peptides identified that come from the previous MS run) we implemented an additional filter similar to what others have described<sup>186</sup> by discarding entries with less than half the spectral counts of the previous run and that are present in less than 30% of all the experiments. To deal with non-unique peptides, we took only the first protein listed for each protein group (usually the one with the most unique peptides) and any other protein in the group that had at least one unique peptide across all replicates.

For the discrimination of protein interactors versus nonspecific binding partners, the SAINTexpress algorithm was used<sup>187</sup> with the sample compression option turned on to 4 samples ('-R4'). Control experiments were not compressed for SAINTexpress analysis as a way to minimize the impact of unremoved carryover in the control samples. SAINTexpress was run independently for each bait vs controls, to prevent the slight penalty that is imposed to proteins present in multiple experiments. A database containing all GO terms with less than 20 members was provided to boost scores of some interactors (as described by Teo et al<sup>187</sup>). For the implementation of the two stage poisson model (TSPM) the relevant script form



Fischer et al<sup>188</sup> was used. We used an FDR cutoff of 0.1 for both algorithms. Only significant interactors identified by SAINTexpress were used in subsequent analyses (see Figure 5.1).

For data manipulations and analyses, R version 3.4.1<sup>168</sup> was used. Quantile normalization was implemented using the Limma package on R<sup>189</sup>. The dot plots for visualization of the BAG and sHSP family were generated using the ProHits-Viz suite<sup>190</sup>. For the comparison of BAG3 variants (Figure 5.6), spectral count values were further normalized to equate BAG3 counts, except for the truncation variants. For visualization of network graphs, Cytoscape<sup>169</sup> version 3.4.0 was used. We used the iRefIndex 14.0<sup>191</sup> database as reference database for physical interactions. For recovery curves (Fig 3.1b-c) we used those baits with the most entries in the database (BAG1-3 and 6, HSPB1, 5 and 8, and DNAJB1). Cytoscape network visualizations were interconnected using the CORUM database of protein complexes<sup>192</sup> (May 2017 release).

### **5.2.6 Sequence analyses**

Sequence alignments were done using ClustalW<sup>193</sup>, and manually adjusted and annotated using JalView<sup>194</sup>. Alignment and tree visualization were performed on JalView.

### **5.2.7 Western blot and immunocytochemistry**

For immunostaining of cells in culture, cells were washed with PBS and fixed with 4% Paraformaldehyde at room temperature for 20 minutes. Then cells were washed three times with PBS supplemented with 0.1% Triton X-100 (PBS-Triton) and incubated with 5% Bovine Albumin Serum in PBS-Triton for 1 hour at room temperature to block and permeabilize. Then cells were incubated overnight at 4C with a solution of primary antibody in PBS-Triton.

The morning after, cells were washed three times with PBS-Triton and then incubated in secondary antibody in PBS-Triton for 1 hour at room temperature. Secondary antibodies used were Goat Anti-Rabbit Alexa Fluor 594 and Goat Anti-Mouse IgG Alexa Fluor 488 (both at a 1:500 dilution; both from Thermo Fisher Scientific). For primary antibodies, we used rabbit anti-BAG3 (Protein Tech, 10599-1-A). Cells were imaged using a BZ-9000 microscope (Keyence).

For western blotting, 20-30ug total protein extracts or 1:20 of total affinity purified protein were used. Samples were prepared in NuPAGE LDS Sample Buffer and NuPAGE Sample Reducing Agent (Thermo Fisher Scientific) following directions by provider. Samples were loaded on NOVEX 4-12% Bis-Tris Protein Gels (1.5mm thick; Thermo Fisher Scientific) and electrophoretic separation was performed on an XCell SureLock mini-cell (Thermo Fisher Scientific) at 150V on NuPAGE MES SDS buffer (Thermo Fisher Scientific). Protein was transferred into nitrocellulose membrane using the iBlot Dry Blotting System (program P3; Invitrogen) and blocked with Odyssey Blocking Buffer (PBS) (LiCor) for 1 hour at room temperature. Then the membrane was incubated overnight at 4C with a primary antibody in PBS supplemented with 0.1% Tween (PBS-Tween). The day after, primary antibody was washed three times with PBS-Tween and membranes were incubated with secondary antibodies for 1 hour at room temperature. Membranes were then imaged using an Odyssey FC Imager (Li-Cor), and bands were analyzed and quantified using the ImageStudio Software (Li-Cor). Secondary antibodies were Goat Anti-Mouse IRDye 680LT (dilution 1:20000) and Donkey Anti-Rabbit IgG IRDye (dilution 1:10000) (both by Li-Cor). Primary

antibodies and their dilutions used were: Anti-BAG3 rabbit polyclonal (1:2000, ProteinTech, 10599-9-Ap), Monoclonal ANTI-FLAG M2 (1:1000, Sigma-Aldrich, F1804).

### **5.2.8 Flow Cytometry Protein Interaction Assay (FCPIA)**

The assay procedure was adopted from previous reports<sup>32</sup>. Briefly, biotinylated Hsp70 was immobilized (1h at room temperature) on streptavidin coated polystyrene beads (Spherotech). After immobilization, beads were washed to remove any unbound protein and then incubated with labeled BAG3 proteins at the indicated concentrations. Binding was detected using an Accuri™ C6 flow cytometer to measure median bead-associated fluorescence. Beads capped with biocytin were used as a negative control, and non-specific binding to beads was subtracted from signal.

### **5.2.9 Luciferase Refolding Assay**

Experiments were performed as described previously<sup>32</sup>. In brief, luciferase (Promega) was denatured in 6 M GnHCl for 1 h at room temperature and then diluted into a working solution of Hsp72 (HSPA1B) in buffer containing an ATP regenerating system (23 mM HEPES, 120 mM KAc, 1.2 mM MgAc, 15 mM DTT, 61 mM creatine phosphate, 35 units/ml creatine kinase, and 5 ng/ l BSA (pH 7.4)). Various concentrations of NEF and J protein were added, and the reaction was initiated with the addition of ATP (1 mM). Sodium phosphate (10 mM) was added as indicated. The assay proceeded for 1 h at 37 °C in white, 96-well plates, and luminescence was measured using SteadyGlo luminescence reagent (Promega).

## 5.3 Results and discussion

### 5.3.1 Data processing for the discrimination of true interactors from nonspecific interactions

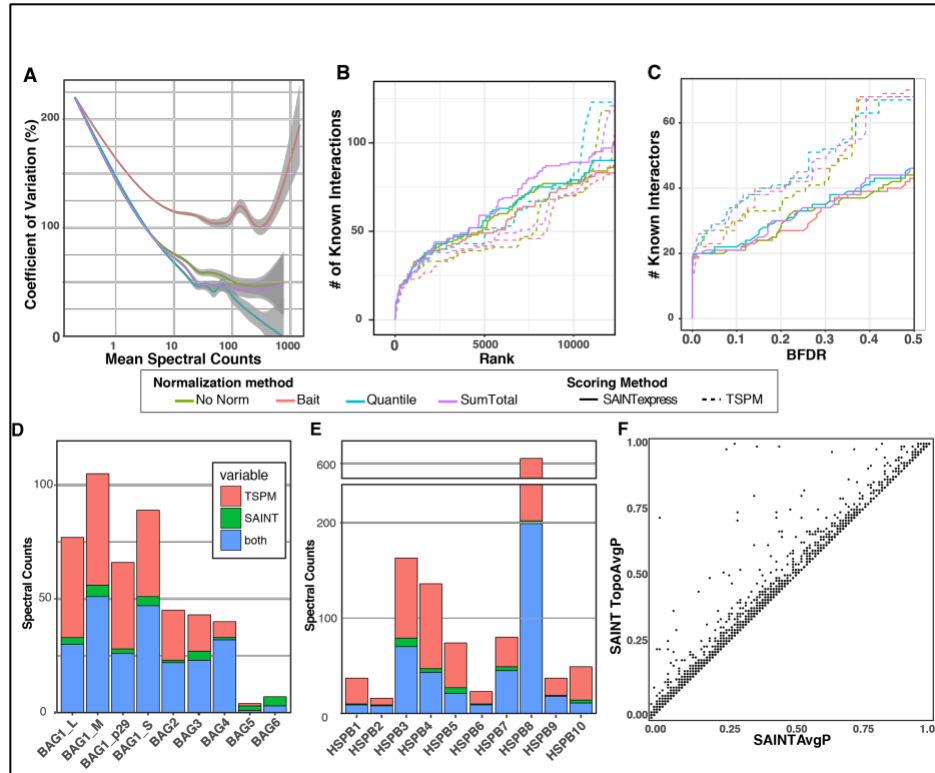
The dataset analyzed here consisted of 158 pulldowns for 30 baits (9 for the BAG family of interactors, 10 small Heat Shock proteins, 11 BAG3 variants; everything with 5 replicates, plus two types of controls with 4 replicates each). The use of 5 replicates is justified here as a way to compensate for the high variability observed in the samples (probably due to the inherent variability of the instrument and column used, the use of spectral counts as quantitative measure, and to the fact that the samples were run on different days and order). Using more replicates also increases stringency of the SAINT analyses<sup>131,187</sup>. As a quantitative measure of protein abundance, we decided to use the number of ‘spectral counts’, which is the number of times a peptide is being selected for fragmentation and subsequent mass analysis. This method is not as sensitive or robust as other methods based on ion current (area under the peak of MS1 spectra), but it is the most quantitative method given the instrument we used and the kind of data we obtained<sup>195,195–197</sup>.

For preprocessing of the data, we first investigated the influence of different normalization methods for the data:

- Quantile normalization, as originally developed for microarray expression data analysis<sup>198</sup> and later implemented by others for analysis of spectral count APMS data<sup>188</sup>. This normalization equalizes the spectral count distributions across replicates.
- ‘SumTotal’ normalization, where the spectral count values are divided by the total spectral counts in the sample, and multiplied by a constant value across all runs. This is the most commonly used form of normalization in AP-MS spectral count data.
- Bait normalization, where the intensity of the bait is equalized for all replicates of the same pulldown. This is *a priori* a good choice for the analysis of AP-MS data since the intensity of the interactors is expected to be proportional to the amount of bait.
- Alternatively, one can skip the normalization step. No normalization is recommended when the replicate variability of the dataset is very low.

A side to side comparison of the result obtained with these four methods reveals that the ‘bait’ normalization method results in inflated variability across replicates (Figure 5.1a). This is probably due to variability in the lysis and wash conditions across experiments, resulting in different interactor recovery, and also due to the nonlinear correlation between spectral counts and protein abundance. This lack of correlation is expected to be stronger highly abundant peptides, such as the ones of the bait (this is due to the mass analyzer being programmed to perform ‘dynamic exclusion’<sup>196,199</sup>). The other methods had comparable levels of variation across replicates, with quantile normalization reducing it the most (as expected). A look at the recovery of known interactions post-scoring reveals that bait normalization

performs poorly, while SumTotal seems to be marginally better than the other two methods (Figure 3.1b).



**Figure 5.1** - Effect of normalization method on protein counts and on recovery of known interactions. (A) Effect of normalization on variation of spectral counts across replicates. Most relevant hits are in the 10-1000 range of spectral counts. At this range, the variability of the ‘bait’ normalization method is significantly higher. Quantile normalization reduces variability the most. Gray shade: loess fitting 95% confidence interval. (B, C) Recovery of known interactions from the iRefIndex database, for the different normalization and scoring methods. SumTotal normalization seems to perform marginally better, and ‘bait’ normalization is clearly the worst at recovering known interactions. The number of hits and BFDR are cut at 10,000 and 0.5, respectively, for the sake of readability. (D and E) Significant interactors (BFDR cutoff of 0.1) for the BAG and sHSP families of proteins. The two-stage Poisson model (TSPM) recovers more interactors but at the cost of being much less stringent with the scoring, potentially increasing false-positives. (F) Influence of including prior information in the scoring of interactions (in this case, if both proteins are in the same small GO term group).

An outstanding issue when analyzing affinity purification-mass spectrometry data for protein interactions is how to tell the difference between a ‘real’ interaction of the protein (happens *in vivo*, functionally relevant) and a nonspecific interaction (‘contaminants’ and ‘sticky’ proteins that bind to the antibody, the beads, or the tag itself, and that do not help understand the function of the protein *in vivo*<sup>130,200</sup>). A number of algorithms are available for this purpose, each with its own assumptions, advantages and drawbacks<sup>131,132,186,188</sup>. We decided to test the use of two: significant analysis of interactome (SAINT) and a two stage Poisson model (TSPM). SAINT was first described for the analysis of a protein kinase and phosphatase interaction network in yeast<sup>201</sup> using spectral counts. Multiple modifications and improvements of the method have been developed since, correcting some assumptions and adapting it to work with other measures of protein abundance<sup>131,187,202–204</sup>. SAINT uses bayesian statistics to compute the posterior probability that a given bait-prey pair is a true interaction, based how significantly higher the spectral counts of the prey are for this bait compared to control pulldowns<sup>131</sup>. SAINT has been used successfully in many studies, and the main limitation that has been found is that it penalizes very highly those interactions with highly variable spectral counts and those with very low spectral counts, regardless of how low representation they have in the control samples (as reported by others<sup>188,205</sup> and observed by us). For our studies we used the SAINTexpress<sup>187</sup> implementation SAINT, as it is better suited for addressing highly interconnected interactomes where a prey can have different average spectral counts for different baits. TSPM models the spectral count data as coming from a Poisson distribution, and generates two different generalized linear models: in one of them there is no association between counts of a specific prey for a bait and for the control, and in

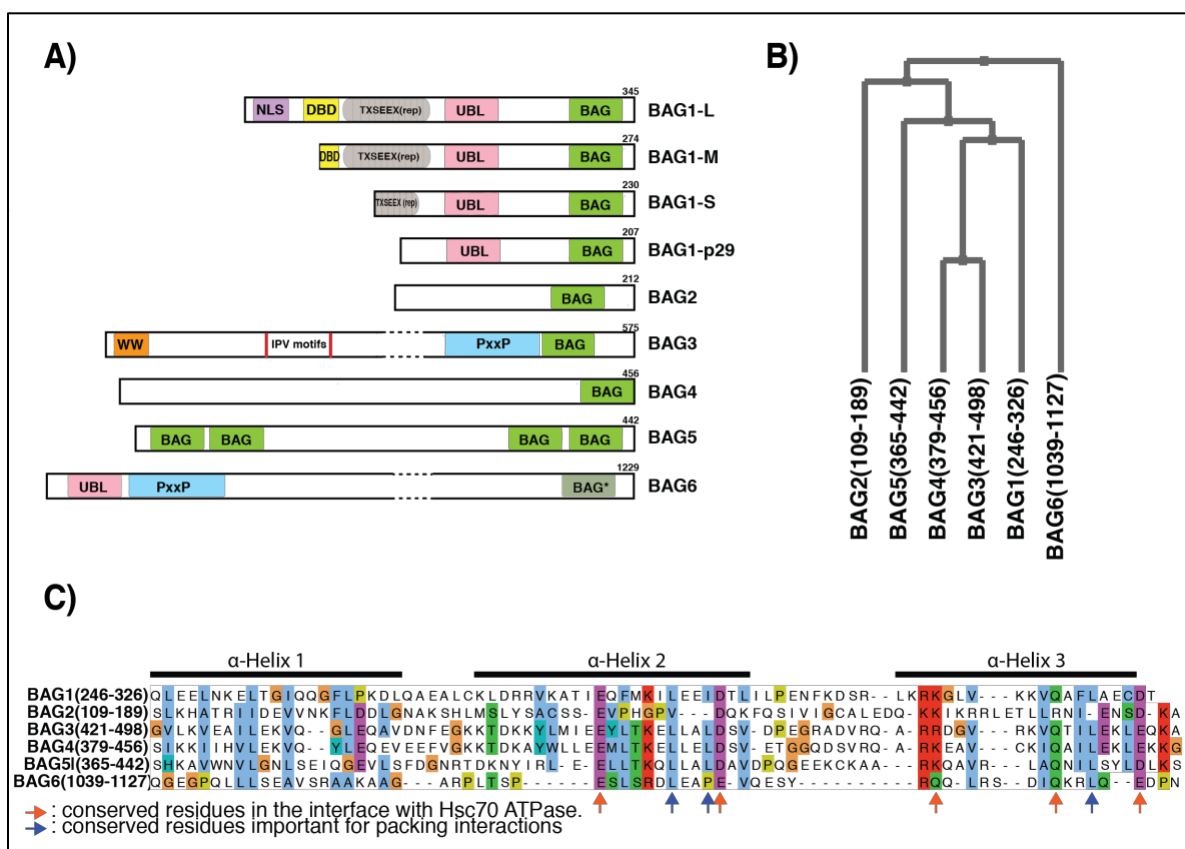
the other one there is such association. A likelihood ratio test is then performed comparing the two models, and an interaction is considered true if the no association model is rejected in favor of the second model. This approach was first described by Fischer et al<sup>188</sup> as a complementary approach to SAINT that addresses the two limitations highlighted earlier. They also found it to be substantially less conservative than SAINT, resulting in more false positives. The TSPM approach has not been used in any other published studies to date, but we decided to test it along with SAINTexpress to see if it improved our results dramatically.

A side to side comparison of the recovery of known interactions using SAINTexpress and TSPM reveals that TSPM yields a better recovery for the same false discovery rate cutoff (BFDR) (Figure 5.1c). However, this is at the cost of increasing the list of significant interactors by almost two-fold compared to SAINT (Figure 5.1d and e). This was more obvious when we looked at the rate of recovery in a list of interactors, ranked by their given score (Figure. 5.1b). Most of the interactors identified by SAINT as significant were also identified by the TSPM, and although it is possible that TSPM identified some true interactors that SAINT missed, we considered that the list was too extensive (especially for some of the baits) and decided to use SAINT only for our analyses. This results in a list of high-confidence interactions at the cost of some potential false-negatives. To improve our predictions, we implemented prior information on known protein complexes and pathways, as allowed by the SAINTexpress algorithm<sup>187</sup>. We used a list of small GO terms (less than 20 members) as a source for prior information. The addition of this prior information only boosted the scores of a small number of the hits significantly (Fig 5.1F).



### 5.3.2 Unbiased analysis of the stable interactions of the BAG family of co-chaperones reveals member-specific functions

The BAG family of proteins is composed of 6 members. All of them have in common having a BAG domain in the C-termini (except BAG5, which has five putative BAG domains). Aside from that, their domain structure is quite diverse and includes some regions typically mediating protein protein interactions (Fig 5.2a). The function of this family has been explored by studies focusing on the BAG1 member of the family, whose main function is acting as a co-chaperone to the HSP70 family of molecular chaperones<sup>31,206,207</sup>. The BAG domain is responsible for this activity<sup>32</sup>. Other studies have linked subsequent members to have the same function<sup>140,208–210</sup>. The HSP70 family of chaperones is expected to be involved in the co-translational or post-translational folding of roughly a 20% of all the proteome, exhibiting preferential binding to stretches of hydrophobic amino acid residues<sup>31,211,212</sup>. To perform their activity on such a broad range of ‘client’ proteins, HSP70 chaperones rely on a complex and extensive network of ‘co-chaperones’ that assist with the binding activity but also direct the chaperone to specific targets<sup>31,32,211,213,214</sup>(Figure 5.2b). As such, the BAG family of co-chaperones act as nucleotide exchange factors (mediating the release of ADP and subsequent release of the client protein<sup>215,216</sup>), probably through their BAG domain, and also bridge HSP70 chaperones with specific clients, potentially through the non-BAG region of the protein<sup>32</sup>.



**Figure 5.2** – The BAG family of co-chaperones. (A) Domain structure of the BAG family of co-chaperones. Note that BAG6 does not contain an active BAG domain. DBD: DNA binding domain; UBL: Ubiquitin-like domain; TXSEEX repeat: repeat of such aminoacids (X=any aminoacid). (B) Tree view of the alignment of BAG domains. The BAG6 being. (C) Alignment of the amino acid sequence for the BAG domain of the different BAG family proteins. For BAG5, only the last BAG domain is shown. Colors represent amino acid categories (Clustal color scheme) for most conserved amino acids. Annotations of alpha helix regions and specific amino acids are taken from Sondermann et al<sup>215</sup>. The third helix for BAG6 BAG domain is substantially different form the other BAGs.

In this study, we performed AP-MS on all the BAG family members, transiently overexpressed on an immortalized cell culture line (HEK293T). Around 25-50 putative interactors were identified for all members except for BAG5 and BAG6, for which only 2 and 6 interactors were identified, respectively (Figure 5.1d). It is important to note that baits were identified with high spectral counts in for the BAG5 and BAG6 AP-MS experiment,

suggesting that the lack of interactions for these members are not due to problems of expression of the bait.

An initial inspection of the local interaction network reveals that a many of the interactions are shared between the BAG1-4 members of the family (Figure 5.3). This core set of interactions is formed mostly by different members of the HSP70 family of chaperones, and a number of its associated co-chaperones. Some of these are members of the HSP40 family of co-chaperones (DNAJ proteins), which typically mediates the hydrolysis of ATP to lock substrates into the chaperone. This would imply that they act in different stages of the HSP70 cycle than nucleotide exchange factors like the BAG proteins<sup>31,211</sup>. It could be that the observed interaction is from the bait being a client of the chaperone (unlikely given that we dont observe this for all of them) or that the proteins are indeed still bound but not performing its enzymatic action at the same time. The presence of HSP70 and other co-chaperones as interaction partners of the BAG proteins validates their well-known role (or at least most of them) as HSP70 co-chaperones. Others putative interactors that are shared between multiple members include the HSP70-associated ubiquitin ligase STUB1/CHIP, the translation elongation factor 2 (EEF2), signaling pathway molecules such as SNW1 and IRS4 and other miscellaneous function proteins (CAD, AMOT, TMPO).

In addition to this canonical role, the dataset also contains a number of member-specific interactions that suggest unique functions that each one of the members performs in the cell, potentially in respect to how the clients of HSP70 are selected or handled post-folding<sup>31,34,212,213</sup> (Figure 5.3).

### *BAG1 and its four alternative translation isoforms*

The BAG1 member was the first BAG protein to be described, and a lot of insight about the nucleotide exchange factor function of the BAG domain has been extracted from this one<sup>206,207</sup>. BAG1 also represents a special case where same mRNA codes for different isoforms that are generated from alternative translation start points. Because of this, the isoforms share the C-terminal BAG domain and a Ubiquitin-like domain (UBL), and they differ in the sequence in the N-termini (Fig. 2.2)<sup>217</sup>. A very comprehensive review of the functions known for these isoforms was written by Gehring eight years ago<sup>217</sup>. In our study we decided to perform AP-MS on the four most common isoforms, hoping to expand our knowledge of the functional differences between isoforms. The most common isoform is BAG1-S, containing the BAG domain and the UBC domain. Two common longer isoforms exist: BAG1-M, which also contains an acidic hexarepeat region and an N-terminal basic region that is predicted to enable interaction with DNA; and BAG1-L, which contains a longer DNA binding domain and a full nuclear localization sequence. Finally, there is also a less abundant isoform that is smaller than BAG1-S and that we will call BAG1-p29. The subcellular location of the smaller isoforms is cytoplasmic, while BAG1-L is mostly nuclear (but also cytoplasmic) and BAG1-M is mostly in the cytoplasm but gets translocated to the nuclei upon heat stress. Both BAG1-M and BAG1-L have been described to influence transcription, potentially through their DNA-binding domain<sup>218,219</sup>.

When looking at our list of significant putative interactors of BAG1, we find that all of them co-precipitate with HSP70 chaperones and associated factors (see above). HSP70 is the

primary interaction of BAG1, for which it works as a nucleotide exchange factor in a dose-dependent manner<sup>32,206,217</sup>. The lower spectral counts for BAG1-L are most likely due to the lower recovery of this bait (probably because of its nuclear localization and size). It is noteworthy that SGTA, a tetratricopeptide co-chaperone that binds HSP70, co-precipitates only with BAG1 in our dataset. Also specific to the BAG1 isoforms and no other BAGs we find a high number of the subunits of the 19S regulatory particle of the proteasome. Indeed, BAG1 has been described to bridge the HSP70 machinery and the proteasome to facilitate the disposal of misfolded clients<sup>220</sup>. This interaction is mediated by the ubiquitin-like domain.

When comparing the BAG1 isoform results, there was no difference between BAG1-S and BAG1-p29. This is not surprising, as they are only different by 13 amino acids that have no assigned function. The functional relevance of the BAG1-p29 isoform remains unknown. As for interactors specific of the BAG1-S isoform (or shared with the also cytoplasmic BAG1-M) we find a couple of proteins involved in protein posttranslational modification (DPH2, ASNS), some proteins involved in RNA and/or the mitochondrial metabolism (RRM2, CSDE1, STRAP, SLC25A, ENDOG), a number of proteins without well known function (Transgelin-2, ARFIP1, ARMC1). Taken together, these hits provide little insight on BAG1-S-specific functions, other than that it might play a role in mitochondria. Last but not least, we find exclusively in BAG1-S (and a bit on BAG1-p29) the STIP1/HOP co-chaperone, known to bridge the HSP70 and HSP90 families. It is surprising that this factor only shows up in these baits. This might be due to a stronger interaction of the shorter forms of BAG1 with HSP70 or simply to a better recovery for these baits, although there is at least one study that links BAG1-S directly to Hop<sup>221</sup>.

As for the longer isoforms of BAG1, the most noticeable thing is that both of them show putative binding to the ribosome. Although these are often considered nonspecific interactions, they were found specifically in these two isoforms. Additionally, a previous study reported an interaction between the fungal BAG1 homolog SNL1 and the ribosome, and predicted a similar behavior for the BAG1-M and BAG1-L (but not BAG1-S) human isoforms<sup>222</sup>. This interaction would be mediated by the short, positively charged sequence in the N-termini of the proteins, and points to the interesting hypothesis that the longer forms of BAG1 are enabling an interaction between translating ribosomes and the HSP70 chaperones. Alternatively, since BAG1-L is mostly nuclear and BAG1-M is inducibly nuclear, it could be that they are involved in quality control of ribosomal proteins inside the nuclei, or assist in the assembly of the ribosome. Agreeing with the latter hypothesis, both nucleolin and nucleophosamin, proteins involved in ribosome synthesis and maturation, co-precipitate with BAG1-M and BAG1-L. On the other hand, in accordance with the co-translational idea, two putative interactors of BAG1-M (SRP68 and SRP72) but not BAG1-L are main members of the signal recognition particle (SRP). This complex is involved in targeting nascent polypeptide chains and shuttling them to the endoplasmic reticulum for assembly in the membrane<sup>223</sup>. Other BAG1-M specific interactors include a couple of proteins involved in mRNA stability and splicing (PTBP1, YTHDF2), the microtubule associated protein MAP4, two histones, and the enzyme PDE8A. As for BAG1-L specific interactions, the most significant is a strong (high spectral counts) interaction with the C1QBP protein. C1QBP has been involved in multiple cellular processes, including but not limited to ribosome biogenesis, transcriptional regulation, inflammation and apoptosis<sup>224</sup>. The fact that it pulls down with

BAG1L so specifically and strongly suggests that it is binding to some motif contained within the first 71 amino acids of the protein. Follow up studies would be required to clarify which one of these functions and localizations is relevant for BAG1 interaction. Another interesting hit for BAG1-L is the Lamin B receptor (LBR), a protein that resides in the inner side of the nuclear membrane and also has been associated with a plethora of functions<sup>225</sup>.

### BAG2

BAG2 is the smallest member of the BAG family of co-chaperones (after BAG1-p29), and no domains other than the BAG one have been identified in the sequence. Our data shows very few significant hits, aside from those associated with the HSP70 co-factor function (see above). One of the hits, Arfaptin-1 (also with BAG1) is a currently uncharacterized protein that has been described in different independent studies to interact with ubiquitin (by AP-MS)<sup>226–228</sup>. This could point to a new link to ubiquitinated proteins and the proteasome.

Just like BAG5 (discussed later), BAG2 has been described to regulate the levels and function of some neurodegeneration-related proteins<sup>229–232</sup>. We do not observe any of these hits in our dataset, probably due to the cell type used in our experiments. Finally, BAG2 has been observed to express at higher levels in some cancers (a common feature of most members of the BAG family and of many chaperones in general)<sup>233</sup>.

### BAG3

BAG3 is arguably the most studied member of the BAG family of co-chaperones. This is mostly due to its implication on multiple diseases<sup>58,234</sup>, mostly dilated cardiomyopathy<sup>17,18,23,35</sup>

and cancer<sup>22,235–238</sup>. BAG3 is a scaffolding protein that integrates multiple pathways through its different interaction domains<sup>22,140,235–239</sup>. Most notably, a BAG3-CHIP/STUB1-HSPB8 interaction has been described to mediate a specific of chaperone-assisted selective autophagy<sup>43,44,240</sup>. In addition to all the interactions shared with BAG1, 2 and 4, our data contains interactions with WW-binding proteins (WBP2 and UBAP2L<sup>241</sup>) and with the lim domain-containing protein PDLIM7. BAG3 also contains two IPV motifs that have been associated with interactions with the members of the small heat shock protein chaperones (sHSP), constituting the only link between this family and the HSP70 system<sup>138,140,242</sup>. Consequent with this observation, we identified the small heat shock protein HSPB1 as one of the hits. Although BAG3 is supposed to interact with multiple members of the sHSP family, HSPB1 is the most abundant in the HEK293T cells used in this study (as shown by the Human Protein Atlas<sup>29</sup> [www.proteinatlas.org](http://www.proteinatlas.org)). Further discussion of the BAG3 interactome can be found later in this chapter.

#### BAG4

A lot of the significant hits identified for BAG4 in our dataset are shared with BAG1-3. This supports the idea that BAG4 develops a function as HSP70 co-chaperone<sup>208</sup>. BAG4 was originally described to play a role in preventing spontaneous activation of the TNF signalling pathway by binding to the TNF receptor 1<sup>243</sup>. It has also been found overexpressed in some forms of cancer<sup>244,245</sup>. Interestingly, our experiments point towards a novel role of the BAG4 co-chaperone in the mRNA degradation pathway. Five of the significant hits are involved in the steps of deadenylation (CNOT1) and decapping of mRNA upon miRNA



binding<sup>246,247</sup>. Another hit, PRDX3, was found in a previous study as to be a representative target of the argonaute complex for miRNA-mediated degradation<sup>248</sup>. This data hints to the interesting hypothesis that BAG4 may constitute the bridge of the mRNA degradation complex and HSP70 protein quality control, either by facilitating assembly or disassembly of the complex or by promoting removal of proteins translated by the degrading mRNA.

### BAG5

We were able to identify only a couple putative interactors for the BAG5 protein. One of these proteins was lipolysis-stimulated lipoprotein receptor (LSR), a tricellular cell tight junction protein that may be involved in the metabolism of lipids and in the invasion of some forms of cancer<sup>249,250</sup>. The other one was MLF2, a protein for which little is known but that we see in most of our experiments where HSP70 is also a significant hit. Most of the previous studies on BAG5 have focused on its involvement in the quality control of various neurodegeneration-associated proteins such as Ataxin, PINK, PRKN or alpha-synuclein<sup>232,251–253</sup>. These are all neuron-specific proteins that we would not be able to recover here. It has also been described to be overexpressed and inhibit apoptosis in some forms of cancer<sup>254</sup>.

The BAG5 protein contains five putative BAG domains, of which on the C-terminal one acts as a nucleotide exchange factor for HSP70<sup>210</sup>. In spite of this, we did not find any HSP70 chaperone in the list of significant hits. Looking at the spectral counts of the BAG5 bait we realized that there was a lot of variability. We reasoned that this high variability could be penalizing BAG5 interactions strongly when using SumTotal normalization. When analyzing the data using quantile or bait normalization, we found that the list of significant

interactors remained the same, except that now it also included both HSPA1A and HSPA8. We did not find this same issue when analyzing data for any other BAG family members, which suggest that there were specific problems expressing or pulling down this bait, and/or that the interaction with HSP70 is weaker than for the other members.

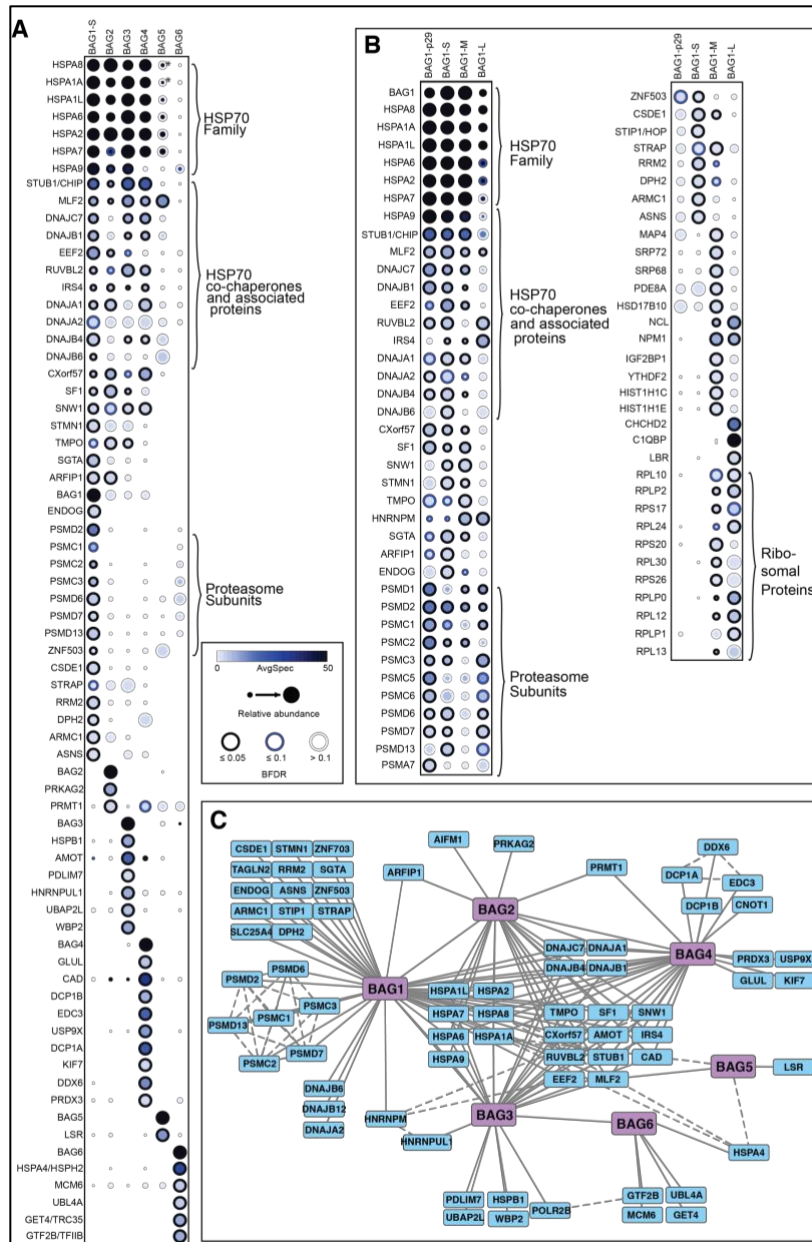
### BAG6

Although the amount of putative interactors identified for BAG6 is small, our list agrees with previous studies on this protein. First, the low recovery of HSP70 family members is consistent with a recent study that revealed that the BAG domain in the BAG6 protein does not mediate an interaction with the HSP70 chaperone, being used instead for binding to other partners<sup>255</sup>. Interestingly, BAG6 was the only protein in the dataset that pulled down with the HSPA4 (member of the HSP110 family) chaperone with high spectral counts. This suggests a previously unreported interaction with this bona fide chaperone that can also bind to HSP70 and has nucleotide exchange factor and disaggregase activity<sup>256,257</sup>.

Along with TRC35/GET4 and UBL4A (also found in this dataset), BAG6 is known to play an important role in the maturation of endoplasmic reticulum transmembrane proteins, by sheltering them through the cytosol until they are relayed to the delivery factor TRC40<sup>258</sup>. Unfortunately, we did not seem to identify any candidate transmembrane protein in our dataset.

The BAG6 protein sequence contains an NLS motif that accounts for its reportedly nuclear localization, although it is also known to carry out functions in the cytoplasm as a

result of the masking of the NLS by GET4/TRC35<sup>255,259</sup>. However, our suggests that BAG6 may bind the transcription initiation factor IIB (TFIIB) and at least one member of the MCM helicase complex. This points towards a novel implication of BAG6 in transcriptional regulation or in the proteostasis of very important nuclear proteins. Finally, BAG6 contains a ubiquitin-like domain just like BAG1, but we did not find any potential interaction partner that pointed towards an involvement in proteasomal degradation (as it has been described by others<sup>260</sup>).



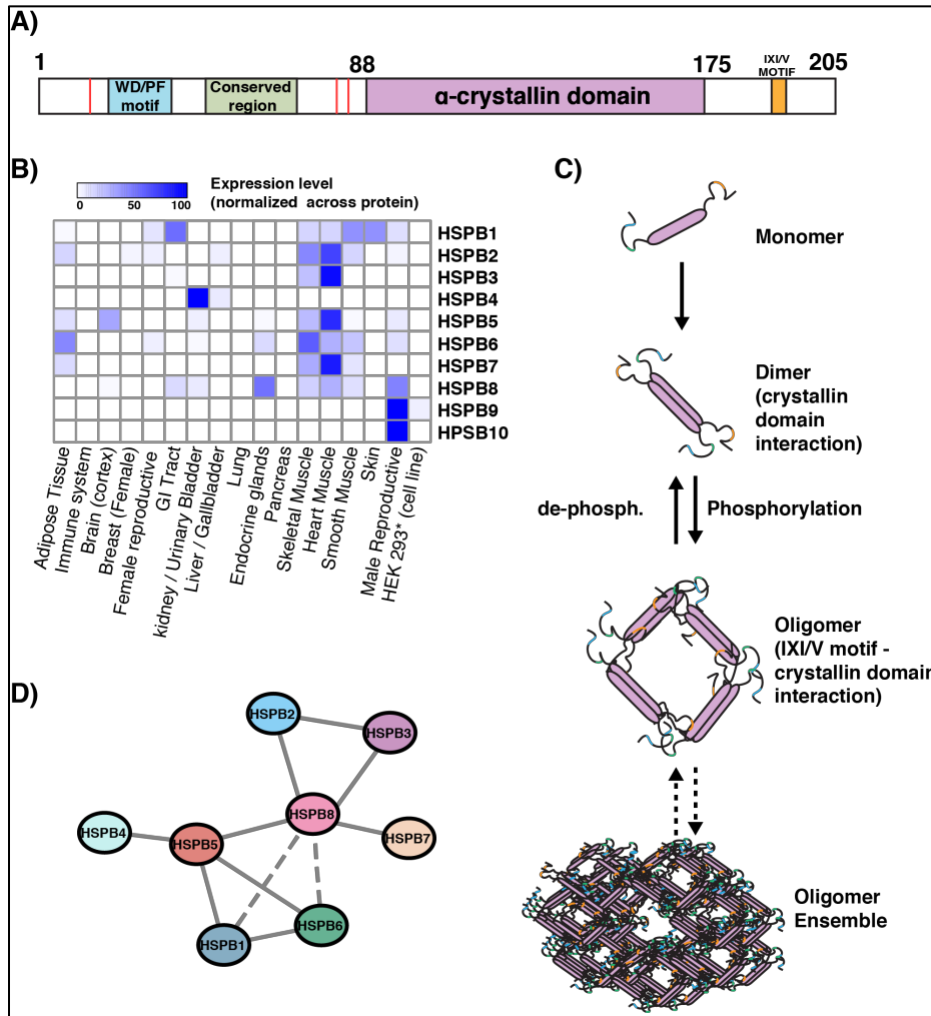
**Figure 5.3** - APMS identifies putative protein interaction partners for the BAG family of co-chaperones in HEK293T. (A) Putative PPIs for the members of the family. The size of the dots represents the relative abundance of the prey in this bait compared to the rest, while the color represents the average spectral counts minus the counts in the controls. The rim represents the BFDR cutoff. Proteins can be significant with low counts or not significant with higher relative counts. (B) Same as A, for the different isoforms of BAG1. (C) Network view of the interactions of the BAG family of proteins. In both (A) and (C), BAG1-S is used for BAG1. Dashed lines represent interactions present in the CORUM database<sup>192</sup>. With an asterisk(\*), those interactions of BAG5 that were recovered using a different normalization method (see text for details).

### 5.3.3 Interactors of the small heat shock protein family provide new insights in the functional diversity of the co-chaperone network

The small heat shock protein family (sHSP henceforth) is a diverse group of proteins that exhibit an ATP-independent chaperone activity (i.e. they are not enzymatic, unlike the HSP70 family of chaperones) by binding to misfolded polypeptides and preventing or

reversing aggregation. The common feature of all the members of this family is an alpha-crystallin domain, which is frequently surrounded by an N-terminal region (mostly intrinsically disordered) and a C-terminal region<sup>261-264</sup>. The family contains 10 members in the human proteome (named HSPB1-10), although members of this family are found in high numbers in most eukaryotes, and many bacteria contain one or two copies. Indeed, sHSPs were already present in the last common ancestor of prokaryotes and eukaryotes<sup>263,264</sup>. Proteins in the sHSP family display a striking ability to oligomerize, binding with other sHSPs through their alpha-crystallin domain to form dimers that then can further bind other dimers to form oligomers (Figure 5.4). Composition and structure of the oligomers is expected to depend on the identity of the proteins forming them, and all the three regions of the protein are thought to be involved in this<sup>262,263,265</sup>. It is speculated that the identity of the client proteins for sHSP could depend on the composition of specific oligomers, although some reports also describe their activity as single units of homodimers<sup>265</sup>. In addition, phosphorylation of sHSP family members is known to have functional implications, mostly described as reducing the ability of the modified subunit to remain in the oligomer although other functions have also been described. Finally, sHSPs are present at different levels in different cell types (Figure 5.4), which probably accounts for the formation of different oligomeric forms that have tissue-specific functions<sup>262</sup>. A combination of the size of the family, their ability to form functionally distinct homo- and hetero- oligomers, the effect of posttranslational modification and the tissue specificity of some members allows us to draw a picture of the sHSP family as being able to bind to a wide range of targets, resulting in an enormous diversity of ubiquitous and tissue-specific functions (recently reviewed elsewhere<sup>265</sup>).

Here, we performed AP-MS on the 10 members of the human sHSP family, overexpressed in HEK-293T cells. In this cell type, only HSPB1 is expressed at reasonable levels (Human Protein Atlas RNA expression data<sup>29</sup>), which means only hetero-oligomers of the overexpressed bait with this protein can be captured (HSPB7, 8 and 9 seem to be expressed at very low levels, probably too low to bind to a significant amount of the massively overexpressed bait). Indeed, our data shows that all the sHSP family members except for HSPB9 and HSPB10 (as expected<sup>265</sup>) are able to pull down with HSPB1.



**Figure 5.4** – The small Heat Shock Protein (sHSP/HSPB) family. (A) Domain structure of HSPB1/HSP27 as a representative member. All the members of the human family have a very similar domain structure. (B) RNA expression of sHSPs members in different human tissues and in the HEK 293 cell line. Heat map intensities are normalized for each protein, across all tissues. Source: the Human Protein Atlas project<sup>29</sup>. \*: our pull-down experiments in HEK 293T suggest HSPB1 is expressed in a high enough level. (C) Polymerization of sHSPs. These proteins rarely exist as monomers, usually forming dimers by means of alpha-crystallin domain interactions. These dimers can form oligomers (regulated by phosphorylation in HSPB1), which at the same time can further group into larger ensembles. Oligomers can be from the same protein or different members (hetero-oligomers). Oligomerization behavior is potentially dependent on composition, as it is which stage is functionally active. (D) Known interactions between sHSPs. Data extracted from Arrigo<sup>265</sup>. Dashed line: conflicting evidence.

When looking at the list of putative interactors (Figure 5.5), we see that multiple members of the tubulin family (alpha and beta) are bound all the sHSP members except for HSPB10 (see below for comment on this). Small heat shock proteins are known to interact with the cytoskeleton<sup>262,266,267</sup>. We were surprised to not find a significant binding to actin in our dataset, as it has been reported by many before (reviewed in <sup>265</sup>). This could be due to insufficient phosphorylation of the sHSPs, as it has been required to be essential for actin interaction of HSPB1<sup>268</sup>. We also found that several members of the HSP70 family and associated co-factors co-precipitated with many of the sHSPs, most strongly with HSPB7 and HSPB9, while they were almost imperceptible in HSPB4 and HSPB5 pulldowns. Coincidentally, HSPB7 and HSPB9 have been described to have the strongest anti-polyQ aggregation activity of the whole family (referenced in <sup>261</sup>). This interaction could be dependant on the BAG co-chaperone, which has been described to bridge the sHSP and HSP70 families of chaperones<sup>140,141,269</sup>, although we did not see a correlation between the sHSPs that had stronger HSP70 and BAG3 spectral counts. We do see significant interaction of BAG3 with HSPB6, HSPB8 and HSPB10. The interaction with HSPB6 and 8 has been reported by others<sup>140,140</sup>. We also found two HSPH/HSP110 chaperones HSPA4/HSPH2 and HSPH1. Finally, we also observe a few of the subunits of the proteasome regulatory subunit. This interaction was stronger for HSPB10 but also present in others (HSPB4, 5 and 8). Indeed, a direct interaction of HSPB10 and HSPB5 with the proteasome has been reported by others<sup>262,270</sup>.

In addition to these functions that are shared by multiple hHSP family members and that generally represent their well-known role as chaperones, we also identified member-



specific interactions that provide new insights into how small heat shock proteins connect to other processes and functions in the cell. The considerations highlighted above on cell type-specific expression of sHSPs and partners should be taken into consideration here.

### HSPB1

No significant putative interactors were found for HSPB1 that were not found in other members of the sHSPB family. As for those shared with other members, nothing seems outstanding. No significant amount of BAG3 was recovered down with HSPB1, but a parallel pulldown using BAG3 as a bait revealed HSPB1 as a clear interactor.

### HSPB2

HSPB2 has been described to localize in cytosolic granules, and to be associated with the membrane components of the mitochondria under stress (<sup>271</sup>). In our study, the mitochondrial proteins ATAD3A and COA7 pulled down with HSPB2, the latter with high spectral counts. In addition, we find an association with the microtubule-associated protein CLASP1, with one ribosomal protein and with the CC2D1A protein, transcription factor also known to be involved in mitotic spindle regulation.

### HSPB3-6

The list of specific putative interactors for HSPB3, HSPB4, HSPB5 and HSPB6 contains proteins of diverse origins without a clear pathway/localization that could be used as substantial evidence of a novel function for these proteins. That being said, there are some

strong, specific hits that suggest potentially interesting associations (e.g. HSPB3 and HCFC1, CRYAB and EIF3D).

### HSPB7

HSPB7 is a small heat shock protein that has been described to have very strong anti-aggregation activity in an HSP70- and proteasome- independent manner, and to not be competent for luciferase refolding, unlike other sHSPs<sup>272</sup>. Our data identifies two clear sets of proteins that interact specifically with HSPB1. Filamins A, B, and C. This is in agreement with a recent study that described HSPB7 interacting with Filamin C and preventing its aggregation toxicity in a chaperone-independent fashion<sup>273</sup>. The second group of proteins are the two subunits of the mitochondrial-processing peptidase (MPP), which is responsible for preprocessing of nuclear-encoded mitochondrial proteins as they enter the mitochondria<sup>274</sup>. This raises the hypothesis that HSPB7 could be involved in the transport or preprocessing of mitochondrial proteins, or in the assembly of the functional MPP heterodimer.

### HSPB8

The list of significant putative interactors specific for HSPB8 was very extensive in our study. The only two complexes identified there were 14-3-3 adapter protein and some of the serine/threonine-protein phosphatase 2A (PP2A) subunits, known to act on microtubule associated proteins, along with their phosphatase target RAF1. A GO enrichment analysis of the rest of the HSPB8 proteins revealed a mild enrichment for phosphoproteins and mitochondrial proteins.

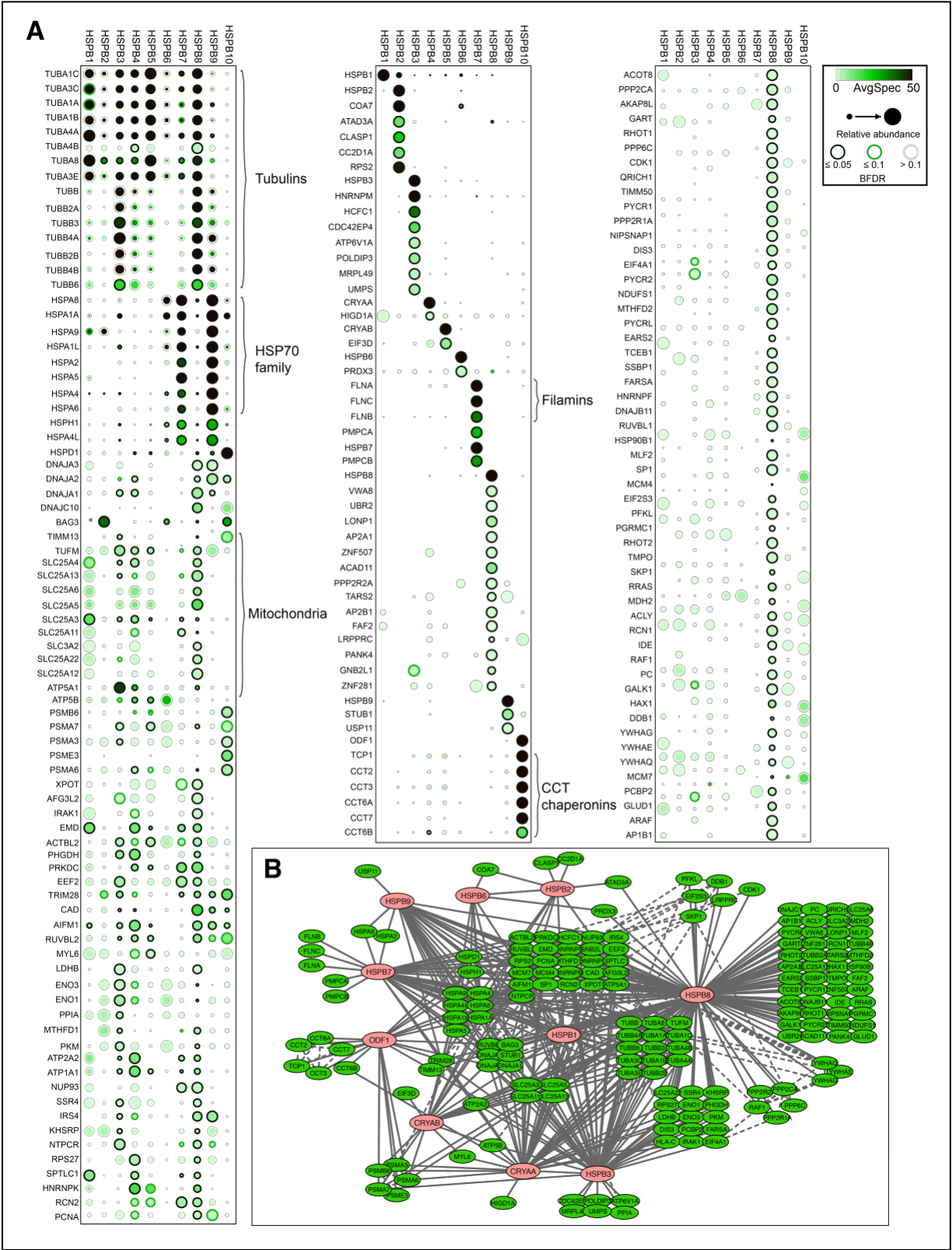
### HPB9

HSPB9 is, along with HSPB7 the member of the sHSP family that showed a stronger binding to HSP70 in our dataset, and also shares with HSPB7 a strong anti-polyQ aggregation activity<sup>261</sup>. In addition, we found it to interact with the ubiquitin protein ligase STUB1/CHIP, and also with the de-ubiquitinase USP11. HSPB9 is known to be specific of testis tissue (more specifically, in the nuclei of spermatogenic cells in later stages<sup>275,276</sup>). The strong interaction with HSP70 and both a ubiquitin ligase and a de-ubiquitinase raises the hypothesis that HSPB9 may be involved in testis-specific protein quality control, refolding and disposal through the proteasome (or protection from it).

### HSPB10

HSPB10 (also known as ODF1) is only found in the outer dense fiber that surrounds the axoneme of the sperm cell, where it is believed to be involved in the maintenance of the recoil and elasticity of the sperm tail<sup>277,278</sup>. Even though the cell line we used in our experiments does not express any sperm specific proteins, we found that the HSPB10 putative interactome presented a few interesting features. First, the lack of coprecipitation with tubulins is striking, given that the axoneme contains microtubules. This could be due to HSPB10 having specialized to not bind to tubulin as a way to prevent undesired interactions with the axoneme core, or that it is adapted to bind only to sperm-specific tubulin variants<sup>279</sup>. Also interesting is the fact that HSPB10 pulled down very strongly most of the subunits of the CCT chaperonin, including CCT6B which is highly enriched in testis but is expressed at low levels in the HEK293T cell line (data from human protein atlas<sup>29</sup>). Finally, HSP10 is the member of the

sHSP family that pulled down HSPD1 with the highest intensity. HSPD1 is a chaperonin located in the mitochondria. The outer dense fiber is surrounded by a mitochondrial sheath, so it could be that HSP10 is interacting with the chaperones inside or surrounding that structure. Both CCT and HSPD1 are known to be highly expressed in the male gametes and have also been found in the surface of the sperm cell, where they get translocated to play a role in the recognition of the zona pellucida<sup>280,281</sup>.



**Figure 5.5** - APMS identifies putative protein interaction partners for the small Heat Shock Protein family in a HEK293T overexpression system. (A) Putative PPIs for the 10 members of the family. Interpretation of the graph should be as in Fig. 3.3. Asterisk (\*): a significant interaction between HSPB1 and BAG3 was observed in a parallel AP-MS experiment we performed. (B) Network view of the interactions of the BAG family of proteins. Only BAG1-S is used for BAG1. Dashed lines represent interactions present in the high confidence database CORUM<sup>192</sup>. For simplicity, only those proteins with BFDR 0.05 are shown in the network view, and in the dot plot only those with at least one bait interaction with BFDR<0.1. With an asterisk(\*), those interactions of BAG5 that were recovered using a different normalization method.

#### **5.3.4 AP-MS study of BAG3 interactors for different disease-causing mutations in an overexpression system on immortalized cells**

Variants in the BAG3 gene have been associated with early-onset dilated cardiomyopathy and muscular myopathy (for more information on this, see<sup>17,23,282,283</sup>and Introduction chapter of this document). These variants include not only loss-of-function but also protein coding variants. This raises the interesting hypothesis that disease isn't caused by a lack of BAG3 protein as a whole, but rather by the perturbation of a specific set of BAG3 interactions<sup>8,9</sup>. Although a few studies have addressed the implications of specific point mutations, this has been mostly based on prior knowledge on direct BAG3 interactions with other members of the chaperone network<sup>45,60</sup>. An unbiased comparison of protein protein interactors for 'healthy' BAG3 *versus* BAG3 disease related variants could potentially provide more knowledge on currently unknown BAG3 interactions. AP-MS could provide such information. In addition, AP-MS can report on indirect interactions that could reveal specific consequences of the specific amino acid transitions beyond an interaction with a chaperone (e.g. if a point mutation affects BAG3-HSP70 interaction, what is the HSP70 interaction that gets affected? Is any client protein not binding to HSP70?).

To address whether such a *differential AP-MS* approach would be able to reveal differences in BAG3 binding partners for different variants of BAG3, we investigated the interactions of 9 different versions of BAG3 when overexpressed in HEK293T cells. This cell type and expression system recapitulates poorly the function and proteome of the cells affected in BAG3-related heart disease (an extended discussion on this can be found in Chapter 2). However, on the other hand it provides an easy platform for the proof-of-concept of the differential AP-MS approach, allowing us to compare multiple variants in parallel. To get additional information on how different regions of BAG3 contribute to the binding to specific putative partners, we also performed APMS on a truncated form of BAG3 missing the BAG domain (amino acids 1-420) and another form only containing the BAG domain (amino acids 399-504).

The results (Fig. 3.6) show that we were able to capture differences between different variants. An analysis of the interactors of the non disease-related variant of BAG3 can be found earlier in the text, along with the other members of the BAG family of chaperones.

The two DCM-causing mutations located in the BAG domain (E455K and L462P) abolished or dramatically reduced the interaction with the HSP70 chaperones and also co-chaperones (DNAJs) and other associated factors (STUB1, MFL2). The implication of this interaction in the development of heart disease was suggested by the higher concentration of disease-causing variants in the BAG domain area, and it also has been explored by others<sup>45</sup>. We confirmed the impact of one of these mutations (E455K) on the chaperone activity inside the cell using luciferase refolding assay (Figure 5.6). Indeed, these two mutations had an almost

identical interaction profile than the BAG3\_ΔBAG truncation variant. Interestingly, these three BAG3 variants also lost their interaction with the also BAG domain-containing protein BAG2, signaling proteins IRS4 and SNW1, and other proteins. This suggests that the BAG3-HSP70 complex is functionally involved with these proteins, potentially targeting them for folding. Interestingly, the L462P variant showed a strong co-precipitation with different members of the chaperone family HSP110. This family is structurally related to HSP70 but known to be involved in different functions and have a slightly different chaperone mechanism<sup>257</sup>. The intensity and specificity of this interaction suggests that maybe the L462 residue is playing an important role in HSP70 binding, and that when switched for proline the affinity of binding tilts it towards the HSP110 family, probably through a family-specific domain or region. The E455K variant seemed to co-precipitate with subunits of the ribosome in an interaction of unknown consequences or origin.

The R218W had been described as pathogenic in a study on a Japanese cohort of patients<sup>35</sup> and in the same study it also showed mislocalization when overexpressed in rat cardiomyocytes. However, this variant showed an interaction profile that was almost indistinguishable from the control, not disease-related variants of BAG3.

The C151R variant does not localize in any known domain of BAG3. A genome wide study associated it with a lower incidence of DCM<sup>18</sup>. Our data showed no differences with the control variants. This could potentially be due to the function of this variant being heart-specific. The P407L variant represents a special case. It hasn't been directly linked to disease, but the same study that identified C151R also found that patients with C151R always had a



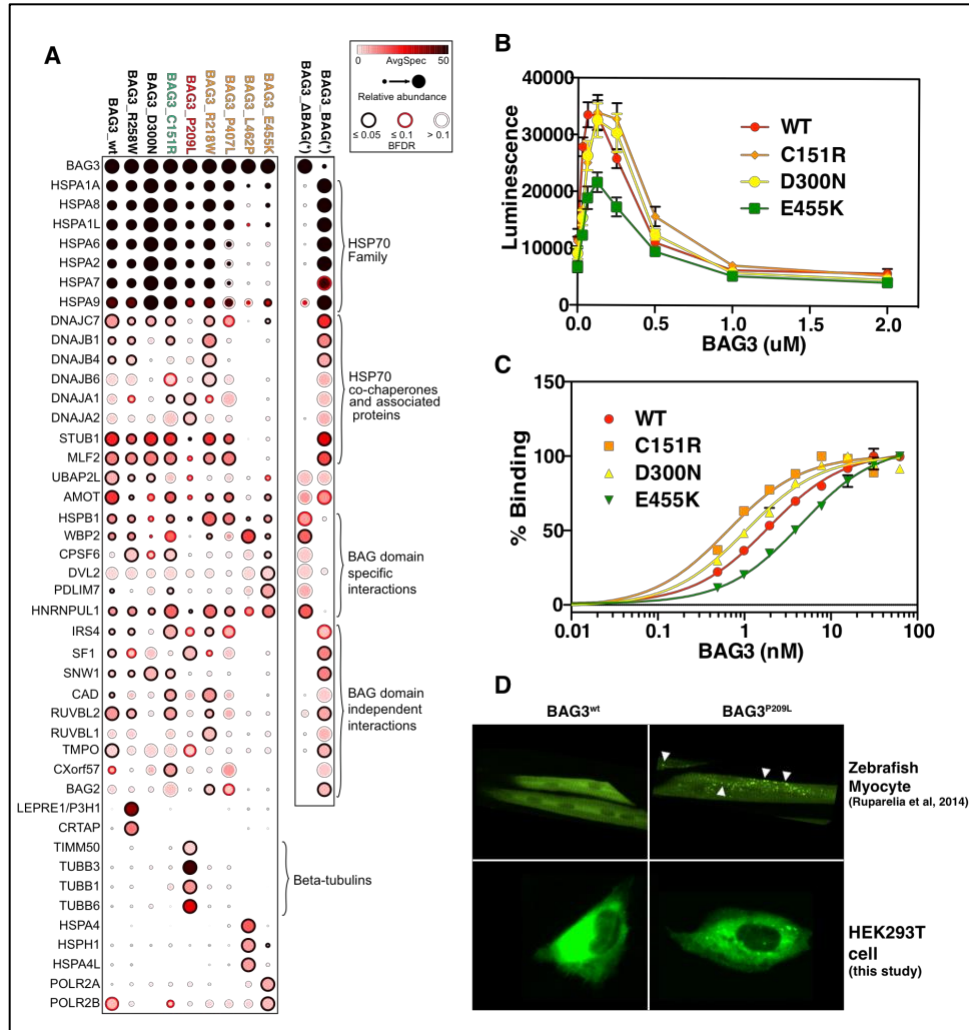
proline at this location, suggesting that maybe this residue played a role in pathogenesis and that patients with a proline (the most common allele) had reduced risk compared to those with a leucine. Our data only showed a small difference when compared to control variants, mostly in the form of lower HSP70-associated interactions. This could indicate that this variant only mildly affects the BAG3 interactions in a similar way to the BAG domain variants, but further studies would be needed to confirm this hypothesis.

The two variants chosen as additional controls for not being known to cause disease (R258W and D300N) showed no dramatic differences from the common form of BAG3 (“BAG3\_wt”). However, it is noteworthy that the R258W variant associated strongly with proteins P3H1/LEPRE1 and CRTAP, known to interact with each other for hydroxylation of collagen<sup>284</sup>. The functional relevance of this is unknown and outside the scope of our study.

The last variant, P209L, has been associated with skeletal myofibrillar myopathy and giant axonal neuropathy<sup>25,27,285</sup>. This variant localizes exactly in one of the IPV motifs that are associated with interaction with small heat shock proteins<sup>140</sup>. This more aggressive disease phenotype has been speculated to be a gain-of-function<sup>60</sup>. Analysis of the APMS results for this variant revealed that it does not result in a dramatic reduction of interactions (although intensity of HSP70s and associated factors and with HSPB1 is mildly reduced). However, a specific feature of this variant was the co-precipitation of beta-tubulins and, with less intensity, the mitochondrial translocase TIMM50. This association with tubulins could be a consequence of the aggregation of the P209L variant, which has been described by others<sup>60</sup>. In fact, when

we performed staining for BAG3 in HEK293T cells overexpressing P209L we did see some cytoplasmic aggregates that agree with previous studies (Fig.3.6).

When comparing the two truncation variants, we observe that the BAG3\_ΔBAG variant is still able to bind a number of factors not associated with the BAG domain. These include HSPB1 (predicted to use the IPV motifs), WBP2 (predicted to use the WW domain), DVL2 or PDLIM7. AP-MS of the BAG domain only was unable to recover these. As mentioned earlier, the BAG3\_ΔBAG is unable to bind the same partners than the E455K and L462P variants.



**Figure 5.6** - APMS is able to capture differences between variants of the BAG3 protein. (A) Comparison of the putative interaction partners for each one of the variants, along with information on the spectral counts obtained. Dot size and color and rim color should be interpreted as in Figure 5.3. Note that for all of the variants an additional step of normalization by the BAG3 spectral counts was performed. This normalization was not performed for the truncation variants, and the dot size represents relative intensity only in between these two. The bottom part of the plot has been removed for these two as it contained only low intensity, non significant values. (B) Luciferase refolding assay for some of the BAG3 variants. (C) Binding affinity of BAG3 variants and HSPA1A (HSP70), as quantified by Flow Cytometry Protein Interaction Assay. (D) Comparison of immunofluorescence images of BAG3<sup>wt</sup> and BAG3<sup>P209L</sup> in both this study and Ruparella et al. <sup>60</sup>. We observe the formation of BAG3<sup>P209L</sup> aggregates in the cytoplasm of HEK293T cells, similar to what they observe in zebrafish myocytes.

# Conclusion

In this section, we explored the use of affinity purification coupled to mass spectrometry for the analysis of interactors and complexes of different proteins involved in the protein quality control network. We focused on comparisons between the members of two protein families: the BAG family of HSP70 co-chaperones and the sHSP family of ATP independent chaperones. The combination of a relatively high replicate number ( $n=5$ ) plus rigorous quality control of samples and the use of spectral counts as a quantitative measure of protein abundance allowed us to obtain quantitative differences between controls and true samples, and also in between different family members. This is especially important when dealing with chaperones, as many of these proteins interact nonspecifically with the affinity matrix or simply aid in the folding of the expressed baits, without adding information on the function of the protein. As an example, in our dataset all the pulldown experiments co-precipitated HSP70 proteins, but by using quantitative analysis we were able to tell specific, functionally relevant interactors from those nonspecific.

Our results validate many of the well known interactions that are shared by most or all members of these families and represent the ‘canonical’ functions that are associated to them. However, we also identified a number of member-specific putative interactions and processes, that provide important clues on the functional diversity of protein chaperone families. Although some functions have been described for specific members of these families, this has mostly been done through educated guesses from disease or knockout phenotypes and focusing on a specific member only. In this study we perform an unbiased, high-throughput

analysis of putative interactions, using all the members of the family in a parallel side-to-side comparison (maybe move this to intro).

In addition, we performed a comparison of putative interactors for a set of disease-associated variants of the BAG3 co-chaperone, and found that APMS was able to capture differences in their putative interactions. This validates this approach as a good platform for the study of the impact of protein coding variation in specific proteins.

Some limitations of this study need to be taken into account when looking at the data. First and foremost, proteins are being expressed at levels substantially higher to their normal expression levels. There is plenty of evidence that chaperones function is highly dose-sensitive. As an example, BAG family of chaperones has been described to inhibit HSP70 activity if expressed at high levels, and only work as effective catalyzers of the folding reaction when expressed at low levels (see for example <sup>32</sup>). Also relevant is our choice of cell line to perform these studies. Many of the studied proteins are enriched or exclusively expressed in a subset of tissues, which points to their role in cell type-specific processes. The immortalized cell line HEK293T we used here expresses a proteome that is poorly representative of many of these tissues. For example, the BAG3 disease variants are known to cause a phenotype in the heart muscle, which suggests that most of the disease-relevant interactions are specific to myocytes. This tissue specificity is also especially important for the sHSP family, for which only 1 out of 10 members (HSPB1) is expressed ubiquitously<sup>261</sup> and at good levels in HEK293T (Figure 5.5). Since sHSPs are known to form hetero-oligomers, we were not able to identify any function derived from the interaction of a specific sHSP with another family member other

than HSPB1. A more detailed discussion on the impact of protein overexpression and cell type used for APMS experiments can be found in Chapter 2. Last but not least, our protocol for APMS is tailored towards the identification of strongly interacting preys, and will miss most of the transient interactors. For chaperones, this means that we expect most of the folding clients to not be identified. On the other hand, this allowed us to identify high-confidence strong interactions of these chaperone families.

In sum, we identified high-confidence strong interactions for different protein chaperones and variants. Our results provide clues on the functional diversity of the protein homeostasis network, and allow us to draw a picture where specific members of protein chaperone families perform distinct functions, rather than all of them being assigned a generic role in the cell. This has been enabled by the parallel, comparative and unbiased analysis of interactions through APMS. In addition, we also describe different interaction profiles for different variants for the same protein. Taken altogether, the results of this study present APMS as a powerful technique for the differential analysis of protein interactions in the proteostasis network, providing interesting insights that can be exploited to improve our understanding of the basic functioning of the cell. Since malfunction in the protein quality control system plays an important role in multiple diseases, the information obtained could also be used to direct the design of therapeutic strategies<sup>212</sup>.

# Chapter 6 - Comparative phenotypic and protein interaction analysis of disease-associated BAG3 variants in iPS-derived cardiomyocytes

## 6.1 Introduction

As I have established in previous chapters, the study of protein-protein interactions can help us understand the function of proteins and gain insight on how genetic variation affects the development of disease. In Chapter 5 we used a “differential APMS” approach to compare different variants for the DCM-related protein BAG3, and successfully identified interacting partners that bind differentially to each variant. However, as I established in Chapter 2 if we want to get insight on how protein variants cause disease in a specific tissue, we need to study their interactions in the context of the cell type affected by disease. This is the case for BAG3, which is expressed in many tissues<sup>184</sup> but for which variants have been associated primarily to heart disease (for a more comprehensive description on this topic, see Introduction and Chapter 5).

Stem cells and genome engineering allow us to address this issue by providing us with a source of heart cells which would bear specific variants of interest. In Chapter 4, I generated a series of cell lines that express selected variants of BAG3 fused to an epitope tag that allows

for APMS. These cell lines have a virtually identical genome, except for the induced modifications. This allows for the dissection of mechanisms minimizing confounding variables.

In this Chapter I present the use of these cell lines for APMS, in an effort to identify cardiac-specific interactors that are modified by disease-related BAG3 variants. I focus mostly in one variant that is most representative of the dilated cardiomyopathy phenotype (BAG3<sup>E455K</sup>) and compare its interactions to those of BAG3<sup>wt</sup>. In addition, I use these cell lines for the phenotypic assay that we described in Chapter 3, where we identify a phenotype for BAG3 insufficiency when iPS-derived heart cells are under proteotoxic stress.

## 6.2 Materials and Methods

### 6.2.1 Differentiation of iPS into iPS-CM and iPS-CM cell culture

The method for differentiation of iPS cells into iPS-CM and further enrichment has been described in Chapter 3. For maintenance, iPS-CM cells were fed with RPMI 1640 (Gibco) supplemented with B-27 Supplement (Thermo Fisher Scientific) (RPMI/B27 henceforth) twice a week. For harvesting, cells were washed once with PBS, and incubated with Trypsin 0.25% (Thermo Fisher Scientific) at 37C until cells detached. After pipetting for cell singularizing, trypsin was quenched with EB-20 media (20% Fetal Bovine Serum (Hyclone; Thermo Fisher Scientific), 1x GlutaMax-I (Life Technologies), 1x MEM Non Essential Amino acids (Gibco), 1/10<sup>6</sup> beta-mercaptoethanol) and cells were centrifuged down. For replating, cells were harvested, counted, and plated as appropriate. For freezing, iPS-CMs were



resuspended in either freezing media (10% DMSO in 90% FBS, supplemented with Rock Inhibitor Y-27632 (10uM, Selleckchem)) or CryoStor Freeze Media (BioLife Solutions) and frozen down in a Mr. Frosty Freezing Container (Thermo Fisher Scientific) before transferring to liquid nitrogen storage. For thawing iPS-CMs, frozen vials were put in a 37C water bath until mostly melt and cells were centrifuged to remove freezing media and resuspended on RPMI/B27 with Rock Inhibitor before plating. All cells used in these experiments were frozen down on day 30 (30 days after addition of CHIR IWP2 inhibitor), and they were thawed in parallel for the different experiments.

### **6.2.2 Flow cytometry**

Flow cytometry was performed as previously<sup>164</sup> described using mouse monoclonal antibody against cardiac troponin-T (clone 13-11, Thermo Fisher Scientific).

### **6.2.3 Bortezomib treatment. Contractility and viability assay**

These experiments were performed as described in Chapter 3. Briefly, day 30 frozen cardiomyocytes were thawed in 6-well plates, and maintained in culture for 10 days. Then they were harvested, counted and seeded on a 96wp at density of  $2e^4$  cells per well. Seven days later, the cells were added bortezomib (Cell Signalling Technologies) at different concentrations. Cells were kept in bortezomib for 48 hours and then the cells were washed and added fresh media without drug. All the cell line/dose combinations were done in triplicate wells, in parallel.

For the contractility assay, a Pulse instrument (Cellogy) was used. Timepoints were taken before adding the drug (baseline), at 24 hours and at 48 hours. Normalized contractility

index was calculated by dividing the contraction peak height at each timepoint by the baseline of each well.

For viability assay, a resazurin-based method was used. Cells were allowed to recover from bortezomib treatment for 5 days, and then media was replaced with 10% PrestoBlue reagent (Thermo Fisher Scientific) in RPMI/B27 and incubated for 1 hour at 37C. Fluorescence intensity was measured with a SpectraMax i3 plate reader (Molecular Devices). Media only wells were used to subtract baseline intensity from all samples. To calculate viability for each well, fluorescence was normalized to control wells treated with no drug.

#### **6.2.4 Affinity Purification coupled to Mass Spectrometry analysis**

Cells were harvested by scraping on ice-cold PBS and pelleted down by centrifugation. Cell pellets were washed twice with ice-cold PBS and flash frozen in dry ice with ethanol.

For iPS-CM samples, day 30 differentiations were used. For each differentiation batch, a separate tissue culture well or plate was used for cell counting, quantification of differentiation efficiency and genotype quality control. Pellets were pooled before affinity purification to the amount of 25-30 million cells per sample. For the Bortezomib treatment condition, bortezomib was added to the media to 100uM final concentration 24 hours prior to the harvesting (on day 29). For iPS cell APMS, three ~90% confluent 15cm dishes were used per sample.

As a negative control for nonspecific binding, WTc iPS or iPS-CM cells not expressing any 3xFLAG affinity epitope were used. All the experimental conditions were performed in

four replicates, each one on a different day. For each replicate all the different conditions were run in parallel. For Mass Spectrometry, all of the samples were run sequentially.

Cell pellets were thoroughly resuspended on lysis buffer (0.1% NP-40, 300 mM NaCl, 20% Glycerol, 2 mM MgCl<sub>2</sub>, 0.5 mM EDTA, 0.5 mM EGTA, 1 mM PMSF, 1 mM DDT in 50 mM HEPES-NaOH pH 8.0, supplemented with complete protease inhibitor cocktail (Roche) and Benzonase Nuclease (50 U/ml, Sigma-Aldrich)). Cells were lysed by four freeze-thaw cycles, followed by incubation for 20 min at 4°C in constant rotation. After clearing lysates by centrifugation, protein extracts were diluted (in 0.5mM EDTA, 0.5mM EGTA, 1mM PMSF in 50mM HEPES-NaOH pH 8.0) three-fold to reduce salt content and incubated with 30 ul anti-FLAG M2 Magnetic Beads slurry (Sigma-Aldrich) for 2-3 hours. Beads were then rinsed in wash buffer (3x washes in 0.01% NP-40, 1 mM PMSF, 0.5% EDTA, 0.5% EGTA in 50 mM HEPES, pH 8.0; 1x wash on buffer without NP-40). Beads loaded with FLAG-enriched proteins were reduced (5 mM TCEP), alkylated (15 mM iodoacetamide), and digested with 1%(w/v) trypsin overnight. Resulting peptides were desalted by OMIX C18 desalting tips (Agilent) following protocol by provider and dried on a speed-vac.

Peptides were resuspended in 0.1% formic acid and analyzed by liquid chromatography tandem mass spectrometry (LC MS/MS) with an Easy-nLC 1000 (Thermo Fisher, San Jose, CA) coupled to an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA). Online LC separation was carried out using a 75 µm x 25 cm fused silica IntregraFrit capillary column (New Objective, Woburn, MA) packed in-house with 1.9 µm Reprisil-Pur C18 AQ reverse-phase resin (Dr. Maisch-GmbH). Peptides were eluted at a flowrate of 300

nL/min using a linear gradient of 5–30% B in 45 min, and 30–95% B for 25 min (mobile phase buffer A: 100 H<sub>2</sub>O/0.1% formic acid; mobile phase buffer B: 100%ACN/0.1% formic acid). Survey scans of peptide precursors from 400 to 1600 m/z were performed at 120K resolution in the Orbitrap, with an AGC target of  $2 \times 10^5$ , and a maximum injection time of 100 ms. Tandem MS (MS<sub>2</sub>) was performed by isolation with the quadrupole, HCD fragmentation with normalized collision energy of 30%, and rapid scan MS analysis in the ion trap. The MS<sub>2</sub> ion count target was set to 104 and the max injection time was 35 ms. Precursors with charge state 2–7 were sampled for MS<sub>2</sub> and dynamically excluded for 20 s (tolerance of 10 ppm). Monoisotopic precursor selection was turned on, and the instrument was run in top speed mode with 3-s cycles.

### **6.2.5 AP-MS Data Analysis**

For protein identification and quantification, MaxQuant software v1.5.3.30 was used<sup>166</sup>. Tandem mass spectrometry (MS/MS) spectra were searched against the November 2016 release of the UniProt complete human proteome sequence database, modified to include the FLAG peptide sequence. MaxQuant was run on default parameters, allowing for 2 maximum missed cleavages, with a first search peptide tolerance of 20 ppm and a main search peptide tolerance of 4.5 ppm. Methionine oxidation and N-terminal acetylation were set as variable modifications, and carbamidomethylation of cysteines as fixed modification. The ‘match between runs’ setting was activated (window of 7 min) to improve peptide identification.

For the analysis of data, proteins with two or fewer peptides identified were discarded, as were typical common contaminant proteins (downloaded from <http://maxquant.org/>). To compare the interactors between BAG3 variants, a two-step analysis was performed using SAINTq<sup>204</sup>. First, a pool of putative BAG3 interactors was obtained by running the algorithm using all the BAG3 variants (in this case, BAG3<sup>WT</sup> and BAG3<sup>E455K</sup>) versus the negative control (WTc cardiomyocytes), and using a Bayesian False Discovery Rate (BFDR) of 10% to filter for significant hits. Then, SAINTq was run again for each one of the two BAG3 variants, using the other one as a control. Only those hits that were significantly present in one variant (10% BFDR) and in another and that were present in the list of interactors for the first step were used. Additionally, we also filtered the list to only consider those hits with a fold change over (average intensity ratios) of 4. We found that, although potentially very conservative, this two-step SAINTq approach was the one that best addressed the high amount of missing data (inherent to Mass Spectrometry analyses).

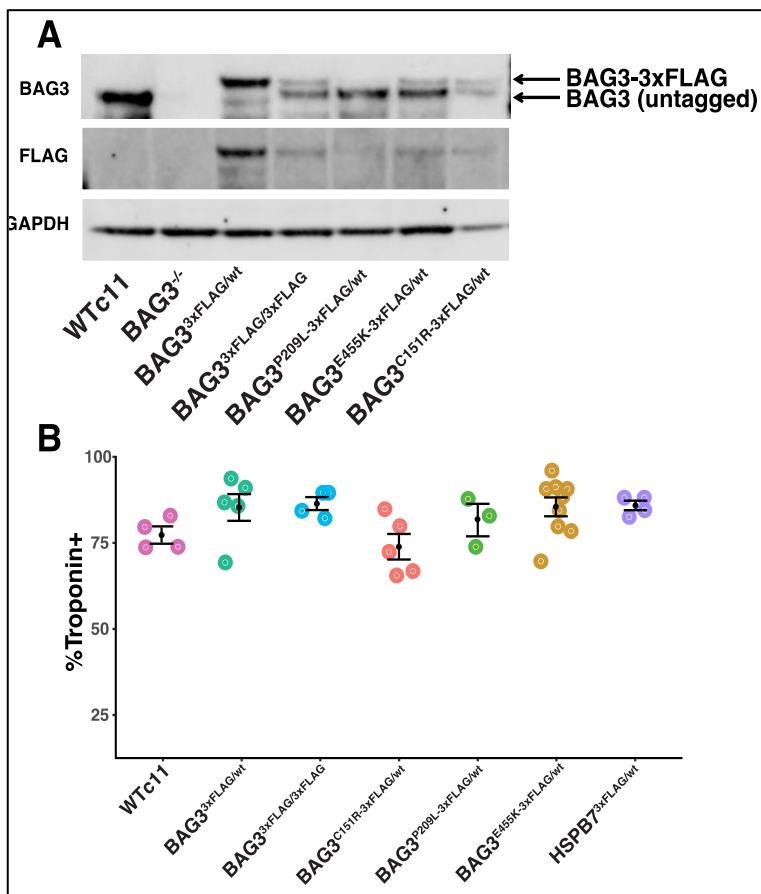
### 6.2.6 Western blotting

An identical procedure to what is described in section 4.2.12 was used.

## 6.3 Results

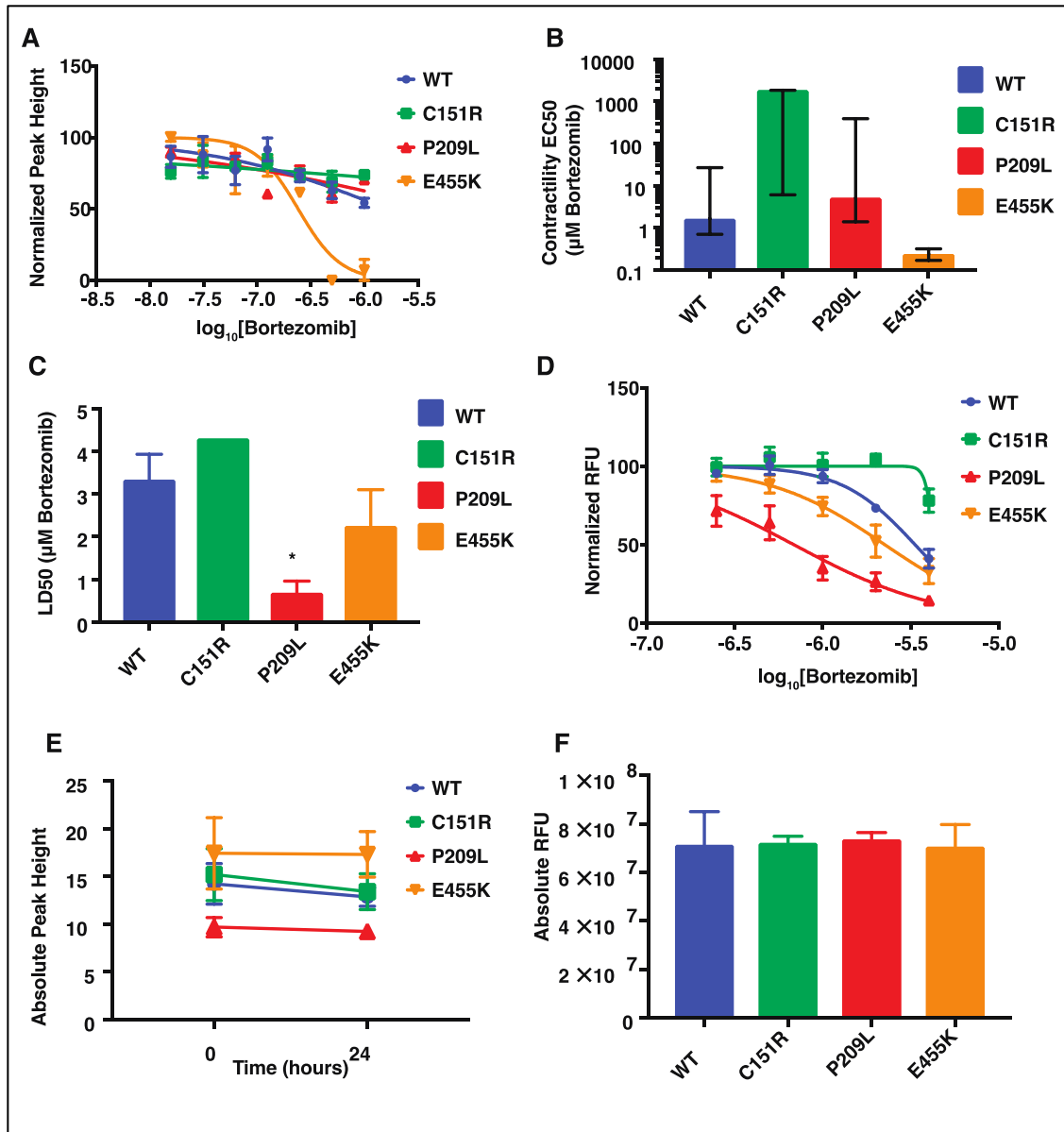
In order to perform the experiments for this chapter, we differentiated the different iPS cell lines into cardiomyocytes using a well characterized protocol<sup>286</sup> and we further purified cardiomyocyte cells<sup>163</sup> to remove other cell types that could affect the downstream analyses. We saw that differentiated cardiomyocytes still express the BAG3-3xFLAG protein fusion

(Figure 6.1a) at similar levels, except for the BAG3<sup>P209L</sup> variant which seems to produce significantly less of the BAG3<sup>3xFLAG</sup> protein. This lack of protein is mostly due to decreased expression of the allele expressing the P209L variant, as we were able to see equal levels of expression in both BAG3 alleles in WTc (original line) and in a cell line derived from a patient bearing the P209L variant (data not shown). Once differentiated and enriched, cardiomyocyte cultures contained a similar level of cardiac troponin-T positive cells (>80%) (Figure 6.2b). We also observed that in general no cells would grow in the cardiomyocyte cultures during the week after enrichment and before harvesting.



**Figure 6.1** - Cardiomyocytes differentiated from BAG3 isogenic lines express BAG3-3xFLAG and display similar differentiation efficiencies. (A) Western blot on iPSc-derived cardiomyocytes from the different BAG3-3xFLAG cell lines used. (B) Percentage of cardiac troponin positive cells after differentiation and lactate enrichment for each one of the cell lines used in this study. Each point represents a separate differentiation batch.

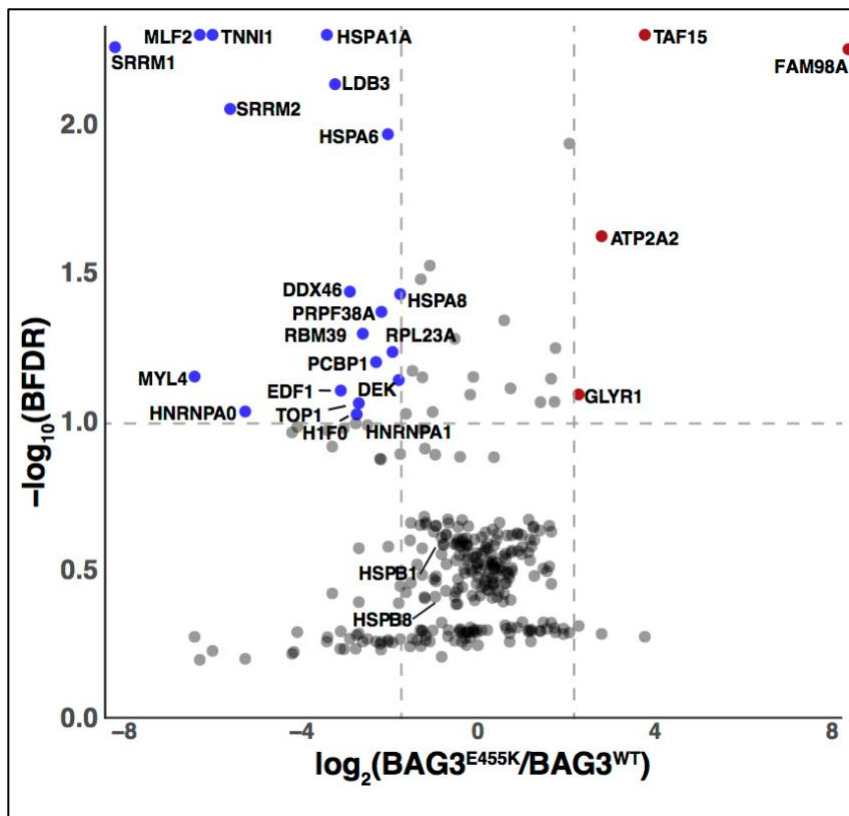
Next, we decided to study whether the different cell lines responded differently to proteotoxic stress, similar to what we described in Chapter 3 (and we published in <sup>30</sup>) for the BAG3<sup>-/-</sup> and BAG3<sup>wt/-</sup> cell lines. We saw that contractility of the cells expressing the BAG3<sup>E455K</sup> variant was significantly more affected by bortezomib than for the other lines, presenting a similar phenotype to what was described for the BAG3 loss of function in Chapter 3 (Figure 6.2a, b). These cells were able to recover after washing the bortezomib with a non-significant loss on viability. On the other hand, BAG3<sup>P209L</sup> cells presented more cell death under bortezomib, as measured by resazurin assay after 5 days recovery (Figure 6.2c, d). The BAG3<sup>C151R</sup> variant showed no significant changes in contractility or viability. However, this was mostly due to the difficulty of calculating EC50 and LD50 as the values did not adjust well to a decay curve. All the cell lines were seeded at very similar densities (Figure 6.2f) and the control wells treated only with vehicle did not change their peak height significantly during the time of the experiment (Figure 6.2e).



**Figure 6.2** – Heart disease-associated cell Bag3 gene variants display a different response to proteotoxic stress. Cells were treated with different concentrations of bortezomib and video was acquired for contractility analysis after 24 hours. (A) Peak height (normalized to untreated) for different concentrations. (B) EC50 values for Bortezomib dose that reduces the peak height by half. (C) Cell viability LD50 for cells treated with bortezomib for 48 hours and allowed to recover for 5 days. (D) Cell viability for different bortezomib concentrations. (E) Absolute contraction peak height for the untreated wells. (F) Amount of cells for each well in this experiment. Note that bars for BAG3<sup>C151R</sup> in (B) and (C) are unusually large because of the lack of curvature of the data.



For elucidation of interaction partners for different BAG3 variants, we performed APMS on cardiomyocytes from each one of the cell lines. When comparing the BAG3<sup>E455K</sup> and BAG3<sup>wt</sup> variants, we found that the E455K version binds less strongly to multiple proteins, which include chaperones of the HSP70 family (and associated proteins) and a number of sarcomeric and muscle specific proteins, along with other proteins associated with RNA processing. Other two well-known binding partners, HSPB1 and HSPB8, co-precipitates equally with both of these variants.



**Figure 6.3.** – BAG3 interaction partners that bind differentially for the BAG3<sup>E455K</sup> variant. As scored by SAINTq (see Materials and Methods Section). Cutoff for relevant interactors is set at >10% Bayesian False Discovery Rate and four-fold fold change in protein intensity. Right side of the graph: proteins that co-precipitate stronger with the BAG3<sup>E455K</sup> variant. Left side: proteins whose interaction is stronger in the BAG3<sup>wt</sup> variant than in the BAG3<sup>E455K</sup> one.

## 6.4 Discussion

In this chapter we performed a molecular and phenotypical comparison of cell lines bearing different variants of the BAG3 protein. Mutations in BAG3 have been described to

affect muscle disease development. While loss of BAG3 expression and multiple variants in the have been associated to early onset dilated cardiomyopathy (DCM), modifications in a specific amino acid have also been described to also cause myofibrillar myopathy. In addition, one common variant was also described in a GWAS study as potentially protecting from the development of DCM. A more comprehensive review of the studies describing the associations can be found in the Introduction chapter. Previously we used genome engineering to modify the endogenous copy of the BAG3 gene and produce cell lines that express a representative variant for each one of the clinical presentations (BAG3<sup>E455K</sup> for early onset DCM, BAG3<sup>P209L</sup> for myofibrillar myopathy, and BAG3<sup>C151R</sup> for the putative protective allele).

When we added bortezomib to cardiomyocytes derived from these cell lines, we observed different responses between cell lines (Figure 6.2). Bortezomib is known to induce accumulation of toxic protein species by inhibiting proteasomal activity. We described previously that BAG3 is required for prevention of cardiotoxicity from bortezomib<sup>30</sup>. Here, we observe that the cell line expressing the BAG3<sup>E455K</sup> protein variant has an impaired response to bortezomib, as indicated by the change in contraction peak height. The viability assay results reveal that the cells are able to recover from 2 days in bortezomib. This phenotype is similar to the one we observed in BAG3 heterozygous null cells (Chapter 3). The affected residue in BAG3 is predicted to lay in the interface of the BAG domain with the ATPase domain of HSP70 proteins, and our data supports that the BAG3<sup>E455K</sup> variant dramatically diminishes the interaction strength with this family of chaperones (Chapter 5 and Figure 6.3). This suggest that the interaction with HSP70 is required for the response to bortezomib, probably by boosting the refolding of protein aggregates or by promoting the autophagic

degradation of misfolded HSP70 substrates, both functions BAG3 is involved in<sup>33,58,141</sup>. The cells expressing the myofibrillar myopathy associated variant BAG3<sup>P209L</sup> did not present a significant decrease of their contractility upon bortezomib treatment. However, the survival of the cells a week after treatment was severely decreased, suggesting they are unable to recover from the stress induced by bortezomib and in contrast to the BAG3<sup>E455K</sup> variant. The BAG3<sup>P209L</sup> variant is located in the middle of a motif that mediates the interaction of BAG3 with small heat shock proteins. Beyond simply affecting this interaction, the specific BAG3<sup>P209L</sup> amino acid transition has been described to form proteotoxic aggregates that result in impaired proteostasis (in<sup>59,60</sup> and Kampinga, Gestwicki et al. unpublished data). Given that the cells do not seem to have a baseline loss in viability (Figure 6.2), it could be that the treatment with bortezomib triggers a protein aggregation episode that the cells do not manage to recover from in the long term.

We also found that the BAG3<sup>C151R</sup> did not show a significant change in EC50 or LD50 for contractility and viability upon treatment as compared to control cells. However, a closer look at the data seems to suggest that this variant could indeed have an improved response to bortezomib, especially for higher doses of the drug. The nucleotide change creating this change in sequence was the most significant hit in a genome wide association study of dilated cardiomyopathy patients<sup>18</sup>, but no functional relevance for this specific residue has been described previously. The only evidence in literature is that of a screening of noncanonical protease sites<sup>287</sup>, which identifies cysteine 151 of BAG3 as a target for cutting. This same cleavage site was later reported by Wiita et al<sup>288</sup> as a non-aspartic acid protease substrate site that cuts very fast upon bortezomib treatment. This, in combination to our data, raises the

interesting hypothesis that BAG3 gets cleaved upon bortezomib treatment, and that the C151R transition could be abolishing this protease site and producing a longer lasting version of the BAG3 chaperone.

Finally, we also performed a comparison of the protein binding partners of the different disease modifying BAG3 variants. We decided to focus mostly on the comparison between the BAG3<sup>WT</sup> variant and the BAG3<sup>E455K</sup> variant. A label-free, data dependent acquisition analysis of the proteins that co-precipitate with these variants in cardiomyocytes provides some interesting clues on potential mechanisms for BAG3 pathogenesis in the heart. As predicted, the BAG3<sup>E455K</sup> variant had substantially lower affinity for HSP70 chaperones (mostly HSPA1A and HSPA6) and other associated proteins (MLF2, for example). In addition, the small heat shock proteins HSPB1 and HSPB8 did not seem to change affinity between these two variants, as we expected given that BAG3 does not bind them through the BAG domain, where BAG3<sup>E455</sup> is located. In addition to that, we also see that BAG3<sup>E455K</sup> binds more weakly to a number of putative interaction partners that are muscle and/or heart-specific. These include DCM-related proteins such as LDB3/ZASP or MYL4, and the troponin form TNNI1 that is expressed in embryonic heart<sup>289</sup>. In addition, there are two members of a family of pre-mRNA processing that also bind more strongly the BAG3<sup>WT</sup> isoform. Although some reports have involved BAG3 in gene expression, mRNA processing is a not so well described function of BAG3. The limitations of the mass spectrometry approach used for quantification between samples allow us to only report with confidence very dramatic changes (here, we are setting our filter to look at changes 4-fold or higher). In the future the list of putative interactors should be used for the production of a transition

library that can be used for data independent acquisition to get more sensitive quantification of these proteins.

## 6.5 Conclusion

In this Chapter, we used engineered iPS cell lines expressing different variants of a heart disease-related gene to study their impact on the response of the cells to proteotoxic stress and on interactions of the BAG3 protein. Overall, these data show that different variants of BAG3 associated with distinct clinical presentations also have a different response to bortezomib stress. In future studies other mutants of BAG3 can be analyzed using these methods, providing further insights on BAG3 biology. In addition, a comparison of the protein binding partners of two representative variants of the protein reveals changes in proteins that are heart specific.

While the two approaches illustrated here provide information of a very different nature, they can be consolidated into a single picture of how BAG3 variation affects heart homeostasis. The knowledge gap between the genotype (the DNA variant) and the phenotype (disease, or in this case the contractility/viability phenotype on a plate) can be closed using protein-protein interactions as a guide. In fact, the PPI analysis provides us with interesting heart-specific leads that we can then test for their necessity to overcome proteasomal overload stress.

# Chapter 7 : Concluding Remarks

In my thesis research I have focused on studying how unbiased studies of protein interactions can be used to identify novel functions of proteins, to pinpoint at specific mechanisms for how genetic variants can affect protein behavior and lead to disease. My work has relied heavily on the use of iPS cells as a source of disease-relevant cells for the study of protein interactions, and on genome engineering to generate the ideal platform for the dissection of the effect of single amino acid transitions on protein function. More specifically, I have focused on studying the effect of variation in the BAG3 gene, a gene that, albeit expressed in many tissues, is only associated to malfunction in heart muscle.

In my research I have optimize methods and procedures to harness affinity purification-mass spectrometry to study disease-associated protein variants. I also produced multiple cell lines that were used for phenotyping and protein interaction analysis. The analysis of these cell lines by affinity purification-mass spectrometry allowed us to identify proteins (some of them cardiac-specific) that bind less strongly to disease-associated variants. Last but not least, we were able to observe a phenotype for these cell lines bearing single amino acid variants in their genomes, in the form of a different response to proteotoxic stress induced by bortezomib. Along the way, I generated a number of independent datasets on putative interaction complexes for two different chaperone families, and explored the limits of genome engineering technologies to engineer small and large modifications and how to phase them.

The combination of stem cell technology, genome engineering and protein-protein interaction by mass spectrometry is a very powerful way to study genetic disease. As showed here for the case of BAG3 and cardiac disease, it allows us to point to specific mechanisms that then can be explored by other means and potentially exploited as therapeutic targets.

These three fields that my thesis work focused have been in constant change, and I have witnessed significant improvements in the methods during my 6 years of work. My work in this project has allowed me to gain advanced knowledge in each of these rapidly progressing fields. Perhaps more importantly, it has allowed me to learn how to thrive in a dynamic, ever-changing environment, where no field of study is too daunting to approach and I am learning every day. I am confident that the combination of these factors has prepared me for a career of exploration of ways to tackle human disease using molecular biology. I am excited to see the challenges and projects the future will bring, and how I will be able to keep contributing to the unstoppable progress against human disease.

# Bibliography

1. Robinson CV, Sali A, Baumeister W (2007) The molecular sociology of the cell. *Nature* **450**, 973–82.
2. Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101–13.
3. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) *Molecular Biology of the Cell*, 4th edn, Garland Science.
4. Xue LC, Dobbs D, Bonvin AMJJ, Honavar V (2015) Computational prediction of protein interfaces: A review of data driven methods. *FEBS Lett* **589**, 3516–26.
5. Xing S, Wallmeroth N, Berendzen KW, Grefen C (2016) Techniques for the Analysis of Protein-Protein Interactions in Vivo. *Plant Physiol* **171**, 727–58.
6. Braun P, Gingras A-C (2012) History of protein–protein interactions: From egg-white to complex networks. *PROTEOMICS* **12**, 1478–98.
7. Dunham WH, Mullin M, Gingras A-C (2012) Affinity-purification coupled to mass spectrometry: Basic principles and strategies. *PROTEOMICS* **12**, 1576–90.
8. Barabási A-L, Gulbahce N, Loscalzo J (2010) Network medicine: a network-based approach to human disease. *Nat Rev Genet* **12**, nrg2918.
9. Zhong Q, Simonis N, Li Q-R, Charloreaux B, Heuze F, Klitgord N, Tam S, Yu H, et al (2009) Edgetic perturbation models of human inherited disorders. *Mol Syst Biol* **5**, 321.
10. Lambert J-P, Ivosev G, Couzens AL, Larsen B, Taipale M, Lin Z-Y, Zhong Q, Lindquist S, et al (2013) Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nat Methods* **10**, 1239–45.
11. Ideker T, Krogan NJ (2012) Differential network biology. *Mol Syst Biol* **8**, 565.
12. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, Ferranti SD de, Floyd J, et al (2017) Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. *Circulation* **135**, e146–603.
13. Jefferies JL (2010) Dilated cardiomyopathy. *The Lancet* **375**, 752–62.
14. Haas J, Frese KS, Peil B, Kloos W, Keller A, Nietsch R, Feng Z, Müller S, et al (2015) Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur Heart J* **36**, 1123–35.



15. Blausencom staff, staff B com (2014) Medical gallery of Blausen Medical 2014. *WikiJournal Med* **1**, 10.
16. Franaszczyk M, Bilinska ZT, Sobieszczka-Małek M, Michalak E, Sleszycka J, Sioma A, Małek Łukasz A, Kaczmarek D, et al (2014) The BAG3 gene variants in Polish patients with dilated cardiomyopathy: four novel mutations and a genotype-phenotype correlation. *J Transl Med* **12**, 192.
17. Norton N, Li D, Rieder MJ, Siegfried JD, Rampersaud E, Züchner S, Mangos S, Gonzalez-Quintana J, et al (2011) Genome-wide Studies of Copy Number Variation and Exome Sequencing Identify Rare Variants in BAG3 as a Cause of Dilated Cardiomyopathy. *Am J Hum Genet* **88**, 273–82.
18. Villard E, Perret C, Gary F, Proust C, Dilanian G, Hengstenberg C, Ruppert V, Arbustini E, et al (2011) A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J* **32**, 1065–76.
19. Feldman AM, Begay RL, Knezevic T, Myers VD, Slavov DB, Zhu W, Gowan K, Graw SL, et al (2014) Decreased Levels of BAG3 in a Family With a Rare Variant and in Idiopathic Dilated Cardiomyopathy. *J Cell Physiol*, n/a-n/a.
20. Chami N, Tadros R, Lemarbre F, Lo KS, Beaudoin M, Robb L, Labuda D, Tardif J-C, et al (2014) Nonsense Mutations in BAG3 are Associated With Early-Onset Dilated Cardiomyopathy in French Canadians. *Can J Cardiol* **30**, 1655–61.
21. d’Avenia M, Citro R, De Marco M, Veronese A, Rosati A, Visone R, Leptidis S, Philippen L, et al (2015) A novel miR-371a-5p-mediated pathway, leading to BAG3 upregulation in cardiomyocytes in response to epinephrine, is lost in Takotsubo cardiomyopathy. *Cell Death Dis* **6**, e1948.
22. Citro R, d’Avenia M, De Marco M, Giudice R, Mirra M, Ravera A, Silverio A, Farina R, et al Polymorphisms of the antiapoptotic protein bag3 may play a role in the pathogenesis of tako-tsubo cardiomyopathy. *Int J Cardiol*. doi:10.1016/j.ijcard.2013.03.050.
23. Selcen D, Muntoni F, Burton BK, Pegoraro E, Sewry C, Bite AV, Engel AG (2009) Mutation in BAG3 causes severe dominant childhood muscular dystrophy. *Ann Neurol* **65**, 83–9.
24. Odgerel Z, Sarkozy A, Lee H-S, McKenna C, Rankin J, Straub V, Lochmüller H, Paola F, et al (2010) Inheritance patterns and phenotypic features of myofibrillar myopathy associated with a BAG3 mutation. *Neuromuscul Disord* **20**, 438–42.
25. Semmler A-L, Sacconi S, Bach JE, Liebe C, Bürmann J, Kley RA, Ferbert A, Anderheiden R, et al (2014) Unusual multisystemic involvement and a novel BAG3 mutation revealed

- by NGS screening in a large cohort of myofibrillar myopathies. *Orphanet J Rare Dis* **9**, 121.
26. Konersman CG, Bordini BJ, Scharer G, Lawlor MW, Zangwill S, Southern JF, Amos L, Geddes GC, et al (2015) BAG3 myofibrillar myopathy presenting with cardiomyopathy. *Neuromuscul Disord* **25**, 418–22.
  27. Lee H, Cherk S, Chan S, Wong S, Tong T, Ho W, Chan A, Lee K, Mak C (2011) BAG3-related myofibrillar myopathy in a Chinese family. *Clin Genet*. doi:10.1111/j.1399-0004.2011.01659.x.
  28. Consortium T 1000 GP (2015) A global reference for human genetic variation. *Nature* **526**, 68.
  29. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, et al (2015) Tissue-based map of the human proteome. *Science* **347**, 1260419.
  30. Judge LM, Perez-Bermejo JA, Truong A, Ribeiro AJS, Yoo JC, Jensen CL, Mandegar MA, Huebsch N, et al (2017) A BAG3 chaperone complex maintains cardiomyocyte function during proteotoxic stress. *JCI Insight* **2**. doi:10.1172/jci.insight.94623.
  31. Mayer MP, Bukau B (2005) Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell Mol Life Sci* **62**, 670–84.
  32. Rauch JN, Gestwicki JE (2014) Binding of Human Nucleotide Exchange Factors to Heat Shock Protein 70 (Hsp70) Generates Functionally Distinct Complexes in Vitro. *J Biol Chem* **289**, 1402–14.
  33. Rauch JN, Zuiderweg ERP, Gestwicki JE (2016) Non-canonical Interactions between Heat Shock Cognate Protein 70 (Hsc70) and Bcl2-associated Anthanogene (BAG) Co-Chaperones Are Important for Client Release. *J Biol Chem* **291**, 19848–57.
  34. Takayama S, Reed JC (2001) Molecular chaperone targeting and regulation by BAG family proteins. *Nat Cell Biol* **3**, E237–41.
  35. Arimura T, Ishikawa T, Nunoda S, Kawai S, Kimura A (2011) Dilated cardiomyopathy-associated BAG3 mutations impair Z-disc assembly and enhance sensitivity to apoptosis in cardiomyocytes. *Hum Mutat*. doi:10.1002/humu.21603.
  36. Garnier S, Hengstenberg C, Lamblin N, Dubourg O, De Groote P, Fauchier L, Trochu J-N, Arbustini E, et al (2015) Involvement of BAG3 and HSPB7 loci in various

- etiologies of systolic heart failure: Results of a European collaboration assembling more than 2000 patients. *Int J Cardiol* **189**, 105–7.
37. Toro R, Pérez-Serra A, Campuzano O, Moncayo-Arlandi J, Allegue C, Iglesias A, Mangas A, Brugada R (2016) Familial Dilated Cardiomyopathy Caused by a Novel Frameshift in the BAG3 Gene. *PLOS ONE* **11**, e0158730.
  38. Rafiq MA, Chaudhry A, Care M, Spears DA, Morel CF, Hamilton RM (2017) Whole exome sequencing identified 1 base pair novel deletion in BCL2-associated athanogene 3 (BAG3) gene associated with severe dilated cardiomyopathy (DCM) requiring heart transplant in multiple family members. *Am J Med Genet A* **173**, 699–705.
  39. Ruppert V, Onda-Ono C, Meyer T, Richter A, Maisch B, Pankuweit S (2013) Genetic variability in the bag3 gene (BCL2-associated athanogene 3) in patients with dilated cardiomyopathy. *Eur Heart J* **34**. doi:10.1093/eurheartj/eh309.P3010.
  40. Marcus JH, Novembre J (2017) Visualizing the geography of genetic variants. *Bioinformatics* **33**, 594–5.
  41. Carra S, Seguin SJ, Lambert H, Landry J (2008) HspB8 Chaperone Activity toward Poly(Q)-containing Proteins Depends on Its Association with Bag3, a Stimulator of Macroautophagy. *J Biol Chem* **283**, 1437–44.
  42. Gamerding M, Kaya AM, Wolfrum U, Clement AM, Behl C (2011) BAG3 mediates chaperone-based aggresome-targeting and selective autophagy of misfolded proteins. *EMBO Rep* **12**, 149–56.
  43. Arndt V, Dick N, Tawo R, Dreiseidler M, Wenzel D, Hesse M, Fürst DO, Saftig P, et al (2010) Chaperone-Assisted Selective Autophagy Is Essential for Muscle Maintenance. *Curr Biol* **20**, 143–8.
  44. Ulbricht A, Arndt V, Hohfeld J (2013) Chaperone-assisted proteostasis is essential for mechanotransduction in mammalian cells. *Commun Integr Biol* **6**. doi:10.4161/cib.24925.
  45. Fang X, Bogomolovas J, Wu T, Zhang W, Liu C, Veevers J, Stroud MJ, Zhang Z, et al (2017) Loss-of-function mutations in co-chaperone BAG3 destabilize small HSPs and cause cardiomyopathy. *J Clin Invest* **127**. doi:10.1172/JCI94310.
  46. Hishiya A, Kitazawa T, Takayama S (2010) BAG3 and Hsc70 interact with actin capping protein CapZ to maintain myofibrillar integrity under mechanical stress. *Circ Res* **107**, 1220–31.

47. Knezevic T, Myers VD, Gordon J, Tilley DG, Iii TES, Wang J, Khalili K, Cheung JY, Feldman AM (2015) BAG3: a new player in the heart failure paradigm. *Heart Fail Rev*, 1–12.
48. Willis MS, Schisler JC, Portbury AL, Patterson C (2009) Build it up–Tear it down: protein quality control in the cardiac sarcomere. *Cardiovasc Res* **81**, 439–48.
49. Willis MS, Patterson C (2013) Proteotoxicity and Cardiac Dysfunction — Alzheimer’s Disease of the Heart? *N Engl J Med* **368**, 455–64.
50. Su H, Wang X (2010) The ubiquitin-proteasome system in cardiac proteinopathy: a quality control perspective. *Cardiovasc Res* **85**, 253–62.
51. McLendon PM, Robbins J (2015) Proteotoxicity and cardiac dysfunction. *Circ Res* **116**, 1863–82.
52. Tarone G, Brancaccio M (2014) Keep your heart in shape: molecular chaperone networks for treating heart disease. *Cardiovasc Res* **102**, 346–61.
53. Bos JM, Ackerman MJ (2010) Z-Disc Genes in Hypertrophic Cardiomyopathy: Stretching the Cardiomyopathies? *J Am Coll Cardiol* **55**, 1136–8.
54. Knöll R, Buyandelger B, Lab M (2011) The Sarcomeric Z-Disc and Z-Discopathies. *J Biomed Biotechnol* **2011**. doi:10.1155/2011/569628.
55. Tayal U, Prasad S, Cook SA (2017) Genetics and genomics of dilated cardiomyopathy and systolic heart failure. *Genome Med* **9**, 20.
56. Merabova N, Sariyer IK, Saribas AS, Knezevic T, Gordon J, Turco MC, Rosati A, Weaver M, et al (2015) WW Domain of BAG3 Is Required for the Induction of Autophagy in Glioma Cells. *J Cell Physiol* **230**, 831–41.
57. Chen Y, Yang L-N, Cheng L, Tu S, Guo S-J, Le H-Y, Xiong Q, Mo R, et al (2013) Bcl2-associated Athanogene 3 Interactome Analysis Reveals a New Role in Modulating Proteasome Activity. *Mol Cell Proteomics* **12**, 2804–19.
58. Behl C (2016) Breaking BAG: The Co-Chaperone BAG3 in Health and Disease. *Trends Pharmacol Sci* **37**, 672–88.
59. Quintana MT, Parry TL, He J, Yates CC, Sidorova TN, Murray KT, Bain JR, Newgard CB, et al (2016) Cardiomyocyte-Specific Human Bcl2-Associated Anthanogene 3 P209L

- Expression Induces Mitochondrial Fragmentation, Bcl2-Associated Anthanogene 3 Haploinsufficiency, and Activates p38 Signaling. *Am J Pathol* **186**, 1989–2007.
60. Ruparelia AA, Oorschot V, Vaz R, Ramm G, Bryson-Richardson RJ (2014) Zebrafish models of BAG3 myofibrillar myopathy suggest a toxic gain of function leading to BAG3 insufficiency. *Acta Neuropathol (Berl)*, 1–13.
  61. Homma S, Iwasaki M, Shelton GD, Engvall E, Reed JC, Takayama S (2006) BAG3 Deficiency Results in Fulminant Myopathy and Early Lethality. *Am J Pathol* **169**, 761–73.
  62. Colman A, Dreesen O (2009) Pluripotent Stem Cells and Disease Modeling. *Cell Stem Cell* **5**, 244–7.
  63. Yoshida Y, Yamanaka S (2010) Recent Stem Cell Advances: Induced Pluripotent Stem Cells for Disease Modeling and Stem Cell–Based Regeneration. *Circulation* **122**, 80–7.
  64. Zhang J, Wilson GF, Soerens AG, Koonce CH, Yu J, Palecek SP, Thomson JA, Kamp TJ (2009) Functional Cardiomyocytes Derived From Human Induced Pluripotent Stem Cells. *Circ Res* **104**, e30–41.
  65. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–76.
  66. Park I-H, Arora N, Huo H, Maherali N, Ahfeldt T, Shimamura A, Lensch MW, Cowan C, et al (2008) Disease-Specific Induced Pluripotent Stem Cells. *Cell* **134**, 877–86.
  67. Reppel M, Pillekamp F, Brockmeier K, Matzkies M, Bekcioglu A, Lipke T, Nguemo F, Bonnemeier H, Hescheler J (2005) The electrocardiogram of human embryonic stem cell–derived cardiomyocytes. *J Electrocardiol* **38**, 166–70.
  68. Gaj T, Gersbach CA, Barbas III CF (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*. doi:10.1016/j.tibtech.2013.04.004.
  69. Carroll D (2011) Genome Engineering With Zinc-Finger Nucleases. *Genetics* **188**, 773–82.
  70. Bogdanove AJ, Voytas DF (2011) TAL Effectors: Customizable Proteins for DNA Targeting. *Science* **333**, 1843–6.

71. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–6.
72. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, et al (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–23.
73. Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096.
74. Ding Q, Lee Y-K, Schaefer EAK, Peters DT, Veres A, Kim K, Kuperwasser N, Motola DL, et al (2013) A TALEN Genome-Editing System for Generating Human Stem Cell-Based Disease Models. *Cell Stem Cell* **12**, 238–51.
75. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási A-L (2015) Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601.
76. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, et al (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, nature22366.
77. Gingras A-C, Gstaiger M, Raught B, Aebersold R (2007) Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* **8**, nrm2208.
78. Mering C von, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**, nature750.
79. Mehta V, Trinkle-Mulcahy L (2016) Recent advances in large-scale protein interactome mapping. *F1000Research* **5**, 782.
80. Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, et al (2015) A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **163**, 712–23.
81. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, et al (2012) A Census of Human Soluble Protein Complexes. *Cell* **150**, 1068–81.
82. Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O’Shea EK (2003) Global analysis of protein localization in budding yeast. *Nature* **425**, nature02026.
83. Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, et al (2002) Subcellular localization of the yeast proteome. *Genes Dev* **16**, 707–19.

84. Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, Snyder M, Oliver SG, et al (2006) Mapping Pathways and Phenotypes by Systematic Gene Overexpression. *Mol Cell* **21**, 319–30.
85. Ratz M, Testa I, Hell SW, Jakobs S (2015) CRISPR/Cas9-mediated endogenous protein tagging for RESOLFT super-resolution microscopy of living human cells. *Sci Rep* **5**, srep09592.
86. Gibson TJ, Seiler M, Veitia RA (2013) The transience of transient overexpression. *Nat Methods* **10**, 715–21.
87. Veitia RA, Potier MC (2015) Gene dosage imbalances: action, reaction, and models. *Trends Biochem Sci* **40**, 309–17.
88. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B (2009) Intrinsic Protein Disorder and Interaction Promiscuity Are Widely Associated with Dosage Sensitivity. *Cell* **138**, 198–208.
89. Singh GP, Dash D (2013) Electrostatic Mis-Interactions Cause Overexpression Toxicity of Proteins in *E. coli*. *PLOS ONE* **8**, e64893.
90. Fink AL (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des* **3**, R9–23.
91. Stefani M (2004) Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochim Biophys Acta BBA - Mol Basis Dis* **1739**, 5–25.
92. Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS (2010) A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* **10**, nrc2771.
93. Shastri BS (1995) Overexpression of genes in health and sickness. A bird's eye view. *Comp Biochem Physiol B Biochem Mol Biol* **112**, 1–13.
94. Prelich G (2012) Gene Overexpression: Uses, Mechanisms, and Interpretation. *Genetics* **190**, 841–54.
95. Lin Y-C, Boone M, Meuris L, Lemmens I, Roy NV, Soete A, Reumers J, Moisse M, et al (2014) Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun* **5**, ncomms5767.
96. Mittelman D, Wilson JH (2013) The fractured genome of HeLa cells. *Genome Biol* **14**, 111.
97. Pan C, Kumar C, Bohl S, Klingmueller U, Mann M (2009) Comparative Proteomic Phenotyping of Cell Lines and Primary Cells to Assess Preservation of Cell Type-specific Functions. *Mol Cell Proteomics* **8**, 443–50.

98. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, et al (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–7.
99. Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-Huntington J, Yang F, Peng J, Weile J, et al (2015) Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders. *Cell* **161**, 647–60.
100. Roberts B, Haupt A, Tucker A, Grancharova T, Arakaki J, Fuqua MA, Nelson A, Hookway C, et al (2017) Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. *Mol Biol Cell* **28**, 2854–74.
101. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–21.
102. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281–308.
103. Thomas KR, Folger KR, Capecchi MR (1986) High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419–28.
104. Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS (1985) Insertion of DNA sequences into the human chromosomal  $\beta$ -globin locus by homologous recombination. *Nature* **317**, 317230a0.
105. Itzhaki JE, Porter ACG (1991) Targeted disruption of a human interferon-inducible gene detected by secretion of human growth hormone. *Nucleic Acids Res* **19**, 3835–42.
106. Kim J-S, Bonifant C, Bunz F, Lane WS, Waldman T (2008) Epitope tagging of endogenous genes in diverse human cell lines. *Nucleic Acids Res* **36**, e127–e127.
107. Hirata R, Chamberlain J, Dong R, Russell DW (2002) Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. *Nat Biotechnol* **20**, nbt0702-735–735.
108. Rago C, Vogelstein B, Bunz F (2007) Genetic knockouts and knockins in human somatic cells. *Nat Protoc* **2**, nprot.2007.408.
109. Song J, Hao Y, Du Z, Wang Z, Ewing RM (2012) Identifying Novel Protein Complexes in Cancer Cells Using Epitope-Tagging of Endogenous Human Genes and Affinity-Purification Mass Spectrometry. *J Proteome Res* **11**, 5630–41.
110. Kamiyama D, Sekine S, Barsi-Rhyne B, Hu J, Chen B, Gilbert LA, Ishikawa H, Leonetti MD, et al (2016) Versatile protein tagging in cells with split fluorescent protein. *Nat Commun* **7**, ncomms11046.



111. Leonetti MD, Sekine S, Kamiyama D, Weissman JS, Huang B (2016) A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc Natl Acad Sci* **113**, E3501–8.
112. Savic D, Partridge EC, Newberry KM, Smith SB, Meadows SK, Roberts BS, Mackiewicz M, Mendenhall EM, Myers RM (2015) CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Res* **25**, 1581–9.
113. Van Nostrand EL, Gelboin-Burkhart C, Wang R, Pratt GA, Blue SM, Yeo GW (2017) CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA binding proteins. *Methods* **118–119**, 50–9.
114. Bressan RB, Dewari PS, Kalantzaki M, Gangoso E, Matjusaitis M, Garcia-Diaz C, Blin C, Grant V, et al (2017) Efficient CRISPR/Cas9-assisted gene targeting enables rapid and precise genetic manipulation of mammalian neural stem cells. *Development*, dev.140855.
115. Kim S, Kim D, Cho SW, Kim J, Kim J-S (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **24**, 1012–9.
116. Liang X, Potter J, Kumar S, Zou Y, Quintanilla R, Sridharan M, Carte J, Chen W, et al (2015) Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* **208**, 44–53.
117. Selbach M, Mann M (2006) Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK). *Nat Methods* **3**, 981–3.
118. Hultquist JF, Schumann K, Woo JM, Manganaro L, McGregor MJ, Doudna J, Simon V, Krogan NJ, Marson A (2016) A Cas9 Ribonucleoprotein Platform for Functional Genetic Studies of HIV-Host Interactions in Primary Human T Cells. *Cell Rep* **17**, 1438–52.
119. Hendel A, Bak RO, Clark JT, Kennedy AB, Ryan DE, Roy S, Steinfeld I, Lunstad BD, et al (2015) Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol* **33**, nbt.3290.
120. Rees JS, Li X-W, Perrett S, Lilley KS, Jackson AP (2015) Protein Neighbors and Proximity Proteomics. *Mol Cell Proteomics* **14**, 2848–56.
121. Miyaoka Y, Chan AH, Judge LM, Yoo J, Huang M, Nguyen TD, Lizarraga PP, So P-L, Conklin BR (2014) Isolation of single-base genome-edited human iPS cells without antibiotic selection. *Nat Methods* **11**, 291–3.

122. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, nature24644.
123. Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, Kim YB, Badran AH, Liu DR (2017) Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **3**, eaao4774.
124. Shi T, Song E, Nie S, Rodland KD, Liu T, Qian W-J, Smith RD (2016) Advances in targeted proteomics and applications to biomedical research. *PROTEOMICS* **16**, 2160–82.
125. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, nbt.2623.
126. Cameron P, Fuller CK, Donohoue PD, Jones BN, Thompson MS, Carter MM, Gradia S, Vidal B, et al (2017) Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods* **14**, 600–6.
127. Chapman JE, Gillum D, Kiani S (2017) Approaches to Reduce CRISPR Off-Target Effects for Safer Genome Editing. *Appl Biosaf* **22**, 7–13.
128. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, et al (2015) In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, nature14299.
129. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK (2016) High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, nature16526.
130. Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, Li T, Miteva YV, Hauri S, et al (2013) The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* **10**, 730–6.
131. Choi H, Larsen B, Lin Z-Y, Breitkreutz A, Mellacheruvu D, Fermin D, Qin ZS, Tyers M, et al (2011) SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Meth* **8**, 70–3.
132. Sowa ME, Bennett EJ, Gygi SP, Harper JW (2009) Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell* **138**, 389–403.
133. Sinz A (2010) Investigation of protein–protein interactions in living cells by chemical crosslinking and mass spectrometry. *Anal Bioanal Chem* **397**, 3433–40.134. Tran JC,

- Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, et al (2011) Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, nature10575.
135. Vicart P, Caron A, Guicheney P, Li Z, Prévost MC, Faure A, Chateau D, Chapon F, et al (1998) A missense mutation in the alphaB-crystallin chaperone gene causes a desmin-related myopathy. *Nat Genet* **20**, 92–5.
  136. Youn D-Y, Lee D-H, Lim M-H, Yoon J-S, Lim JH, Jung SE, Yeum CE, Park CW, et al (2008) Bis deficiency results in early lethality with metabolic deterioration and involution of spleen and thymus. *Am J Physiol - Endocrinol Metab* **295**, E1349–57.
  137. Franceschelli S, Rosati A, Lerosé R, De Nicola S, Turco MC, Pascale M (2008) bag3 gene expression is regulated by heat shock factor 1. *J Cell Physiol* **215**, 575–7.
  138. Fuchs M, Poirier DJ, Seguin SJ, Lambert H, Carra S, Charette SJ, Landry J (2010) Identification of the key structural motifs involved in HspB8/HspB6–Bag3 interaction. *Biochem J* **425**, 245–57.
  139. Shemetov AA, Gusev NB (2011) Biochemical characterization of small heat shock protein HspB8 (Hsp22)–Bag3 interaction. *Arch Biochem Biophys* **513**, 1–9.
  140. Rauch JN, Tse E, Freilich R, Mok S-A, Makley LN, Southworth DR, Gestwicki JE BAG3 is a modular, scaffolding protein that physically links heat shock protein 70 (Hsp70) to the small heat shock proteins. *J Mol Biol.* doi:10.1016/j.jmb.2016.11.013.
  141. Minoia M, Boncoraglio A, Vinet J, Morelli F, Brunsting J, Poletti A, Krom S, Reits E, et al (2014) BAG3 induces the sequestration of proteasomal clients into cytoplasmic puncta: implication for a proteasome-to-autophagy switch. *Autophagy* **10**, 54–53.
  142. Heidrsbach A, Saxby C, Carver-Moore K, Huang Y, Ang Y-S, Jong PJ de, Ivey KN, Srivastava D (2013) microRNA-1 regulates sarcomere formation and suppresses smooth muscle gene expression in the mammalian heart. *eLife* **2**, e01323.
  143. Ribeiro AJS, Ang Y-S, Fu J-D, Rivas RN, Mohamed TMA, Higgs GC, Srivastava D, Pruitt BL (2015) Contractility of single cardiomyocytes differentiated from pluripotent stem cells depends on physiological shape and substrate stiffness. *Proc Natl Acad Sci* **112**, 12705–10.
  144. Rapino F, Jung M, Fulda S (2014) BAG3 induction is required to mitigate proteotoxicity via selective autophagy following inhibition of constitutive protein degradation pathways. *Oncogene* **33**, 1713.

145. Nowis D, Mączewski M, Mackiewicz U, Kujawa M, Ratajska A, Wieckowski MR, Wilczyński GM, Malinowska M, et al (2010) Cardiotoxicity of the Anticancer Therapeutic Agent Bortezomib. *Am J Pathol* **176**, 2658–68.
146. Orciuolo E, Buda G, Cecconi N, Galimberti S, Versari D, Cervetti G, Salvetti A, Petrini M (2007) Unexpected cardiotoxicity in haematological bortezomib treated patients. *Br J Haematol* **138**, 396–7.
147. Maddah M, Heidmann JD, Mandegar MA, Walker CD, Bolouki S, Conklin BR, Loewke KE (2015) A Non-invasive Platform for Functional Characterization of Stem-Cell-Derived Cardiomyocytes with Applications in Cardiotoxicity Testing. *Stem Cell Rep* **4**, 621–31.
148. Onyx Pharmaceuticals Inc Thousand Oaks, CA (2012) Kyprolis (carfilzomib) for injection [package insert].
149. Millennium Pharmaceuticals, Inc Cambridge, MA (2005) Velcade (bortezomib) for injection [package insert].
150. Huebsch N, Loskill P, Deveshwar N, Spencer CI, Judge LM, Mandegar MA, B Fox C, Mohamed TMA, et al (2016) Miniaturized iPS-Cell-Derived Cardiac Muscles for Physiologically Relevant Drug Response Analyses. *Sci Rep* **6**. doi:10.1038/srep24726.
151. Wyles SP, Li X, Hrstka SC, Reyes S, Oommen S, Beraldi R, Edwards J, Terzic A, et al (2016) Modeling structural and functional deficiencies of RBM20 familial dilated cardiomyopathy using human induced pluripotent stem cells. *Hum Mol Genet* **25**, 254–65.
152. Sun N, Yazawa M, Liu J, Han L, Sanchez-Freire V, Abilez OJ, Navarrete EG, Hu S, et al (2012) Patient-Specific Induced Pluripotent Stem Cells as a Model for Familial Dilated Cardiomyopathy. *Sci Transl Med* **4**, 130ra47.
153. Du Z-X, Zhang H-Y, Meng X, Gao Y-Y, Zou R-L, Liu B-Q, Guan Y, Wang H-Q (2009) Proteasome inhibitor MG132 induces BAG3 expression through activation of heat shock factor 1. *J Cell Physiol* **218**, 631–7.
154. Kaake RM, Wang X, Huang L (2010) Profiling of Protein Interaction Networks of Protein Complexes Using Affinity Purification and Quantitative Mass Spectrometry. *Mol Cell Proteomics* **9**, 1650–65.

155. Ganassi M, Mateju D, Bigi I, Mediani L, Poser I, Lee HO, Seguin SJ, Morelli FF, et al (2016) A Surveillance Function of the HSPB8-BAG3-HSP70 Chaperone Complex Ensures Stress Granule Integrity and Dynamism. *Mol Cell* **63**, 796–810.
156. Burridge PW, Li YF, Matsa E, Wu H, Ong S-G, Sharma A, Holmström A, Chang AC, et al (2016) Human induced pluripotent stem cell-derived cardiomyocytes recapitulate the predilection of breast cancer patients to doxorubicin-induced cardiotoxicity. *Nat Med* **advance online publication**. doi:10.1038/nm.4087.
157. Liang P, Lan F, Lee AS, Gong T, Sanchez-Freire V, Wang Y, Diecke S, Sallam K, et al (2013) Drug Screening Using a Library of Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes Reveals Disease Specific Patterns of Cardiotoxicity. *Circulation*, CIRCULATIONAHA.113.001883.
158. Wang G, McCain ML, Yang L, He A, Pasqualini FS, Agarwal A, Yuan H, Jiang D, et al (2014) Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nat Med* **advance online publication**. doi:10.1038/nm.3545.
159. Meyer H, Wehl CC (2014) The VCP/p97 system at a glance: connecting cellular function to disease pathogenesis. *J Cell Sci* **127**, 3877–83.
160. Okita K, Matsumura Y, Sato Y, Okada A, Morizane A, Okamoto S, Hong H, Nakagawa M, et al (2011) A more efficient method to generate integration-free human iPSCs. *Nat Methods* **8**, 409–12.
161. Grau J, Boch J, Posch S (2013) TALENoffer: genome-wide TALEN off-target prediction. *Bioinformatics* **29**, 2931–2.
162. Lian X, Zhang J, Azarin SM, Zhu K, Hazeltine LB, Bao X, Hsiao C, Kamp TJ, Palecek SP (2013) Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/ $\beta$ -catenin signaling under fully defined conditions. *Nat Protoc* **8**, 162–75.
163. Tohyama S, Hattori F, Sano M, Hishiki T, Nagahata Y, Matsuura T, Hashimoto H, Suzuki T, et al (2013) Distinct Metabolic Flow Enables Large-Scale Purification of Mouse and Human Pluripotent Stem Cell-Derived Cardiomyocytes. *Cell Stem Cell* **12**, 127–37.
164. Mandegar MA, Huebsch N, Frolov EB, Shin E, Truong A, Olvera MP, Chan AH, Miyaoka Y, et al (2016) CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell Stem Cell* **18**, 541–53.
165. Ribeiro AJ, Schwab O, Mandegar MA, Ang Y-S, Conklin BR, Srivastava D, Pruitt BL (2017) Multi-Imaging Method to Assay the Contractile Mechanical Output of

Micropatterned Human iPSC-Derived Cardiac Myocytes. *Circ Res*, CIRCRESAHA.116.310363.

166. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–72.
167. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol* **57**, 289–300.
168. R Core Team (2017) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria Available at: <https://www.R-project.org/>.
169. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **13**, 2498–504.
170. Hulsen T, de Vlieg J, Alkema W (2008) BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, 488.
171. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, et al (2011) Genetic engineering of human ES and iPS cells using TALE nucleases. *Nat Biotechnol* **29**, 731–4.
172. Regan JF, Kamitaki N, Legler T, Cooper S, Klitgord N, Karlin-Neumann G, Wong C, Hodges S, et al (2015) A Rapid Molecular Approach for Chromosomal Phasing. *PLOS ONE* **10**, e0118270.
173. Yang L, Mali P, Kim-Kiselak C, Church G (2014) CRISPR-Cas-Mediated Targeted Genome Editing in Human Cells. In: *Gene Correction*, Methods in Molecular Biology, Humana Press, Totowa, NJ, pp 245–67.
174. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, et al (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–32.
175. Basu AS (2017) Digital Assays Part I: Partitioning Statistics and Digital PCR. *SLAS Technol Transl Life Sci Innov* **22**, 369–86.
176. Oyer JA, Chu A, Brar S, Turker MS (2009) Aberrant Epigenetic Silencing Is Triggered by a Transient Reduction in Gene Expression. *PLOS ONE* **4**, e4832.
177. Einhauer A, Jungbauer A (2001) The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *J Biochem Biophys Methods* **49**, 455–65.

178. Stark K, Esslinger UB, Reinhard W, Petrov G, Winkler T, Komajda M, Isnard R, Charron P, et al (2010) Genetic Association Study Identifies HSPB7 as a Risk Gene for Idiopathic Dilated Cardiomyopathy. *PLoS Genet* **6**, e1001167.
179. Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, Garitaonandia I, Müller F-J, et al (2012) Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* **10**, 620–34.
180. Narsinh KH, Sun N, Sanchez-Freire V, Lee AS, Almeida P, Hu S, Jan T, Wilson KD, et al (2011) Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells. *J Clin Invest* **121**, 1217–21.
181. HGVS Recommendations for the Description of Sequence Variants: 2016 Update - Dunnen - 2016 - Human Mutation - Wiley Online Library Available at: <http://onlinelibrary.wiley.com/doi/10.1002/humu.22981/abstract> [Accessed September 8, 2017].
182. Lehner B (2013) Genotype to phenotype: lessons from model organisms for human genetics. *Nat Rev Genet* **14**, 168.
183. Harper AR, Topol EJ, Nayee S (2015) Protective alleles and modifier variants in human health and disease. *Nat Rev Genet* **16**, 689.
184. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, et al (2010) Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* **28**, 1248–50.
185. Guan S, Price JC, Prusiner SB, Ghaemmaghami S, Burlingame AL (2011) A Data Processing Pipeline for Mammalian Proteome Dynamics Studies Using Stable Isotope Metabolic Labeling. *Mol Cell Proteomics MCP* **10**. doi:10.1074/mcp.M111.010728.
186. Verschueren E, Von Dollen J, Cimermanic P, Gulbahce N, Sali A, Krogan N (2015) Scoring Large Scale Affinity Purification Mass Spectrometry Datasets with MIST. *Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al* **49**, 8.19.1-8.19.16.
187. Teo G, Liu G, Zhang J, Nesvizhskii AI, Gingras A-C, Choi H (2014) SAINTexpress: Improvements and additional features in Significance Analysis of INTeractome software. *J Proteomics* **100**, 37–43.
188. Fischer M, Zilkenat S, Gerlach RG, Wagner S, Renard BY (2014) Pre- and Post-Processing Workflow for Affinity Purification Mass Spectrometry Data. *J Proteome Res* **13**, 2239–49.

189. Smyth GK, Speed T (2003) Normalization of cDNA microarray data. *Methods* **31**, 265–73.
190. Knight JDR, Choi H, Gupta GD, Pelletier L, Raught B, Nesvizhskii AI, Gingras A-C (2017) ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nat Methods* **14**, 645–6.
191. Razick S, Magklaras G, Donaldson IM (2008) iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinformatics* **9**, 405.
192. Ruepp A, Waegelé B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes H-W (2010) CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res* **38**, D497–501.
193. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539–539.
194. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–91.
195. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol Cell Proteomics MCP* **13**, 2513–26.
196. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG (2005) Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics. *Mol Cell Proteomics* **4**, 1487–502.
197. Lundgren DH, Hwang S-I, Wu L, Han DK (2010) Role of spectral counting in quantitative proteomics. *Expert Rev Proteomics* **7**, 39–53.
198. Bolstad BM, Irizarry RA, Åstrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–93.
199. Zhang Y, Wen Z, Washburn MP, Florens L (2009) Effect of Dynamic Exclusion Duration on Spectral Count Based Quantitative Proteomics. *Anal Chem* **81**, 6317–26.
200. Morris JH, Knudsen GM, Verschueren E, Johnson JR, Cimermancic P, Greninger AL, Pico AR (2014) Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions. *Nat Protoc* **9**, 2539–54.



201. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin Z-Y, Breitkreutz B-J, et al (2010) A Global Protein Kinase and Phosphatase Interaction Network in Yeast. *Science* **328**, 1043–6.
202. Skarra DV, Goudreault M, Choi H, Mullin M, Nesvizhskii AI, Gingras A-C, Honkanen RE (2011) Label-free quantitative proteomics and SAINT analysis enable interactome mapping for the human Ser/Thr protein phosphatase 5. *PROTEOMICS* **11**, 1508–1516.
203. Choi H, Glatter T, Gstaiger M, Nesvizhskii AI (2012) SAINT-MS1: Protein–Protein Interaction Scoring Using Label-free Intensity Data in Affinity Purification-Mass Spectrometry Experiments. *J Proteome Res* **11**, 2619–24.
204. Teo G, Koh H, Fermin D, Lambert J-P, Knight JDR, Gingras A-C, Choi H (2016) SAINTq: scoring protein-protein interactions in affinity purification – mass spectrometry experiments with fragment or peptide intensity data. *PROTEOMICS*, n/a-n/a.
205. Choi H, Liu G, Mellacheruvu D, Tyers M, Gingras A-C, Nesvizhskii AI (2012) Analyzing Protein-Protein Interactions from Affinity Purification-Mass Spectrometry Data with SAINT. In: *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi0815s39/abstract> [Accessed March 19, 2013].
206. Höhfeld J, Jentsch S (1997) GrpE-like regulation of the Hsc70 chaperone by the anti-apoptotic protein BAG-1. *EMBO J* **16**, 6209–16.
207. Doong H, Vrailas A, Kohn EC (2002) What’s in the ‘BAG’? – a functional domain analysis of the BAG-family proteins. *Cancer Lett* **188**, 25–32.
208. Briknarová K, Takayama S, Homma S, Baker K, Cabezas E, Hoyt DW, Li Z, Satterthwait AC, Ely KR (2002) BAG4/SODD Protein Contains a Short BAG Domain. *J Biol Chem* **277**, 31172–8.
209. Brockmann C, Leitner D, Labudde D, Diehl A, Sievert V, Büsow K, Kühne R, Oschkinat H (2004) The solution structure of the SODD BAG domain reveals additional electrostatic interactions in the HSP70 complexes of SODD subfamily BAG domains. *FEBS Lett* **558**, 101–6.
210. Arakawa A, Handa N, Ohsawa N, Shida M, Kigawa T, Hayashi F, Shirouzu M, Yokoyama S (2010) The C-Terminal BAG Domain of BAG5 Induces Conformational Changes of the Hsp70 Nucleotide- Binding Domain for ADP-ATP Exchange. *Structure* **18**, 309–19.

211. Vabulas RM, Raychaudhuri S, Hayer-Hartl M, Hartl FU (2010) Protein Folding in the Cytoplasm and the Heat Shock Response. *Cold Spring Harb Perspect Biol* **2**, a004390.
212. Va A, At G, Jn R, Je G (2013) Hsp70 protein complexes as drug targets. *Curr Pharm Des* **19**, 404, 404–17.
213. Caplan AJ (2003) What is a co-chaperone? *Cell Stress Chaperones* **8**, 105–7.
214. Radons J (2016) The human HSP70 family of chaperones: where do we stand? *Cell Stress Chaperones* **21**, 379–404.
215. Sondermann H, Scheufler C, Schneider C, Höhfeld J, Hartl F-U, Moarefi I (2001) Structure of a Bag/Hsc70 Complex: Convergent Functional Evolution of Hsp70 Nucleotide Exchange Factors. *Science* **291**, 1553–7.
216. Bracher A, Verghese J (2015) The nucleotide exchange factors of Hsp70 molecular chaperones. *Front Mol Biosci* **2**. doi:10.3389/fmolb.2015.00010.
217. Gehring U (2009) Multiple, but Concerted Cellular Activities of the Human Protein Hap46/BAG-1M and Isoforms. *Int J Mol Sci* **10**, 906–28.
218. Zeiner M, Niyaz Y, Gehring U (1999) The hsp70-associating protein Hap46 binds to DNA and stimulates transcription. *Proc Natl Acad Sci* **96**, 10194–9.
219. Niyaz Y, Zeiner M, Gehring U (2001) Transcriptional activation by the human Hsp70-associating protein Hap50. *J Cell Sci* **114**, 1839–45.
220. Lüders J, Demand J, Höhfeld J (2000) The Ubiquitin-related BAG-1 Provides a Link between the Molecular Chaperones Hsc70/Hsp70 and the Proteasome. *J Biol Chem* **275**, 4613–7.
221. Kanelakis KC, Morishima Y, Dittmar KD, Galigniana MD, Takayama S, Reed JC, Pratt WB (1999) Differential Effects of the hsp70-binding Protein BAG-1 on Glucocorticoid Receptor Folding by the hsp90-based Chaperone Machinery. *J Biol Chem* **274**, 34134–40.
222. Verghese J, Morano KA (2012) A Lysine-Rich Region within Fungal BAG Domain-Containing Proteins Mediates a Novel Association with Ribosomes. *Eukaryot Cell* **11**, 1003–11.
223. Elvekrog MM, Walter P (2015) Dynamics of co-translational protein targeting. *Curr Opin Chem Biol* **29**, 79–86.
224. Zhang X, Zhang F, Guo L, Wang Y, Zhang P, Wang R, Zhang N, Chen R (2013) Interactome Analysis Reveals that C1QBP (complement component 1, q

- subcomponent binding protein) Is Associated with Cancer Cell Chemotaxis and Metastasis. *Mol Cell Proteomics MCP* **12**, 3199–209.
225. Olins AL, Rhodes G, Welch DBM, Zwerger M, Olins DE (2010) Lamin B receptor. *Nucleus* **1**, 53–70.
226. Choudhary C, Poulsen JW, Sylvestersen KB, Povlsen LK, Nielsen ML, Mailand N, Beli P, Poulsen SL, et al (2012) Systems-wide analysis of ubiquitylation dynamics reveals a key role for PAF15 ubiquitylation in DNA-damage bypass. *Nat Cell Biol* **14**, 1089.
227. Altun M, Kramer HB, Willems LI, McDermott JL, Leach CA, Goldenberg SJ, Kumar KGS, Konietzny R, et al (2011) Activity-Based Chemical Proteomics Accelerates Inhibitor Development for Deubiquitylating Enzymes. *Chem Biol* **18**, 1401–12.
228. Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, Sowa ME, Rad R, et al (2011) Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. *Mol Cell* **44**, 325–40.
229. Song Z, Xu S, Song B, Zhang Q (2015) Bcl-2-Associated Athanogene 2 Prevents the Neurotoxicity of MPP+ via Interaction with DJ-1. *J Mol Neurosci* **55**, 798–802.
230. Che X, Tang B, Wang X, Chen D, Yan X, Jiang H, Shen L, Xu Q, et al (2013) The BAG2 protein stabilises PINK1 by decreasing its ubiquitination. *Biochem Biophys Res Commun* **441**, 488–92.
231. Song Z, Xu S, Song B, Zhang Q (2015) Bcl-2-Associated Athanogene 2 Prevents the Neurotoxicity of MPP+ via Interaction with DJ-1. *J Mol Neurosci* **55**, 798–802.
232. Che X-Q, Tang B-S, Wang H-F, Yan X-X, Jiang H, Shen L, Xu Q, Wang G-H, et al (2015) The BAG2 and BAG5 proteins inhibit the ubiquitination of pathogenic ataxin3-80Q. *Int J Neurosci* **125**, 390–4.
233. Yue X, Zhao Y, Liu J, Zhang C, Yu H, Wang J, Zheng T, Liu L, et al (2015) BAG2 promotes tumorigenesis through enhancing mutant p53 protein levels and function. *eLife* **4**, e08401.
234. McCollum AK, Casagrande G, Kohn EC (2010) Caught in the middle: the role of Bag3 in disease. *Biochem J* **425**, e1–3.
235. Boiani M, Daniel C, Liu X, Hogarty MD, Marnett LJ (2013) The Stress Protein BAG3 Stabilizes Mcl-1 Protein and Promotes Survival of Cancer Cells and Resistance to Antagonist ABT-737. *J Biol Chem* **288**, 6980–90.
236. Zhu H, Liu P, Li J (2012) BAG3: a new therapeutic target of human cancers? *Histol Histopathol* **27**, 257–61.

237. Falco A, Festa M, Basile A, Rosati A, Pascale M, Florenzano F, Nori SL, Nicolin V, et al (2012) BAG3 controls angiogenesis through regulation of ERK phosphorylation. *Oncogene* **31**, 5153–61.
238. Gentilella A, Passiatore G, Deshmane S, Turco MC, Khalili K (2008) Activation of BAG3 by Egr-1 in response to FGF-2 in neuroblastoma cells. *Oncogene* **27**, 5011–8.
239. Rosati A, Graziano V, De Laurenzi V, Pascale M, Turco MC (2011) BAG3: a multifaceted protein that regulates major cell pathways. *Cell Death Dis* **2**, e141.
240. Ulbricht A, Eppler FJ, Tapia VE, van der Ven PFM, Hampe N, Hersch N, Vakeel P, Stadel D, et al (2013) Cellular Mechanotransduction Relies on Tension-Induced and Chaperone-Assisted Autophagy. *Curr Biol* **23**, 430–5.
241. Ingham RJ, Colwill K, Howard C, Dettwiler S, Lim CSH, Yu J, Hersi K, Raaijmakers J, et al (2005) WW Domains Provide a Platform for the Assembly of Multiprotein Networks. *Mol Cell Biol* **25**, 7092–106.
242. Carra S (2009) The stress-inducible HspB8-Bag3 complex induces the eIF2alpha kinase pathway: implications for protein quality control and viral factory degradation? *Autophagy* **5**, 428–9.
243. Jiang Y, Woronicz JD, Liu W, Goeddel DV (1999) Prevention of Constitutive TNF Receptor 1 Signaling by Silencer of Death Domains. *Science* **283**, 543–6.
244. Annunziata CM, Kleinberg L, Davidson B, Berner A, Gius D, Tchabo N, Steinberg SM, Kohn EC (2007) BAG-4/SODD and Associated Antiapoptotic Proteins Are Linked to Aggressiveness of Epithelial Ovarian Cancer. *Clin Cancer Res* **13**, 6585–92.
245. Ozawa F, Friess H, Zimmermann A, Kleeff J, Büchler MW (2000) Enhanced Expression of Silencer of Death Domains (SODD/BAG-4) in Pancreatic Cancer. *Biochem Biophys Res Commun* **271**, 409–13.
246. Chen Y, Boland A, Kuzuoğlu-Öztürk D, Bawankar P, Loh B, Chang C-T, Weichenrieder O, Izaurralde E (2014) A DDX6-CNOT1 Complex and W-Binding Pockets in CNOT9 Reveal Direct Links between miRNA Target Recognition and Silencing. *Mol Cell* **54**, 737–50.
247. Rouya C, Siddiqui N, Morita M, Duchaine TF, Fabian MR, Sonenberg N (2014) Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1. *RNA* **20**, 1398–409.
248. Szczyrba J, Jung V, Beitzinger M, Nolte E, Wach S, Hart M, Sapich S, Wieshöfer M, et al (2017) Analysis of Argonaute Complex Bound mRNAs in DU145 Prostate

- Carcinoma Cells Reveals New miRNA Target Genes. *Prostate Cancer* **2017**. doi:10.1155/2017/4893921.
249. Yen FT, Roitel O, Bonnard L, Notet V, Pratte D, Stenger C, Magueur E, Bihain BE (2008) Lipolysis Stimulated Lipoprotein Receptor A NOVEL MOLECULAR LINK BETWEEN HYPERLIPIDEMIA, WEIGHT GAIN, AND ATHEROSCLEROSIS IN MICE. *J Biol Chem* **283**, 25650–9.
250. Shimada H, Satohisa S, Kohno T, Konno T, Takano K-I, Takahashi S, Hatakeyama T, Arimoto C, et al (2017) Downregulation of lipolysis-stimulated lipoprotein receptor promotes cell invasion via claudin-1-mediated matrix metalloproteinases in human endometrial cancer. *Oncol Lett* **14**, 6776–82.
251. Kalia LV, Kalia SK, Chau H, Lozano AM, Hyman BT, McLean PJ (2011) Ubiquitinylation of  $\alpha$ -Synuclein by Carboxyl Terminus Hsp70-Interacting Protein (CHIP) Is Regulated by Bcl-2-Associated Athanogene 5 (BAG5). *PLOS ONE* **6**, e14695.
252. Kalia SK, Lee S, Smith PD, Liu L, Crocker SJ, Thorarinsdottir TE, Glover JR, Fon EA, et al (2004) BAG5 Inhibits Parkin and Enhances Dopaminergic Neuron Degeneration. *Neuron* **44**, 931–45.
253. Wang X, Guo J, Fei E, Mu Y, He S, Che X, Tan J, Xia K, et al (2014) BAG5 Protects against Mitochondrial Oxidative Damage through Regulating PINK1 Degradation. *PLOS ONE* **9**, e86276.
254. Bruchmann A, Roller C, Walther TV, Schäfer G, Lehmusvaara S, Visakorpi T, Klocker H, Cato ACB, Maddalo D (2013) Bcl-2 associated athanogene 5 (Bag5) is overexpressed in prostate cancer and inhibits ER-stress induced apoptosis. *BMC Cancer* **13**, 96.
255. Mock J-Y, Chartron JW, Zaslaver M, Xu Y, Ye Y, Clemons WM (2015) Bag6 complex contains a minimal tail-anchor-targeting module and a mock BAG domain. *Proc Natl Acad Sci* **112**, 106–11.
256. Genest O, Doyle SM, Wickner S (2013) Protein rescue from aggregates by powerful molecular chaperone machines. *Nat Rev Mol Cell Biol* **14**, 617.
257. Mattoo RUH, Sharma SK, Priya S, Finka A, Goloubinoff P (2013) Hsp110 is a bona fide chaperone using ATP to unfold stable misfolded polypeptides and reciprocally collaborate with Hsp70 to solubilize protein aggregates. *J Biol Chem*, jbc.M113.479253.

258. Mateja A, Sharma A, Mariappan M, Hegde RS, Keenan RJ, Stefanovic S, Li X (2010) A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* **466**, 1120.
259. Binici J, Koch J (2014) BAG-6, a jack of all trades in health and disease. *Cell Mol Life Sci* **71**, 1829–37.
260. Minami R, Hayakawa A, Kagawa H, Yanagi Y, Yokosawa H, Kawahara H (2010) BAG-6 is essential for selective elimination of defective proteasomal substrates. *J Cell Biol*, jcb.200908092.
261. Tanguay RM, Hightower LE (2015) *The Big Book on Small Heat Shock Proteins* | | Springer Available at: [//www.springer.com/us/book/9783319160764](http://www.springer.com/us/book/9783319160764) [Accessed November 28, 2017].
262. Garrido C, Paul C, Seigneuric R, Kampinga HH (2012) The small heat shock proteins family: the long forgotten chaperones. *Int J Biochem Cell Biol* **44**, 1588–92.
263. Haslbeck M, Vierling E (2015) A First Line of Stress Defense: Small Heat Shock Proteins and Their Function in Protein Homeostasis. *J Mol Biol* **427**, 1537–48.
264. Kriehuber T, Rattei T, Weinmaier T, Bepperling A, Haslbeck M, Buchner J (2010) Independent evolution of the core domain and its flanking sequences in small heat shock proteins. *FASEB J* **24**, 3633–42.
265. Arrigo A-P (2013) Human small heat shock proteins: Protein interactomes of homo- and hetero-oligomeric complexes: An update. *FEBS Lett* **587**, 1959–69.
266. Almeida-Souza L, Asselbergh B, d'Ydewalle C, Moonens K, Goethals S, Winter V de, Azmi A, Irobi J, et al (2011) Small Heat-Shock Protein HSPB1 Mutants Stabilize Microtubules in Charcot-Marie-Tooth Neuropathy. *J Neurosci* **31**, 15320–8.
267. Hino M, Kurogi K, Okubo M-A, Murata-Hori M, Hosoya H (2000) Small Heat Shock Protein 27 (HSP27) Associates with Tubulin/Microtubules in HeLa Cells. *Biochem Biophys Res Commun* **271**, 164–9.
268. Guay J, Lambert H, Gingras-Breton G, Lavoie JN, Huot J, Landry J (1997) Regulation of actin filament dynamics by p38 map kinase-mediated phosphorylation of heat shock protein 27. *J Cell Sci* **110**, 357–68.
269. Hishiya A, Salman MN, Carra S, Kampinga HH, Takayama S (2011) BAG3 Directly Interacts with Mutated alphaB-Crystallin to Suppress Its Aggregation and Toxicity. *PLoS ONE* **6**, e16828.

270. Boelens WC, Croes Y, de Jong WW (2001) Interaction between  $\alpha$ B-crystallin and the human 20S proteasomal subunit C8/ $\alpha$ 7. *Biochim Biophys Acta BBA - Protein Struct Mol Enzymol* **1544**, 311–9.
271. Nakagawa M, Tsujimoto N, Nakagawa H, Iwaki T, Fukumaki Y, Iwaki A (2001) Association of HSPB2, a Member of the Small Heat Shock Protein Family, with Mitochondria. *Exp Cell Res* **271**, 161–8.
272. Vos MJ, Zijlstra MP, Kanon B, Waarde-Verhagen V, Awh M, Brunt ERP, Oosterveld-Hut HMJ, Carra S, et al (2010) HSPB7 is the most potent polyQ aggregation suppressor within the HSPB family of molecular chaperones. *Hum Mol Genet* **19**, 4677–93.
273. Juo L-Y, Liao W-C, Shih Y-L, Yang B-Y, Liu A-B, Yan Y-T (2016) HSPB7 interacts with dimerized FLNC and its absence results in progressive myopathy in skeletal muscles. *J Cell Sci*, jcs.179887.
274. Teixeira PF, Glaser E (2013) Processing peptidases in mitochondria and chloroplasts. *Biochim Biophys Acta BBA - Mol Cell Res* **1833**, 360–70.
275. de Wit NJW, Verschuure P, Kappé G, King SM, de Jong WW, van Muijen GNP, Boelens WC (2004) Testis-specific human small heat shock protein HSPB9 is a cancer/testis antigen, and potentially interacts with the dynein subunit TCTEL1. *Eur J Cell Biol* **83**, 337–45.
276. Kappé G, Verschuure P, Philipsen RLA, Staalduinen AA, Van de Boogaart P, Boelens WC, De Jong WW (2001) Characterization of two novel human small heat shock proteins: protein kinase-related HspB8 and testis-specific HspB9. *Biochim Biophys Acta BBA - Gene Struct Expr* **1520**, 1–6.
277. Yang K, Meinhardt A, Zhang B, Grzmil P, Adham IM, Hoyer-Fender S (2012) The Small Heat Shock Protein ODF1/HSPB10 Is Essential for Tight Linkage of Sperm Head to Tail and Male Fertility in Mice. *Mol Cell Biol* **32**, 216–25.
278. Baltz JM, Oneeka Williams P, Cone RA (1990) Dense Fibers Protect Mammalian Sperm Against Damage. *Biol Reprod* **43**, 485–91.
279. Alper JD, Decker F, Agana B, Howard J (2014) The Motility of Axonemal Dynein Is Regulated by the Tubulin Code. *Biophys J* **107**, 2872–80.
280. Naaby-Hansen S, Herr JC (2010) Heat shock proteins on the human sperm surface. *J Reprod Immunol* **84**, 32–40.

281. Dun MD, Aitken RJ, Nixon B (2012) The role of molecular chaperones in spermatogenesis and the post-testicular maturation of mammalian spermatozoa. *Hum Reprod Update* **18**, 420–35.
282. Esslinger U, Garnier S, Korniat A, Proust C, Kararigas G, Müller-Nurasyid M, Empana J-P, Morley MP, et al (2017) Exome-wide association study reveals novel susceptibility genes to sporadic dilated cardiomyopathy. *PLOS ONE* **12**, e0172995.
283. Charron P, Villard E, Cambien F (2012) In Vitro Diagnosis Method for Predicting a Predisposition to Cardiomyopathy. Available at: [http://patentscope.wipo.int/search/en/detail.jsf?docId=WO2012107580&recNum=5&docAn=EP2012052352&queryString=\(elisa%20AND%20method%20AND%20diagnostics%20and%20\(kit%20or%20kits\)\)&maxRec=35687](http://patentscope.wipo.int/search/en/detail.jsf?docId=WO2012107580&recNum=5&docAn=EP2012052352&queryString=(elisa%20AND%20method%20AND%20diagnostics%20and%20(kit%20or%20kits))&maxRec=35687) [Accessed April 3, 2013].
284. Chang W, Barnes AM, Cabral WA, Bodurtha JN, Marini JC (2010) Prolyl 3-hydroxylase 1 and CRTAP are mutually stabilizing in the endoplasmic reticulum collagen prolyl 3-hydroxylation complex. *Hum Mol Genet* **19**, 223–34.
285. Jaffer F, Murphy SM, Scotto M, Healy E, Rossor AM, Brandner S, Phadke R, Selcen D, et al (2012) BAG3 mutations: another cause of giant axonal neuropathy. *J Peripher Nerv Syst JPNS* **17**, 210–6.
286. Lian X, Hsiao C, Wilson G, Zhu K, Hazeltine LB, Azarin SM, Raval KK, Zhang J, et al (2012) Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc Natl Acad Sci*. doi:10.1073/pnas.1200250109.
287. Crawford ED, Seaman JE, Agard N, Hsu GW, Julien O, Mahrus S, Nguyen H, Shimbo K, et al (2013) The DegraBase: A Database of Proteolysis in Healthy and Apoptotic Human Cells. *Mol Cell Proteomics* **12**, 813–24.
288. Wiita AP, Ziv E, Wiita PJ, Urisman A, Julien O, Burlingame AL, Weissman JS, Wells JA (2013) Global cellular response to chemotherapy-induced apoptosis. *eLife* **2**. doi:10.7554/eLife.01236.
289. Sasse S, Brand NJ, Kyprianou P, Dhoot GK, Wade R, Arai M, Periasamy M, Yacoub MH, Barton PJ (1993) Troponin I gene expression during human cardiac development and in end-stage heart failure. *Circ Res* **72**, 932–8.



## Publishing Agreement

*It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.*

*I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.*

Author Signature \_\_\_\_\_

A handwritten signature in black ink, appearing to be 'W. S.', written over a horizontal line.

Date \_\_\_\_\_

12/22/17