

Application of Proteomics to Human Hypertrophic Cardiomyopathy

A thesis submitted for the degree of Doctor of Philosophy

December 2016

Institute of Cardiovascular Science, University College London

Supervisors: Professor Perry Elliott, Professor William McKenna

I, Caroline Coats confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed  _____

Date _____ 18th December 2016 _____

Abstract

This work describes a preliminary study to evaluate the use of proteomics in the study of human hypertrophic cardiomyopathy. Both gel and mass spectrometry techniques were used for the identification and analysis of myocardial proteins in whole tissue lysate. Qualitative and quantitative proteomic methods were used to understand disease mechanisms and identify and validate novel biomarkers.

Disease caused by mutations in beta myosin heavy chain, *MYH7* and myosin binding protein-C, *MYBPC3* were studied alongside patients where no genetic variant could be identified. Left ventricular septal myectomy samples showed changes in protein expression compared with control tissue. Novel mutations in *MYH7* were confirmed by identification of mutant peptide sequences. Disease mechanisms were investigated by studying interactions between up- and down-regulated proteins involved in various pathways. Enriched protein groups included those involved in cytoskeletal protein binding and energy production. Novel findings included the identification of carbonic anhydrase III in cardiomyocytes.

A targeted and multiplexed MRM-MS assay was developed to validate potential biomarkers in tissue and correlated with clinical phenotype. The assay was further applied to screen these biomarkers in urine. Novel findings included increased expression of lumican, a small leucine-rich proteoglycan that controls the assembly of collagen fibres in the extracellular matrix. Lumican concentration was highest in a sub-group of patients with evidence of scarring on cardiac magnetic resonance imaging, making it a potential marker of progressive disease.

This is the first comprehensive global proteomic study of human hypertrophic cardiomyopathy. It has identified differences in the expression of several proteins in the myocardium not described previously, but highly relevant to pathophysiology of this disease. A targeted proteomic translational assay, capable of quantitating 35 peptides in less than 10 minutes has been developed.

Abbreviations

1D	One dimensional
2D	Two dimensional
2D-E	Two-dimensional electrophoresis
2D-DIGE	Two-dimensional fluorescence difference gel electrophoresis
2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
α -gal A	Alpha-galactosidase A
AmBic	Ammonium Bicarbonate
ACN	Acetonitrile
AD	Autosomal dominant
ADP	Adenosine diphosphate
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
AR	Autosomal recessive
AS	Aortic stenosis
ASB-14	Amidosulfobetaine14 (ASB-14)
ASD	Atrial septal defect
AST	Aspartate amino transferase
BCA	Bicinchoninic acid
BNP	Brain natriuretic peptide
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
cDNA	Complementary DNA
Ca ²⁺	Calcium
C-18	Octadecylsilane
CE	Capillary Electrophoresis
CHAPS	3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate hydrate
CID	Collision induced dissociation
CMR	Cardiac magnetic resonance
CPK	Creatine phosphokinase
CRP	C-reactive protein
dpi	Dots per inch
Da	Daltons
DCM	Dilated cardiomyopathy
DNA	Deoxyribonucleic acid
DTE	Dithioerythritol
DTT	Dithiothreitol
EBI	European Bioinformatics Institute
ECG	Electrocardiogram
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionisation
ETC	Electron transport chain
FA	Formic acid
FFA	Free fatty acids

FFPE	Formalin fixed paraffin embedded
FGP	Fetal gene program
FWHM	Full width half maximum
GSD	Glycogen storage disease
H&E	Haematoxylin and eosin
HCl	Hydrochloric acid
HCM	Hypertrophic cardiomyopathy
HPLC	High performance liquid chromatography
HSP	Heat shock proteins
IAA	Indole-3-Acetic Acid
ICD	Implantable cardiac defibrillator
IEF	Isoelectric focusing
IL-6	Interleukin 6
IPA	Ingenuity pathway analysis
IPG	Immobilized pH gradient
IQR	Interquartile range
IS	Internal standard
kDa	Kilo Daltons
kV	Kilo Volts
LC	Liquid chromatography
LC-MS	Liquid chromatography mass spectrometry
LDH	Lactate dehydrogenase
LGE	Late gadolinium enhancement
LOD	Limits of detection
LV	Left ventricular
LVH	Left ventricular hypertrophy
LVOTO	Left ventricular outflow tract obstruction
m/z	Mass to charge ratio
mA	Milli Amps
min	Minutes
MALDI	Matrix assisted laser desorption/ionization
MMP	Matrix metalloproteinases
MyBP-C	Cardiac myosin binding protein-C
MW	Molecular weight
mRNA	Messenger RNA
miRNA	Micro RNA
MRM	Multiple reaction monitoring
MRS	Magnetic resonance spectroscopy
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
mtDNA	Mitochondrial DNA
NHS	National Health Service
NT-proBNP	N-terminal pro brain natriuretic peptide
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCr	Phosphocreatine
PCR	Polymerase chain reaction
pI	Isoelectric point

PICP	C-terminal propeptide of collagen type I
PINP	N-terminal propeptide of collagen type I
PIIICP	C-terminal propeptide of collagen type III
PIIINP	N-terminal propeptide of collagen type III
PLGS	ProteinLynx Global SERVER™
PTM	Post-translational modification
QC	Quality control
Q-ToF	Quadrupole time of flight
R&D	Research and Development
RF	Radiofrequency
RNA	Ribonucleic acid
ROC	Receiver operating characteristics
ROS	Reactive oxygen species
RT	Retention time
RT-qPCR	Real-time reverse transcription-PCR
RV	Right ventricular
RVH	Right ventricular hypertrophy
Sec	Second
SCD	Sudden cardiac death
SCX	Strong cation exchange resin
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SRM	Selected reaction monitoring
TCA	Tricarboxylic acid
TIMP	Tissue inhibitors of metalloproteinases
TFA	Trifluoroacetic acid
TMA	Tissue microarray
TNF- α	Tissue necrosis factor alpha
ToF	Time of Flight
UPLC	Ultra High Performance Liquid Chromatography
V	Volts

Contents

Abbreviations.....	4
Acknowledgements.....	13
Personal Contribution	15
Funding and Grants.....	17
Abstracts and Publications.....	18
1 Aims of Thesis.....	19
2 Background.....	22
2.1 Hypertrophic cardiomyopathy	22
2.2 Proteomics.....	48
2.3 Cardiovascular Proteomics.....	70
2.4 Cardiovascular Biomarkers	75
3 Materials and Methods.....	81
3.1 Materials.....	81
3.2 Proteomics Methods.....	82
3.3 Histopathology	95
3.4 Clinical Evaluation.....	97
3.5 Statistical Analysis	98
4 Analysis of the Cardiac Proteome	99
4.1 Introduction.....	99
4.2 Comparison of proteomic profiling techniques.....	99
4.3 Exploration of potential pre-analytical variation	111
4.4 Summary	114
5 The Proteome of Human HCM.....	115
5.1 Methods.....	115
5.2 Results	115
5.3 Discussion	132
5.4 Summary	139
6 Validation and Verification of Proteins	140
6.1 Study design and rationale	140
6.2 Development of a multiplexed biomarker assay	141
6.3 Results	147
6.4 Validation and Relevance of Specific Proteins	150

6.5	Conclusion.....	170
7	Genomics and Proteomics	171
7.1	Genotype-phenotype relationship in HCM	171
7.2	Exploration of a genotype specific proteome.....	173
7.3	Application of next generation sequencing.....	178
7.4	Identification of mutant peptides by MS.....	183
7.5	Conclusion.....	188
8	Clinical Application of Novel Biomarkers	189
8.1	Visualisation of proteins in myocardium	189
8.2	Screening biomarkers in urine	195
8.3	Correlation with clinical parameters	198
8.4	Conclusion.....	202
9	Discussion.....	203
9.1	Proteome coverage	203
9.2	Differences between HCM and Controls	206
9.3	Biomarker Discovery	209
9.4	Future work	212
10	Conclusion.....	215
11	References	216
12	Appendix.....	239

List of Figures

Figure 1.1 Overall project design.....	20
Figure 1.2 Flow diagram summarising laboratory and bioinformatics methods	21
Figure 2.1 The clinical phenotype of HCM	29
Figure 2.2. Macroscopic features of hypertrophic cardiomyopathy.	31
Figure 2.3 Microscopic features of hypertrophic cardiomyopathy	32
Figure 2.4 Cardiac magnetic resonance imaging in HCM.....	34
Figure 2.5 Composition of the sarcomere.	37
Figure 2.6 Complexity of biological information	44
Figure 2.7 Proteomics methods and applications.....	51
Figure 2.8 Overview of quantitative proteomics methods	53
Figure 2.9 2D-DIGE Method	56
Figure 2.10 Schematic view of a generalised mass spectrometer	59
Figure 2.11 Schematic representation of an electrospray ionisation (ESI) source. ...	62
Figure 2.12 Schematic representation of a Q-ToF mass spectrometer	64
Figure 2.13 Example of an MS/MS spectrum.....	65
Figure 2.14 Schematic representation of a tandem mass spectrometer	67
Figure 2.15 The key stages of biomarker development	76
Figure 2.16 Prognostic role of NT-pro BNP in HCM.....	79
Figure 4.1 Description of study cohort	100
Figure 4.2 Illustration of three techniques to investigate the HCM proteome.....	102
Figure 4.3 2D-DIGE map of the pH 3-10 region of human myocardium.....	104
Figure 4.4 Coomassie blue stained 1D SDS-PAGE of heart homogenate.....	106
Figure 4.5 Venn diagram representing overlap in protein identifications.	109
Figure 4.6 The top 20 most abundant proteins in heart.....	110

Figure 4.7 Protein abundance after delayed freezing	112
Figure 4.8 Effect of sub-optimal sample storage on number of proteins identified	114
Figure 5.1 Summary of number of proteins identified in heart samples.....	116
Figure 5.2 Bar chart illustrating proteins identified per fraction	116
Figure 5.3 The proteome of human HCM.....	118
Figure 5.4 Proteins identified according to GO protein class	119
Figure 5.5 Proteins identified according to GO molecular function	120
Figure 5.6 Proteins identified according to GO sub-cellular compartment	121
Figure 5.7 Proteins identified according to GO biological function	121
Figure 5.8 Principle component analysis of myocardial protein expression.	123
Figure 5.9 Proteins detected as showing >2-fold alteration in heart tissue.....	124
Figure 5.10 Significantly altered protein expression in HCM	127
Figure 5.11 Significantly enriched canonical pathways identified by IPA.....	129
Figure 5.12 Proteins involved in canonical pathway "mitochondrial dysfunction".	131
Figure 5.13 The electron transport chain	135
Figure 5.14 Proteins involved in canonical pathway "protein ubiquination"	138
Figure 6.1 Validation study design	140
Figure 6.2 Overlaid chromatogram of the internal standard peptides.....	143
Figure 6.3 Typical UPLC MS/MS chromatogram.....	144
Figure 6.4 MRM transitions of cardiac and skeletal actin.	151
Figure 6.5 Percentage of cardiac actin detected in myocardial samples.....	152
Figure 6.6 FHL1 expression in HCM	153
Figure 6.7 Desmin expression in HCM.....	155
Figure 6.8 Spectrum of peptide marker for carbonic anhydrase III.....	158
Figure 6.9 Carbonic anhydrase III expression in HCM	159
Figure 6.10 Correlation between two unique peptide markers for lumican.....	160

Figure 6.11 Lumican expression in HCM.....	161
Figure 6.12 Phosphoglycerate mutase expression in HCM	163
Figure 6.13 Aldolase expression in HCM.....	164
Figure 6.14 Creatinine kinase M-type in HCM	166
Figure 6.15 Gelsolin expression in HCM	168
Figure 6.16 Transthyretin expression in HCM	170
Figure 7.1 Expression of haemoglobin subunits in myocardial samples.....	175
Figure 7.2 Haemoglobin concentration in the discovery cohort.	175
Figure 7.3 Protein expression according to genetic variants.....	180
Figure 7.4 Myosin binding protein C secondary structure.....	181
Figure 7.5 Myosin binding protein C expression in HCM.....	182
Figure 7.6 Spectra from mutant and normal peptide in V606M mutation.....	186
Figure 7.7 Spectra from mutant and normal peptide in E1555G mutation.....	187
Figure 8.1 Design of tissue microarray	190
Figure 8.2 Annexin staining in HCM.....	191
Figure 8.3 Histological validation of CA III in HCM.....	192
Figure 8.4 Localisation of CA III within cytoplasm of cardiomyocytes	192
Figure 8.5 Quantification of CA III staining intensity	193
Figure 8.6 Immunohistochemical validation of myocardial proteins	194
Figure 8.7 Biomarker screening in urine using UPLC-MS/MS assay	195
Figure 8.8 Targeted MRM LC-MS/MS analysis in HCM urine samples	196
Figure 8.9 Lumican concentration and fibrosis on CMR.....	199

List of Tables

Table 2.1 Genes associated with familial hypertrophic cardiomyopathy.	24
Table 2.2 Genetic diseases and syndromes associated with HCM	25
Table 2.3 Summary of transcriptomic studies relevant to HCM.	46
Table 3.1 LC inlet parameters.....	92
Table 3.2 Instrument settings for optimising peptide tuning	93
Table 4.1 Characteristics of study cohort.....	100
Table 4.2 Protein spots \geq 2-fold difference between HCM and Control	105
Table 4.3 Comparison of proteome separation methods.....	107
Table 5.1 Proteins up-regulated $>$ 2-fold in HCM.....	125
Table 5.2 Proteins down-regulated $>$ 2-fold in HCM	126
Table 5.3 Table summarising the number of proteins within each complex.	135
Table 6.1 The internal standard optimised for UPLC-MS/MS assay	143
Table 6.2 MRM transitions for the UPLC-MS/MS analysis of heart proteins	146
Table 6.3 Characteristics of patients included in the validation study.....	148
Table 6.4 Comparison of targeted MS/MS with LC-MS analysis.....	149
Table 6.5 Unique peptides identified by LC-MS that correlated with CA III	157
Table 7.1 Proteins $>$ 1.5 fold higher in patients with MYBPC3 mutation.....	176
Table 7.2 Proteins $>$ 1.5 fold higher in patients with MYH7 mutation	177
Table 7.3 Prediction of pathogenicity for the E1555G mutation in MYH7.....	184
Table 7.4 Tryptic peptides related to E1555G mutation in MYH7.....	185
Table 8.1 Working dilutions of antibodies used and product codes	189
Table 8.2 Potential urinary biomarkers in HCM.....	197

Acknowledgements

This work would not have been possible without the advice, support and collaboration of many individuals. Firstly, I would like to thank my supervisors: Professor Perry Elliott and Professor William McKenna, who have encouraged and supported me throughout my research, and focussed my energies when needed. Professor Michael Frenneaux has been an external mentor and fantastic sounding board. I am indebted to Dr Kevin Mills, Senior Lecturer in Biochemistry and Dr Wendy Heywood, Research Associate in Translational Biomarker Discovery, who provided the day to day guidance, training, and facilities to carry out this work. Their expertise, sincerity and enthusiasm have underpinned the project.

Many other individuals at the Institute of Child Health were influential in training me and facilitating specific scientific aspects of this work including: Dr Sarah Pressey, Dr Vicky Manwaring, Mr Ernestas Sirka, Miss Emily Bliss, Dr James McCormick and Mr Ivan Doykov (Biological mass spectrometry Unit); Professor Neil Sebire, Dr Michael Ashworth, Dr Jeremy Pryce, Dr Andrew Bamber and Mr Alex Virsami (Histopathology); Professor Paul Riley and Dr Louisa Petchey (Molecular medicine unit); Dr Mike Hubank (Genetics).

The Inherited Heart Disease Unit at UCLH formed the backbone of my clinical research into hypertrophic cardiomyopathy and I am thankful to the many fellows, nurses, genetic counsellors and clerical staff whom I have worked with over the years. Particularly important to this project were: Ms Shaughan Dickie, Dr Antonis Pantazis, Dr Maite Tome, Dr Juan Gimeno and Dr Vimal Patel (Medical); Professor Christopher McGregor (Surgery); Dr Petros Syrris, Dr Luis Lopes, Ms Sharon Jenkins and Miss Demetra Georgiou (Genetics); Ms Linda Moss, Mr Simon Waller, Miss Myra O'Donovan and Mrs Joanna Mander (Nursing); Professor James Moon, and Dr Silvia Castelletti (Imaging) and Dr Anne Dawnay (Clinical Biochemistry).

All translational research relies on the altruistic involvement of patients. The human tissue and urine samples were obtained from individuals undergoing cardiac surgery or other treatments at the Heart Hospital and I am hugely thankful to them for their readiness to participate. Professor Christobel dos Remedios, University of Sydney,

collaborated by providing unused healthy donor heart samples for comparison. Professor James Moon and Dr Thomas Treibel provided aortic stenosis specimens which enabled further validation and hypothesis generation.

Last but not least, I owe much to my family, friends and husband John for their love, support and encouragement throughout this journey.

Personal Contribution

The basis of this thesis is an exploratory study using proteomics to identify potential biomarkers in hypertrophic cardiomyopathy. The project was conceptualised by Professor Perry Elliott and Dr Kevin Mills and a preliminary study performed by Dr Wendy Heywood. I was responsible for the subsequent conduct of all aspects of the study, including obtaining ethical and R&D approval to study biomarkers in patients with inherited heart disease. I took charge of collection, processing and storage of all clinical data, and human tissue, blood and urine samples in accordance with the Data Protection Act 1998 and the Human Tissue Act 2004. I completed Good Clinical Practice training to comply with the NHS Research Governance framework. I performed all laboratory work and bioinformatics analyses myself.

CHAPTER 2:

As part of the background to this work, I conducted a study to evaluate the clinical role of brain natriuretic peptide (BNP), an established cardiovascular biomarker which had been relatively unexplored in hypertrophic cardiomyopathy. Serum N-terminal BNP was measured by a two-site electrochemiluminescence immunoassay on a Roche E170 analyser in collaboration with Dr Anne Dawney and the Clinical Biochemistry department at University College London Hospitals.

CHAPTER 3:

The laboratory methodology was an evolution of previous proteomics work conducted in Dr Mill's laboratory. I adapted standard protocols for myocardial tissue in accordance with published literature and recommendations from experts in the field. I performed all the wet laboratory work myself, including protein separation, quantification, and digestion. In the mass spectrometry laboratory, I performed the tuning and optimisation of synthetic peptides and developed the multiplexed assay. I was taught the practical techniques by Dr Wendy Heywood under the supervision of Dr Kevin Mills. I attended the Wellcome Trust Advanced Course in Proteomics Bioinformatics which was invaluable in obtaining skills to perform data analyses, which I did with guidance from Dr Wendy Heywood.

CHAPTER 7

Genotyping was performed in collaboration with Dr Petros Syrris and Dr Mike Hubank, in all patients who provided consent, using a next generation sequencing platform designed at UCL to screen 41 genes associated with cardiomyopathy and inherited arrhythmia. Bioinformatic analyses and in silico prediction of pathogenicity were conducted by Dr Vincent Plagnol and Dr Luis Lopes who taught me about the use of variant effect prediction tools.

CHAPTER 8

I developed the concept of creating a tissue array to perform immunohistochemistry and obtained additional funding for this. I reviewed all histology and selected cases for inclusion with Dr Michael Ashworth, cardiac histopathologist. The microarray was constructed by Dr Jeremy Pryce and Mr Alex Virsami who assisted me with antibody optimisation. I collated all clinical data from patients' medical records. I performed the urinary proteomics with Miss Si Wang as part of her MSc Res project under supervision of Dr Kevin Mills. Creatinine quantification was performed by Mr Ernestas Sirka.

Funding and Grants

The British Heart Foundation supported this work, by funding my salary for a three year Clinical Research Training Fellowship (FS/10/027/28248). During the project I was awarded additional funding to progress this research:

1. BRITISH CARDIOVASCULAR SOCIETY TRAVEL AWARD 2012, £500

This funding allowed me to attend The European Society of Cardiology Meeting in Munich in 2012, where I presented my data on the role of natriuretic peptides in hypertrophic cardiomyopathy.

2. ROSETREES TRUST PROJECT GRANT 2013, 1 YEAR, £5,000

CJ COATS, PM ELLIOTT, MA ASHWORTH

This grant subsidised the cost of constructing a tissue microarray and purchasing antibodies for immunohistochemistry, in order to validate potential biomarkers identified from the proteomic study.

3. GOSH CHILDREN'S CHARITY STARTER GRANT, 2014, 2 YEARS, £70,000

J KASKI, CJ COATS, PM ELLIOTT, K MILLS.

This grant has enabled the continuation of my work, by extending the study population to children with pre-phenotypic disease. It will establish a bio-resource for paediatric cardiomyopathy, providing a base for longitudinal biomarker studies.

Abstracts and Publications

The work submitted in this thesis has been presented at two national and one international meeting to date. I have been invited to give two lectures and a write a review article on the subject:

1. Coats CJ, Gallagher M, Foley M, O'Mahony C, Critoph C, Gimeno J, Dawnay A, McKenna WJ, Elliott PM, Relation between serum N-terminal pro brain natriuretic peptide and prognosis in patients with hypertrophic cardiomyopathy. *Eur Heart J.* 2013;34(32):2529-37. Also presented as an Abstract at the ESC Congress.
2. Coats CJ, Heywood WE, McKenna WJ, Frenneaux MP, Mills K, Elliott P. Novel Insights into the Pathophysiology of Hypertrophic Cardiomyopathy Using Label free Quantitative Proteomics. Abstract published in *Heart.* 2014 Jun;100 Suppl 3:A89-90.
3. Coats CJ, "Biomarkers in Cardiomyopathy – What's New?" Cardiomyopathy Association's Annual Medical Conference, London, May 2014.
4. Coats CJ, "Biomarkers: Past, Present and Future", by invitation, The Royal Society of Medicine Cardiology Training Day, November 2014.
5. Coats CJ, Heywood WE, Mills K, Elliott PM, Current Applications of Biomarkers in Cardiomyopathies, *Expert Rev. Cardiovasc. Ther.* 13(7), 2015.

Two manuscripts describing a) the global proteomic profiling results and b) the identification of lumican and its role as a potential clinical biomarker to measure disease progression in hypertrophic cardiomyopathy have been submitted.

1 Aims of Thesis

This is the first human study using mass spectrometry based proteomics in hypertrophic cardiomyopathy (HCM). The specific aims are to:

1. Use proteomic methodology to understand disease mechanisms in HCM
2. Use proteomic methodology to identify potential biomarkers in HCM
3. Develop and validate quantitative assays for any potential biomarkers.

The overall project design is shown in Figure 1.1 and Figure 1.2.

In the “discovery” phase, a small cohort of samples was used to perform protein expression profiling and identify potential biomarkers. This included development and optimisation of methods for high resolution detection, identification, and analysis of proteins in heart tissue. In the “validation” phase, a larger cohort of tissue samples was used and a targeted approach applied to quantify these proteins in patients with HCM of known and unknown genetic aetiology.

There are two main expected outcomes from proteomic analyses of heart tissue. The first is to decipher mechanisms underlying development of disease. The second is to discover molecular markers for diagnosis and staging of HCM. Such data could provide the knowledge base for the identification development of innovative treatment strategies.

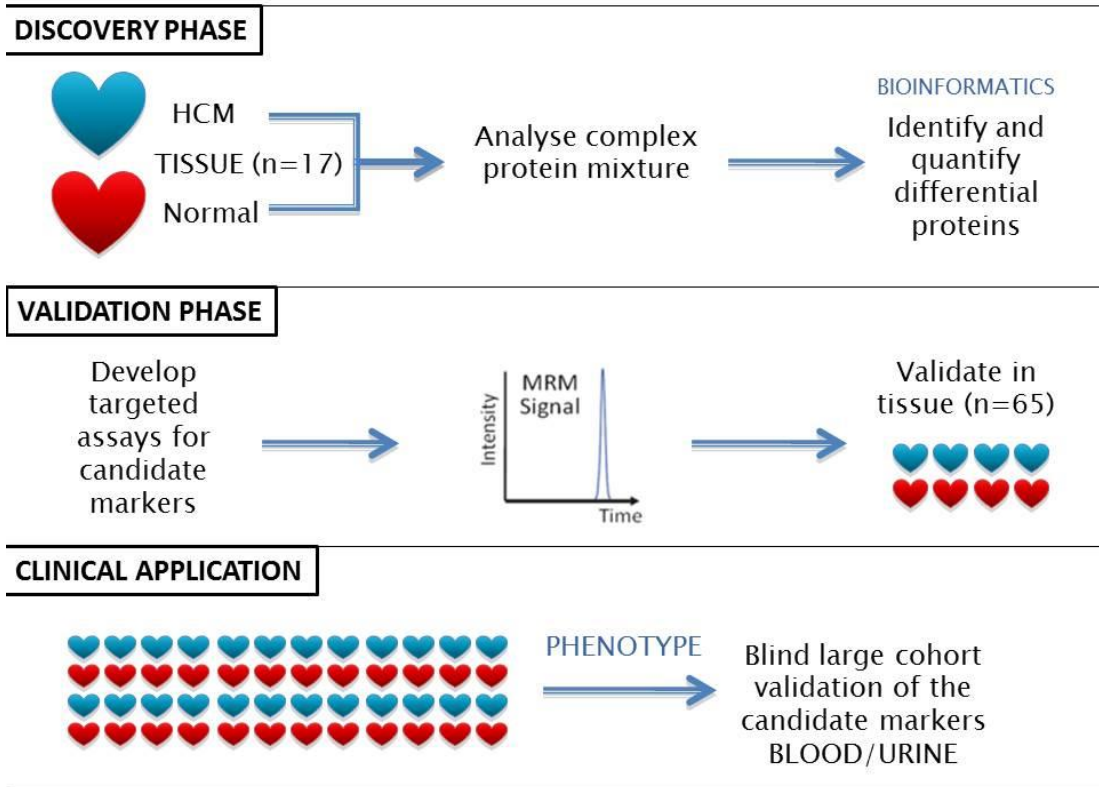


Figure 1.1 Overall project design

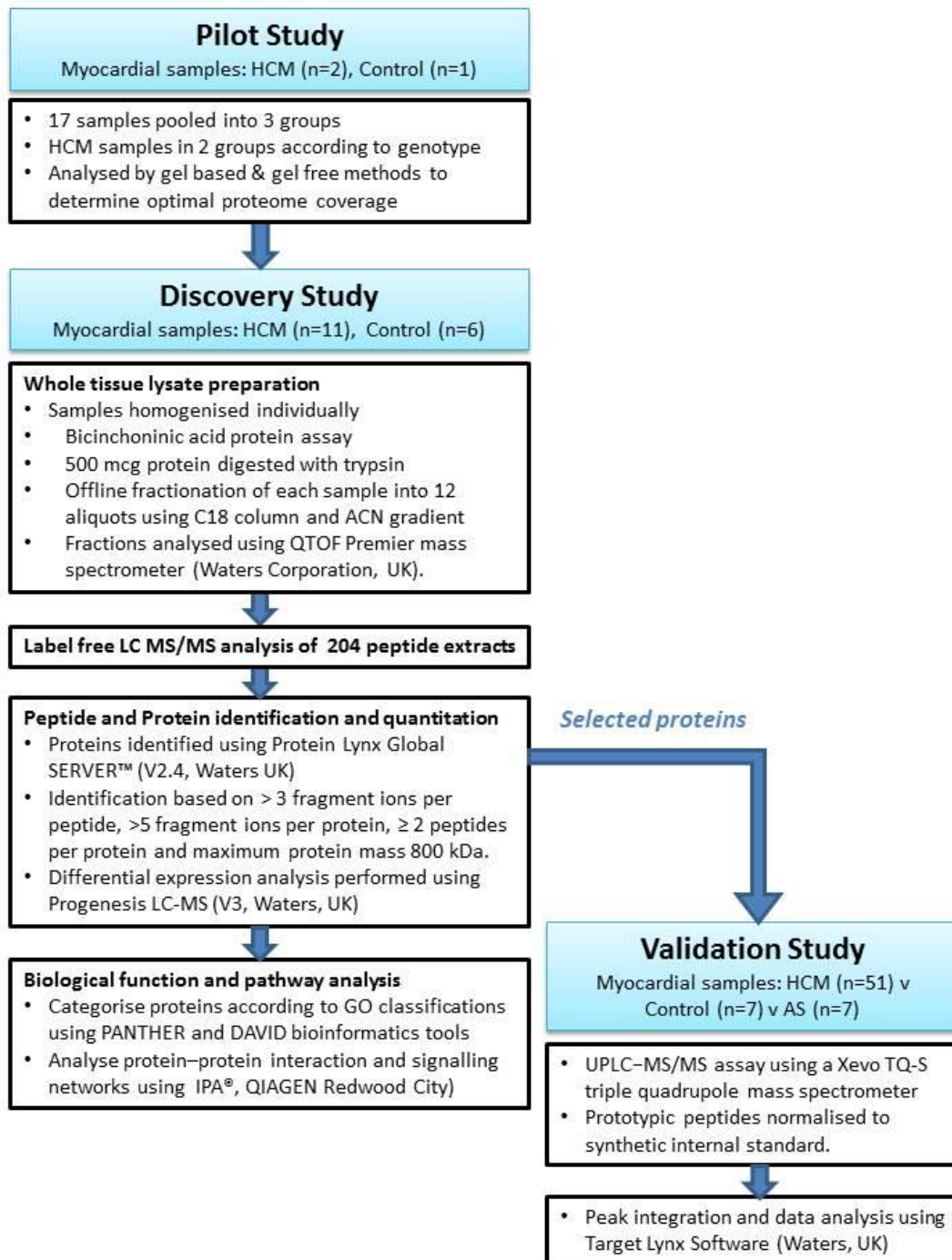


Figure 1.2 Flow diagram summarising laboratory and bioinformatics methods

2 Background

2.1 Hypertrophic cardiomyopathy

Hypertrophic cardiomyopathy (HCM) is an autosomal dominant inherited heart muscle disorder with a prevalence of 1 in 500 (Maron, Gardin et al. 1995). Clinically it is defined as left ventricular hypertrophy (LVH) unexplained by loading conditions (Elliott, Andersson et al. 2008). The disease may be discovered incidentally or following presentation with symptoms of breathlessness, chest pain, palpitations or syncope. HCM is an important cause of sudden cardiac death in young people and may be complicated by arrhythmias and premature heart failure (Maron and Maron 2013).

Over the past fifty years clinical management has focussed on strategies to protect those at risk of sudden death and the treatment of outflow tract obstruction caused by septal hypertrophy (Maron, Olivotto et al. 2003, Elliott, Gimeno et al. 2006, Maron, Spirito et al. 2007). Implantable cardiac defibrillators (ICD) and septal reduction procedures have contributed to improved survival (Ommen, Maron et al. 2005, Elliott, Gimeno et al. 2006). There is currently no treatment which specifically targets the underlying molecular or genetic defect and no pharmacologic therapy has been shown to alter the prognosis of HCM.

2.1.1 Genetics of HCM

As the first inherited heart disease to be genetically characterised, HCM has been a prototype for cardiovascular research. The first genetic locus associated with HCM was discovered by linkage analysis in 1989 and a responsible missense mutation in the beta myosin heavy chain gene, *MYH7* later identified (Jarcho, McKenna et al. 1989, Geisterfer-Lowrance, Kass et al. 1990). Over 1400 mutations, in a variety of genes, have been documented in the ensuing years (Seidman and Seidman 2011).

2.1.1.1 A disease of the sarcomere

Most disease-causing genes encode for proteins of the cardiac sarcomere. Mutations in genes encoding beta-myosin heavy chain, *MYH7*, and myosin-binding protein C, *MYBPC3*, account for the majority of human disease. Less commonly mutations are

identified in cardiac troponin I, *TNNI3*, cardiac troponin T, *TNNT2* and alpha-tropomyosin, *TPM1*. Rarely familial disease can be attributed to mutations in genes coding for Z-disc proteins or proteins involved in cellular calcium cycling.

Table 2.1 lists all genes associated with familial HCM published to date on OMIM®, the Online Mendelian Inheritance in Man database (Coats and Elliott 2013). Although the published prevalence varies according to the population studied, a mutation in a cardiac sarcomere protein gene can be found in approximately 60% of patients with familial HCM and 40% of those with seemingly sporadic disease (Ho and Seidman 2006, Morita, Rehm et al. 2008). Even though access to resources is variable across the world, genetic testing is recommended as part of routine clinical evaluation (Charron, Arad et al. 2010, Gersh, Maron et al. 2011, Elliott, Anastasakis et al. 2014).

2.1.1.2 Genetic complexity and heterogeneity

Advances in sequencing technology and systematic family screening have facilitated early recognition of the disease and understanding of its natural history. But despite whole exome and whole genome studies, the genetic causes of HCM remains incompletely resolved. While genotype–phenotype associations exist, the gene mutation alone fails to explain the wide disease spectrum. For example the same Arg145Trp missense mutation in *TNNI3* has been associated with phenotypes ranging from severe biventricular hypertrophy to minor ECG abnormalities in non-hypertrophic hearts (Mogensen, Murphy et al. 2004). This variable clinical expression within and between families defies any ostensible mutation or gene-specific disease pattern. Heterogeneity has largely been explained by incomplete penetrance, variable expression and multi-genic inheritance.

There are also a number of important inherited multisystem disorders and syndromes associated with HCM where disease is typically not confined to the heart e.g. neuromuscular, metabolic and mitochondrial diseases. The commonest are listed in Table 2.2 with their genetic cause and mode of inheritance. These have been important in understanding disease pathophysiology and are fundamental to developing final common pathway hypotheses (see 2.1.3.7).

Protein	Gene	Location	Freq.
Myosin-7 or beta myosin heavy chain	<i>MYH7</i>	14q11.2	40%
Myosin-binding protein C, cardiac-type	<i>MYBPC3</i>	11p11.2	40%
Troponin T, cardiac	<i>TNNT2</i>	1q32.1	3-5%
Troponin I, cardiac	<i>TNNI3</i>	19q.13.42	<5%
Tropomyosin 1 alpha chain	<i>TPM1</i>	15q22.1	<5%
Less common:			
Troponin C, slow skeletal and cardiac muscles	<i>TNNC1</i>	3p21.1	1%
Myosin regulatory light chain 2	<i>MYL2</i>	12q23-q24	<1%
Myosin essential light chain	<i>MYL3</i>	3p21.3	<1%
Actin, alpha cardiac muscle 1	<i>ACTC1</i>	15q11-q14	<1%
Cysteine and glycine-rich protein 3	<i>CSRP3</i>	11p15.1	<1%
Vinculin	<i>VCL</i>	10q22.2	<1%
Myozenin	<i>MYOZ2</i>	4q26	<1%
Myosin-6	<i>MYH6</i>	14q11.2	<1%
Titin	<i>TTN</i>	2q31.2	<1%
Junctophilin 2	<i>JPH2</i>	20q13.12	<1%
Caveolin 3	<i>CAV3</i>	3p25.3	<1%
Phospholamban	<i>PLN</i>	6q22.31	<1%
Calreticulin-3	<i>CALR3</i>	19p13.11	<1%
Nexilin	<i>NEXN</i>	1p31.1	<1%
AMP activated protein kinase subunit gamma 2	<i>PRKAG2</i>	7q36.1	<1%
Solute carrier family 25	<i>SLC25A4</i>	4q35.1	<1%
Myosin light chain kinase 2	<i>MYLK2</i>	20q11.21	<1%
Unknown*	<i>unknown</i>	7p12.1-q21	<1%
Calsequestrin †	<i>CASQ2</i>	1p13.1	<1%
Telethonin †	<i>TCAP</i>	17q12	<1%
α-actinin-2 †	<i>ACTN2</i>	1q43	<1%

Table 2.1. Genes associated with familial hypertrophic cardiomyopathy.

The molecular basis of the disease is known in all cases except. (OMIM® Phenotypic series, 192600). Genes reported in the literature but not on OMIM® are marked †*

Disease	Gene	Inheritance
DISORDERS OF CARBOHYDRATE METABOLISM		
Glycogen storage disease (GSD) Type II (Pompe's)	GAA	AR
GSD Type III (Cori or Forbes disease)	AGL	AR
GSD Type IX (cardiac phosphorylase kinase deficiency)	PHKB, PHKG2 PHKA1, PHKA2	AR X-linked
Congenital disorder of glycosylation Type 1a	PMM2	AR
DISORDERS OF AMINO ACID METABOLISM		
Type 1 Tyrosinaemia	FAH	AR
Dihydrolipoamide dehydrogenase deficiency	DLD	AR
DISORDERS OF FATTY ACID METABOLISM		
Very long chain acyl Co-A dehydrogenase deficiency (VLCHAD)	ACAVLD	AR

Malonyl-CoA Decarboxylase Deficiency	MLYCD	AR
Systemic Primary Carnitine Deficiency	SLC22A5	AR
LYSOSOMAL STORAGE DISEASES		
Fabry Disease (alpha galactosidase deficiency)	GLA	AD
Mucopolidosis type II alpha/beta	GNPTAB	AR
Mucopolysaccharidosis Type VII (Sly Syndrome)	GUSB	AR
Danon disease	LAMP2	X-linked
Gangliosidosis Type 1	GLB1	AR
MITOCHONDRIAL DISORDERS		
Complex I, II, III, IV and V deficiency	mtDNA	maternal
ACAD9 deficiency (complex I)	ACAD9	AR
ATP Synthase deficiency (complex V)	ATPAF2, TMEM70	AR
Cytochrome C oxidase deficiency (complex IV)	COX6B1	AR
Combined oxidative phosphorylation deficiency (Types 3, 5, 8, 9 and 10)	MTO1, AARS2, TSFM, MRPS22, MRPL3, GFM1	AR
3-methylglutaconic aciduria (Barth syndrome)	TAZ	X-linked
Kearns-Sayre Syndrome	mtDNA	maternal
Leber's hereditary optic neuropathy	mtDNA	maternal
Sengers syndrome	AGK	AR
Leigh Syndrome (pyruvate dehydrogenase complex deficiency)	mtDNA SURF1 PDHA1	maternal AR X-linked
Friedrichs ataxia	FXN	AR
Sengers syndrome	AGK	AR
Pyruvate dehydrogenase lipoic acid synthetase deficiency	LIAS	AR
Primary Coenzyme Q10 deficiency	COQ2	AR
CARDIOCUTANEOUS SYNDROMES or RASopathies		
LEOPARD syndrome	PTPN11	AD
Noonan syndrome	PTPN11, RAF1, SOS1, KRAS, NRAS, BRAF	AD
Costello syndrome	HRAS	AD
Cardiofaciocutaneous Syndrome	KRAS, BRAF, MEK1 & 2 MEK2	AD
Neurofibromatosis Type 1	NF1	AD
CONGENITAL LIPODYSTROPHIC SYNDROMES		
Type 1 (Berardinelli-Seip Syndrome)	AGPAT2	AR
Type 2 (Seip Syndrome)	BSCL2	AR
NEUROMUSCULAR DISORDERS		
Myofibrillar myopathy type 2	CRYAB	AD
Myofibrillar myopathy type 1	DES	AD, AR
Emery Dreifuss muscular dystrophy 6	FHL1	X-linked
Nemaline Myopathy 3	ACTA1	AD
AMYLOIDOSIS		
Transthyretin-related amyloidosis	TTR	AD

Table 2.2. Genetic diseases and syndromes associated with HCM

AD: autosomal dominant; AR: autosomal recessive. Adapted from (Coats and Elliott 2013)

2.1.2 Clinical phenotype of HCM

The clinical spectrum of HCM is wide and overlaps with non-genetic conditions associated with left ventricular hypertrophy (LVH) such as hypertension, obesity, and athletic heart. Although genetic markers underpin the diagnosis of HCM, they currently have a limited role in predicting the clinical phenotype or guiding individual patient management.

2.1.2.1 Disease demographics

Hypertrophic cardiomyopathy may present at any age, from early childhood to old age and many patients will have a normal life expectancy. Historically the disease was thought to be a rare and potentially fatal condition but this belief reflected referral bias of the severest cases to specialist centres (Elliott, Gimeno et al. 2006). Family screening forms an integral part of clinical assessment, and as a result, the disease spectrum now extends to asymptomatic adults and children who carry a pathogenic (or disease-causing) mutation in one of the known cardiomyopathy genes. These patients may have a sub-clinical phenotype with mild or no clinically detectable abnormalities (McKenna, Spirito et al. 1997). Most studies report a greater number of affected males which is unexplained, but might be due to genetic and hormonal modifiers. The prevalence of HCM in different racial groups is similar (Authors/Task Force, Elliott et al. 2014).

2.1.2.2 Diagnosis and Investigations

The practicalities of diagnosis and management of HCM are based on international guidelines created from a consensus of expert opinion rather than large randomised controlled trials (Gersh, Maron et al. 2011, Elliott, Anastakis et al. 2014). The mainstay of clinical evaluation of an individual patient with HCM, or a gene carrier, is a personal and family history and physical examination. Investigations include resting and ambulatory electrocardiography (ECG), echocardiography, exercise testing and cardiac magnetic resonance (CMR) imaging (Figure 2.1). A panel of blood tests is usually performed to screen for disorders that cause or exacerbate myocardial dysfunction (e.g. thyroid disease, diabetes, and anaemia) and assess secondary organ dysfunction. Additional biomarkers (e.g. creatine phosphokinase (CPK), alpha-galactosidase A (α -gal A) and lactate) can be helpful to ascertain the

presence of a genocopy (Table 2.2) and should be measured if there is a clinical suspicion of disease associated with secondary HCM (Section 2.4.3). For example elevated CPK can be found in mitochondrial diseases, glycogen storage disease and Danon disease (Rapezzi, Arbustini et al. 2013).

Imaging features of HCM

Using echocardiography, HCM is defined by a left ventricular (LV) maximal wall thickness ≥ 15 mm, or 13 to 14 mm wall thickness in the presence of a family history and other features of the disease (McKenna, Spirito et al. 1997). The distribution of LV hypertrophy is typically regional and asymmetric affecting the ventricular septum. Other patterns of LV hypertrophy have been described including apical, concentric and sigmoidal, with right ventricular hypertrophy (RVH) reported in around 20 % of patients (Binder, Ommen et al. 2006, Lopes, Syrris et al. 2015). Left ventricular outflow tract obstruction (LVOTO) affects up to two thirds of patients and is associated with an unfavourable prognosis (Maron, Olivotto et al. 2003, Elliott, Gimeno et al. 2006). The presence of LVOTO is determined using Doppler echocardiography to record a pressure gradient ≥ 30 mmHg at rest or during physiological provocation such as Valsalva manoeuvre, standing or exercise. A gradient of ≥ 50 mmHg is considered to be the threshold at which LVOTO becomes haemodynamically important (Wigle, Sasson et al. 1985). The mechanism of obstruction is usually due to anterior motion of the mitral valve causing systolic septal contact and obliteration of the outflow tract.

The introduction of routine cardiac magnetic resonance (CMR) imaging over the past decade has revealed additional morphological features including segmental hypertrophy, hypertrabeculation, myocardial clefts and apical aneurysms (Maron, Maron et al. 2009). The greatest advantage of CMR is that it enables characterisation of myocardial tissue using late gadolinium enhancement (LGE) imaging to determine the presence of fibrosis. There are important relationships between LGE and cardiovascular mortality, heart failure death, and all-cause mortality in HCM (Green, Berger et al. 2012). However, it is difficult to quantify the extent of LGE to monitor disease progression.

2.1.2.3 Treatment

The aims of treatment are to relieve symptoms and prevent complications of heart failure, sudden death, thromboembolism, and infective endocarditis. Pharmacological therapy for the relief of symptoms is pragmatic and based on the use of non-vasodilating beta-blockers or calcium channel blockers which slow the heart rate, prolong diastole, and suppress arrhythmias (Authors/Task Force, Elliott et al. 2014). Drug refractory symptoms related to LVOTO can be treated with surgery (ventricular septal myectomy), percutaneous alcohol septal ablation or dual chamber pacing. Heart failure affects 10-20 % of patients and a small proportion will require cardiac transplantation (Melacini, Basso et al. 2010, Kato, Takayama et al. 2012). Particular clinical characteristics are used to identify patients at highest risk of sudden cardiac death (SCD) and guide recommendations for implantable cardioverter-defibrillators (O'Mahony, Jichi et al. 2014).

2.1.2.4 Future Challenges

Although there has been considerable progress in understanding the genetic basis of HCM, it remains an incurable chronic disease. By conducting family screening over several decades, we have now identified a large population of patients with either early disease or a genetic predisposition to disease. The aim of much clinical research is to better understand the molecular pathways that lead from the gene mutation to a diverse clinical phenotype and thereby identify new drug targets that can prevent disease development or slow progression of the phenotype in patients at a more advanced stage of its natural history.

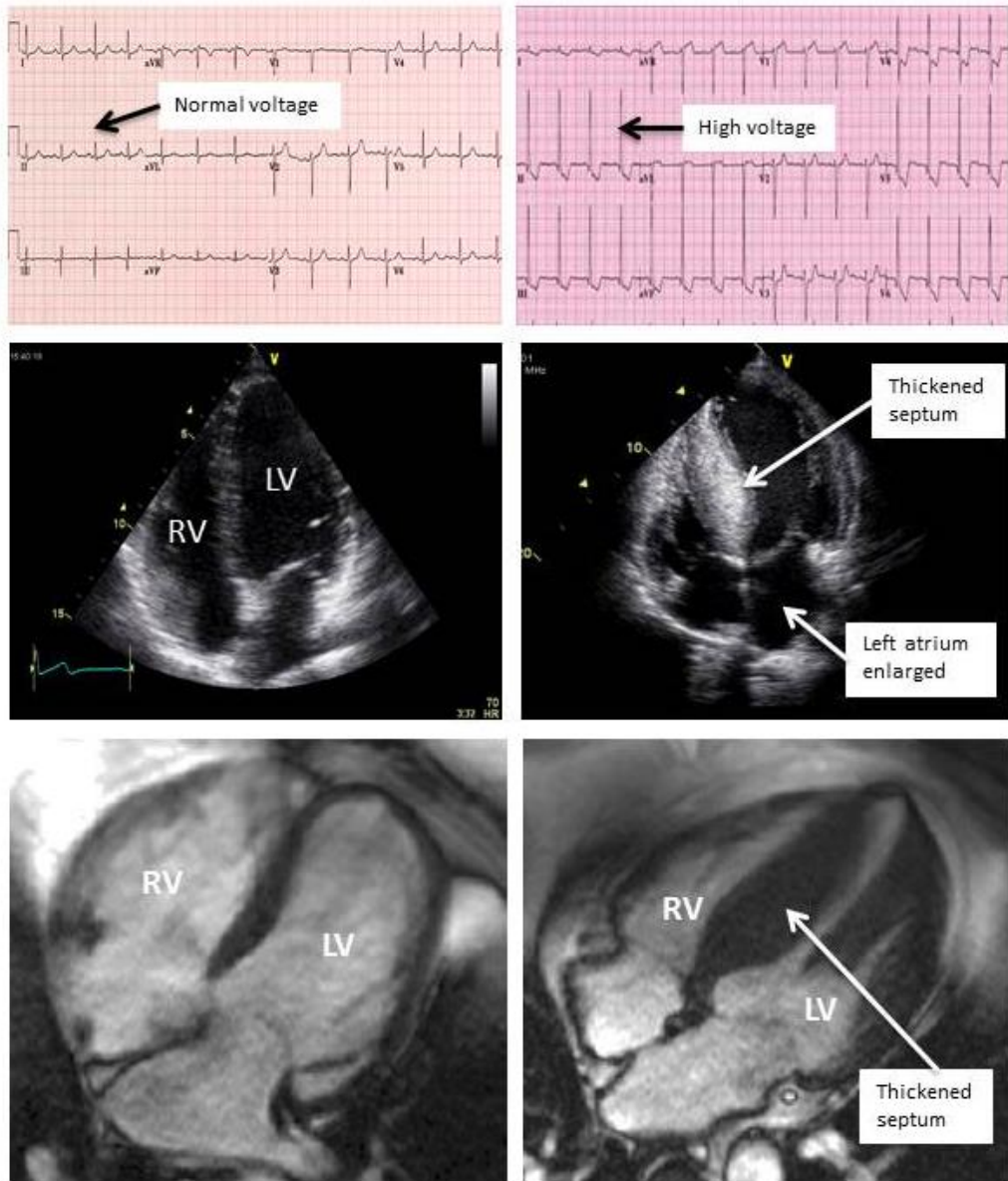


Figure 2.1. The clinical phenotype of HCM

The phenotype of HCM is described by a) an electrocardiogram b) two-dimensional echocardiography and c) CMR imaging. The left panel shows these investigations in a normal individual and the right panel a patient with HCM. Major abnormalities are highlighted with arrows. LV = left ventricle, RV = right ventricle.

2.1.3 Pathophysiology of HCM

Despite its global prevalence, the pathophysiology of HCM is still poorly understood. Once believed to be an archetypal single-gene disorder, HCM is now recognised as a complex condition with considerable biological heterogeneity.

Cardiac hypertrophy

Cardiac hypertrophy is an adaptive response to pressure or volume stress, genetic mutations, or loss of muscle mass from prior myocardial infarction (Frey, Katus et al. 2004). Pressure overload (e.g. from hypertension) leads to concentric LV chamber hypertrophy, which histologically is characterised by an increase in myocyte cross-sectional area. Volume overload (e.g. from aortic regurgitation) leads to a proportional increase in chamber diameter/wall thickness, which is characterised histologically by both increased myocyte length and cross-sectional area. Excessive myocyte lengthening without further changes in area appears to be the primary cellular feature leading to chamber dilatation and progression to heart failure (Gerdes 2002, Dorn, Robbins et al. 2003). There has been substantial effort to identify the key mechanisms leading to cardiomyocyte hypertrophy which is believed to have a compensatory function by diminishing wall stress and oxygen consumption. Potential pathways include an abnormal mechanical stress response, altered calcium cycling within the sarcomere, increased fibrosis, and impaired energy homeostasis.

Hypertrophic cardiomyopathy has been a paradigm for studying mechanisms of cardiac hypertrophy. Whilst some final common pathways may explain pathogenesis of HCM, there are certain aspects of the disease that are not shared by more common diseases associated with left ventricular hypertrophy. The macroscopic, microscopic, and molecular features of HCM are discussed below, with focus on the proposed mechanisms to explain the development of both cellular and gross cardiac hypertrophy resulting from single sarcomere gene mutations.

2.1.3.1 Histopathology of HCM

The modern description of hypertrophic cardiomyopathy is credited to the London pathologist, Robert Donald Teare who likened the disease to “*a tumour of the heart*” (Teare 1958). Although thick and heavy hearts had been of interest to artists, anatomists and pathologists for centuries, it wasn’t until 1958 that Teare made the

clinic-pathological link between particular features of the disease and premature or sudden death in families (Coats and Hollman 2008). The common features he recognised were “*bizarre disorganised muscle bundles and individual fibre hypertrophy*”. Nowadays the clinical phenotype is much broader, but these original pathological features have remained the cornerstone of post mortem diagnosis (Davies and McKenna 1994).

Macroscopic

The average adult human heart weighs about 300 grams in men, and about 200 grams in women. The average heart weight in autopsy studies of HCM is 500 to 600 grams with hearts over 1000 grams reported in some young sudden deaths (Varnava, Elliott et al. 2000). At whole organ level, myocardial hypertrophy in HCM is typically asymmetric, rather than concentric as in aortic stenosis (AS) and hypertension. This results in marked regional variation in wall stress as well as systolic and diastolic dysfunction. The hypertrophied myocardium causes a reduction in LV cavity size and narrowing of the left ventricular outflow tract (Figure 2.2). The change in geometry can cause the mitral valve leaflets to move towards the septum when the heart contracts, leading to obstruction to outflow and regurgitation of blood backwards into the left atrium. This dynamic phenomenon of left ventricular outflow tract obstruction (LVOTO) results in increased ventricular afterload and reduced myocardial perfusion and is understood to be the mechanism responsible for many patients’ symptoms (Wigle, Sasson et al. 1985).

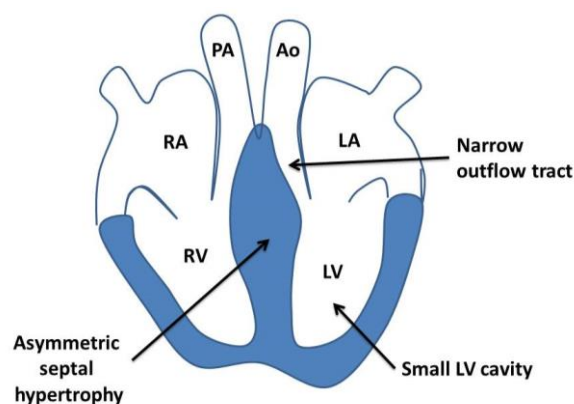


Figure 2.2. Macroscopic features of hypertrophic cardiomyopathy.

Ao = aorta, LA = left atrium, LV=left ventricle, PA = pulmonary artery, RA = right atrium, RV = right ventricle.

Microscopic

Histological features include myocyte hypertrophy and disarray, interstitial and/or replacement fibrosis, and thickening of the walls of the small mural vessels (Figure 2.3). However, the extent and distribution of these findings is extremely variable between patients and within affected individuals during their lifespan (Becker and Caruso 1982). It has been postulated that this variation in pathological phenotype may reflect the genetic heterogeneity that underlies the disease. Patients with mutations in *TNNT2* gene, for example, have been reported to exhibit severe disarray, in the context of only mild hypertrophy and fibrosis (Varnava, Elliott et al. 2001). Few studies have explored the relationship between genotype, protein expression and histology in human disease.

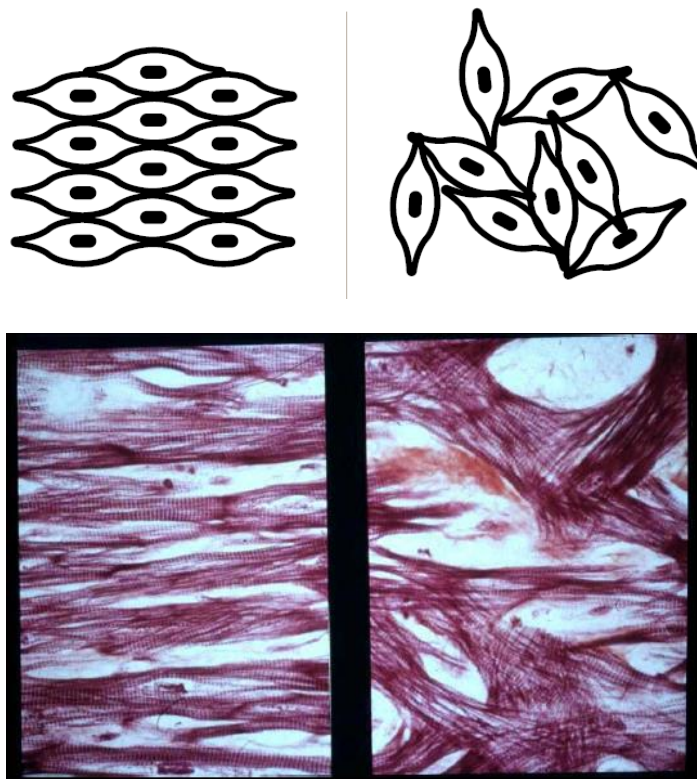


Figure 2.3. Microscopic features of hypertrophic cardiomyopathy

Left: Normal myocardium. Right: Hypertrophic cardiomyopathy. Note the irregular shape and disorganised arrangement of muscle cells in HCM (Davies, McKenna 1994).

2.1.3.2 Fibrosis

Fibrosis occurs in a variety of tissues and is a key feature of disease in many organs. In the heart, myocardial fibrosis is categorised as either interstitial fibrosis or replacement fibrosis. Both types are observed in HCM and are linked to morphological and functional disease severity.

Interstitial (or reactive) fibrosis occurs in perivascular spaces and corresponds to similar fibrogenic responses in other tissues; replacement fibrosis occurs at the site of myocyte loss (Rockey, Bell et al. 2015). Comparison is often made between HCM and hypertensive heart disease, but in hypertension the percentage area of fibrosis correlates with heart weight implying the widespread fibrosis in HCM cannot be explained by cardiac hypertrophy alone, and that disarray and other factors are also important in fibrogenesis (Tanaka, Fujiwara et al. 1986). Histopathological studies have shown an eight-fold increase in the amount of matrix collagen in HCM patients who die suddenly or develop progressive heart failure compared to normal controls (Shirani, Pick et al. 2000, Lombardi, Betocchi et al. 2003). Although fibrosis is not specific to HCM, its quantification and regional distribution provide information that is useful in exploring the cellular and molecular processes underlying this disorder.

Imaging fibrosis

The current gold standard for non-invasive measurement of fibrosis is by cardiac magnetic resonance (CMR) imaging (Paya, Marin et al. 2008). Gadolinium-based contrast agents can be used to detect expansion of the myocardial interstitium, an imaging technique termed late gadolinium enhancement (LGE). The typical distribution pattern of LGE found in HCM is patchy or confluent, generally located in wall segments with greatest thickening, and at the junction points between right and left ventricles in the septum, and largely reflects replacement fibrosis. It is evident from both histology and gadolinium enhanced CMR imaging that myocardial fibrosis is an important determinant of disease progression particularly in the development of heart failure (Figure 2.4). Interstitial fibrosis is more challenging to distinguish with LGE, but emerging T1 mapping techniques can detect and quantify diffuse fibrosis which is recognised in gene carriers prior to developing hypertrophy (Hinojar, Varma et al. 2015). Tissue Doppler echocardiography measurements and

grades of diastolic function correlate with CMR documented fibrosis and may also be useful to track disease progression in larger populations. (Moon, McKenna et al. 2003, O'Hanlon, Grasso et al. 2010).

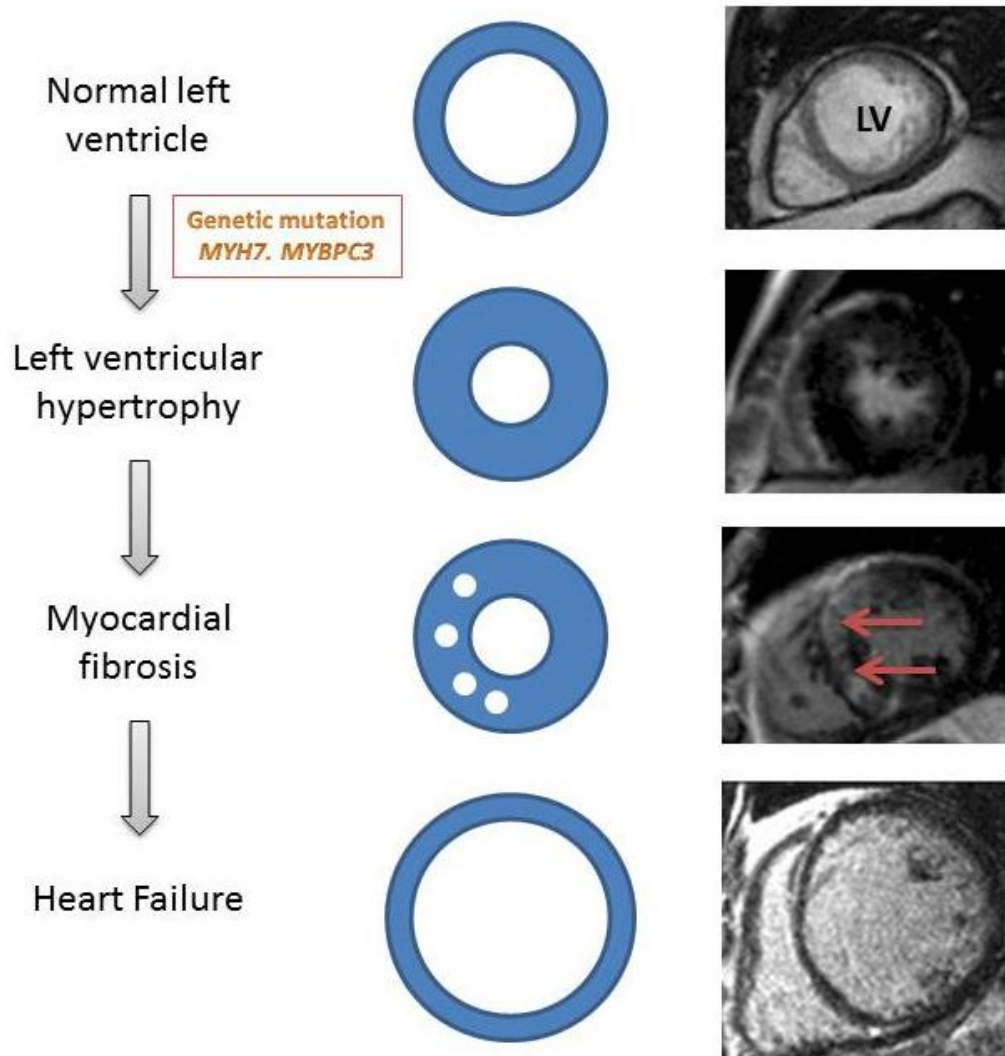


Figure 2.4 Cardiac magnetic resonance imaging in HCM

Increased amounts of fibrosis can be detected using cardiac magnetic resonance late gadolinium enhanced imaging. The areas of fibrosis are indicated with arrows. Progressive remodelling leads to chamber dilatation and heart failure. LV = left ventricle.

Biomarkers of fibrosis

Fibroblasts and myofibroblasts are key fibrosis effectors in many organs, and are responsible for the synthesis of extracellular matrix (ECM) and regulatory proteins (Hinz, Phan et al. 2007). Collagen types I and III are the main collagen fractions within the heart, while collagen types IV, V and VI are less abundant (Camelliti, Borg et al. 2005). The matrix metalloproteinases (MMPs), a family of zinc-dependent enzymes, and their inhibitors, the tissue inhibitors of metalloproteinases (TIMPs) maintain the correct balance between collagen synthesis and degradation. The C- or N-terminal pro-peptides of collagen type I (PICP, PINP) and collagen type III (PIIICP, PIIINP) that are released during collagen biosynthesis are measured as biomarkers of collagen accumulation.

There is increased activity of fibroblasts in HCM and collagen I synthesis prevails over degradation (Lombardi, Betocchi et al. 2003, Spinale 2007). Both PICP and PIIINP correlate with scar burden on CMR and are elevated in sarcomere-mutation carriers without visible hypertrophy, suggesting a potential diagnostic role for these biomarkers in detecting a diffuse interstitial profibrotic state that precedes the classical clinical phenotype (Ho, Lopez et al. 2010).

2.1.3.3 Inflammation

Acute and chronic inflammation is associated with fibrosis in various disease states. C-reactive protein (CRP) is a non-specific marker of inflammation and its prognostic role has been demonstrated in chronic heart failure and dilated cardiomyopathy, however it is not very specific levels do not always correlate with histological signs of myocardial inflammation (Zimmermann, Bienek-Ziolkowski et al. 2009). CRP levels are higher in HCM patients compared with controls, particularly in the subgroup of patients with LVOT obstruction (Dimitrow, Undas et al. 2008). Tissue necrosis factor alpha (TNF- α) and interleukin 6 (IL-6) have also been found elevated in the plasma of patients with HCM, compared to healthy controls, and are highest in patients with end-stage disease (Hogye, Mandi et al. 2004, Zen, Irie et al. 2005). However, these studies are small with no serial data or correlation with tissue, and the role of inflammation in development and progression of HCM is indeterminate.

2.1.3.4 Mechanical stress

Mechanical stress is an important trigger for inducing a hypertrophic growth response in the pressure or volume overloaded heart. Numerous pathways are involved from sensing mechanical stimuli to signal transduction enabling modulation of gene expression and protein synthesis. An early hypothesis in HCM was that the sarcomere mutations resulted in decreased contractile activity and that hypertrophy was a compensatory response (Sata and Ikebe 1996). We now appreciate that sarcomere mutations can both increase and decrease the force of contractility at a molecular level. Although not proven, there is a common view that “hypercontractile” sarcomere protein gene mutations result in HCM whereas “hypocontractile” mutations lead to dilated cardiomyopathy (DCM) (Debold, Schmitt et al. 2007, Spudich 2014). Work in this area is ongoing.

Anatomy of the sarcomere

The sarcomere is made up of several myofilament proteins organised into myofibrils in such a way that muscle can contract and relax (Figure 2.5). The main components are myosin which forms the thick filament and actin which forms the thin filament. Myosin is made up of two protein units of β - or α -myosin heavy chain and four myosin light chain molecules: two essential light chain and two regulatory light chains. Cardiac myosin binding protein-C (cardiac MyBP-C) is a myosin-associated protein that binds at intervals along the myosin thick filament backbone. The thin filament is composed of repeating actin molecules and is associated with the troponin complex (made up of three different troponins I, C and T) and alpha-tropomyosin.

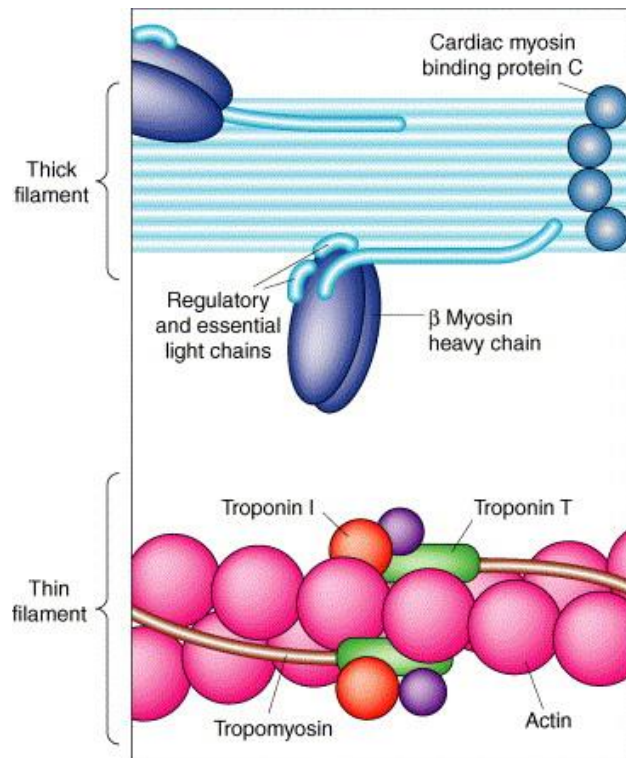


Figure 2.5 Composition of the sarcomere.

Image reproduced from (Ashrafian, Redwood et al. 2003). The cardiac sarcomere comprises the contractile proteins of the thick and thin filaments. Mutations in genes encoding each of the proteins labelled in this figure can cause HCM.

Normal muscle contraction

Normal myocardial contraction depends on oxygen for high-energy phosphate production by oxidative phosphorylation. Myosin contraction sites are exposed when calcium binds to troponin molecules causing tropomyosin to move. Cross bridges are formed when the myosin heads bind to the exposed attachment sites on the actin allowing phosphates to be released from myosin heads. Stored energy from myosin heads is used to cause movement of actin past myosin which shortens the sarcomere. Adenosine diphosphate (ADP) molecules are then released from the myosin heads. Next adenosine triphosphate (ATP) molecules bind to the myosin heads. When ATP is broken down to ADP and phosphates, cross bridges are released. Finally, in recovery the heads of the myosin molecule are returned to their resting position and energy is stored in them. This known as the sliding filament theory (Huxley and Niedergerke 1954).

Mutant sarcomere proteins

Mutant proteins can alter sarcomere function by either decreasing or increasing the translocating filament sliding force. In HCM an increased force of contraction is typically observed. Although the mechanisms are not fully elucidated, several studies illustrate the importance of characterising the individual response of HCM mutations and integrating information at genetic, transcript and protein level.

Theis et al studied expression and localization of myofilament proteins in surgical myectomy specimens from 47 patients with HCM (45% had either *MYBPC3* (n=12) or *MYH7* (n=9) mutations). Myofilament protein levels were quantified by Western blotting, and localisation graded from immunohistochemical staining of tissue sections. Increased levels of myofilament proteins in patients carrying a mutation in either gene was found suggesting a dominant negative mechanism. Patients with a frameshift mutation predicted to truncate cardiac MyBP-C showed clear disturbances in protein localization compared with missense mutations implying the mechanism of dysfunction may vary according to the genetic subgroup (Theis, Bos et al. 2009).

Poison peptide

The majority of *MYH7* mutations are missense and are proposed to cause disease by a 'poison peptide' mechanism whereby the mutant protein is expressed in the heart and impairs or overrides the function of the wild-type protein from the normal allele (Seidman and Seidman 2001). More recently, the notion of allelic imbalance (where gene expression levels differ between the two alleles) was suggested to explain the variation between genotype and phenotype. Tripathi et al demonstrated that the fraction of gene product (at both mRNA and protein levels) in human tissue is identical among relatives but the total amount varies depending on the specific mutation (Tripathi, Schultz et al. 2011). Furthermore in a HCM mouse (*MYH6*, R403Q) it has been shown that mutant transcripts can be selectively silenced by an RNA interference cassette delivered by an adeno-associated virus vector, a potential strategy for therapeutic benefit (Jiang, Wakimoto et al. 2013).

Haploinsufficiency

Most pathogenic *MYBPC3* mutations are frameshift or nonsense, leading to premature truncation of the protein. Marston et al studied ventricular muscle from 37

patients undergoing surgical myectomy with control samples from donor hearts. They quantified cardiac MyBP-C protein and mRNA levels using immunoblotting and RT-PCR and were unable to detect any truncated cardiac MyBP-C in whole muscle homogenates arguing against a dominant negative effect. They also found a reduced overall level of cardiac MyBP-C in samples containing either a truncation or missense *MYBPC3* mutation compared with donor and non-*MYBPC3* myocardial samples, implying “haploinsufficiency” was the mechanism of HCM (Marston, Copeland et al. 2009). Findings were reproduced in an independent patient group including 7 patients with a founder mutation. The authors also demonstrated reduced force generating capacity of the cardiomyocytes and evidence of increased Ca^{2+} sensitivity due to hypophosphorylation of troponin I (van Dijk, Dooijes et al. 2009). More recently, a larger study of 46 genotyped patients with HCM found an unexpected 9-fold upregulation of total *MYBPC3* mRNA in samples containing *MYBPC3* truncating mutations, and proposed this might be compensating for the mutant transcript to maintain normal protein levels. Using multiple reaction monitoring, they demonstrated the fraction of mutant protein relative to wild type ranges from 30 to 84 %, varying with individual mutations. The authors argue against a single mechanism of haploinsufficiency and propose allelic imbalance contributes to disease progression and phenotype variation (Helms, Davis et al. 2014).

The Myosin Mesa

More recently Spudich et al, recognizing that many classical HCM mutations are localised to the S1 domain on human β -cardiac myosin, proposed that cardiac MyBP-C mainly interacts with this region, and designated it “the myosin mesa”. He showed that mutations which reduce the affinity of cardiac MyBP-C for the myosin head, allowed those heads to become involved in contraction thus explaining the consequential hyper-contractile response (Spudich 2015). He theorised that small molecules binding to the mutant sarcomeric protein complex should therefore be able to mitigate the effects of mutations at their sources, and Phase 1 trials have just begun (ClinicalTrials.gov Identifier: NCT02329184).

2.1.3.5 Myocardial metabolism

Energy is required for muscle contraction. Adenosine triphosphate (ATP) is produced primarily by the metabolism of free fatty acids (FFAs) and carbohydrates, with FFAs accounting for approximately 70% of ATP production (Stanley, Recchia et al. 2005). Importantly, FFAs are less efficient as a source of myocardial energy as they require approximately 10% more oxygen than glucose in order to produce an equivalent amount of ATP (Stanley, Recchia et al. 2005). Several types of pathology including dilated and hypertrophic cardiomyopathies are associated with a reduction in cardiac energetic status. Mechanisms responsible include impaired energy generation and/or inefficient energy utilisation (Kalsi, Smolenski et al. 1999, Siddiqi, Singh et al. 2013).

The energy depletion hypothesis

The observation that sarcomere mutations have differing contractile properties led to the hypothesis that inefficient use of ATP, rather than abnormal force generation, may be the unifying pathogenic mechanism in HCM. This is supported by identifying mutations in genes (e.g. *PRKAG2*, 5'-AMP-activated protein kinase subunit gamma-2) involved in energy homeostasis in families with HCM (Blair, Redwood et al. 2001). Patients with inherited mitochondrial or metabolic disorders may also develop a HCM phenotype; although rare, these conditions explain a proportion of non-autosomal dominant disease, more commonly with childhood onset (Obayashi, Hattori et al. 1992, Colan, Lipshultz et al. 2007).

Evidence for energetic impairment in HCM

By applying magnetic resonance spectroscopy (MRS) techniques to humans with HCM a disproportionate reduction in phosphocreatine (PCr) is observed compared with ATP. The reduction in PCr/ATP ratio is also evident in sarcomere mutation carriers with little or no hypertrophy implying that energetic impairment occurs early in the development of HCM (Crilley, Boehm et al. 2003). This is supported by *in vitro* measurements in human cardiac muscle strips where ATPase activity during force development, is higher in sarcomere mutation-positive compared with sarcomere mutation-negative HCM hearts, indicating that energetic alterations are primarily related to the sarcomere mutation itself, rather than a non-specific feature

of hypertrophied myocardium (Witjas-Paalberends, Guclu et al. 2014). Clinical studies add further support since in HCM left ventricular relaxation (a highly energy dependent process) paradoxically prolongs on exercise (Lele, Thomson et al. 1995). Pharmacological modulation using perhexiline, an inhibitor of mitochondrial fatty acid uptake, leads to improved myocardial energetics and exercise capacity (Abozguia, Elliott et al. 2010).

2.1.3.6 Calcium homeostasis

Electrical activation and contraction of the heart are coupled through the intracellular movement of calcium. Increased calcium (Ca^{2+}) sensitivity has been reported in all contractile protein genes that cause HCM, and is recognised to contribute to the observed increase in force production and impaired relaxation (Somura, Izawa et al. 2001). It is also notable that the risk of sudden cardiac death appears to be higher with certain mutations (e.g. *TNNT2*, cardiac troponin T) and correlates with the degree of increased Ca^{2+} sensitivity measured in vitro (Haim, Dowell et al. 2007). Patient-specific induced pluripotent stem cell-derived cardiomyocytes carrying a R663H mutation in *MYH7* display elevated Ca^{2+} which induces both hypertrophy and arrhythmia. In the same model pharmacological treatment of Ca^{2+} imbalance prevents the HCM phenotype development (Lan, Lee et al. 2013). In mouse models of HCM (an alpha cardiac myosin heavy-chain *MYH6* R403Q missense mutant mouse and a transgenic mouse model expressing I79N *TNNT2* mutation) inhibition of plasma membrane L-type Ca^{2+} channels by diltiazem normalised calcium storage protein levels and prevented fibrosis and cardiac dysfunction (Semsarian, Ahmad et al. 2002, Westermann, Knollmann et al. 2006). Clinical trials of diltiazem (a Ca^{2+} -channel blocker) in humans are currently underway (Ho, Lakdawala et al. 2015).

2.1.3.7 Abnormal development

The heart contains different cell types including cardiomyocytes, endocardial cells (covering the inner layer of the heart), epicardial cells (covering the outer layer of the heart) and smooth muscle cells. During development, these cells must migrate and proliferate to form the coronary arteries, heart valves and myocardium. The diverse morphological abnormalities observed in HCM (e.g. myocardial bridging, abnormal mitral valves, diffuse fibrosis and microvascular remodeling) led researchers to

consider a primary developmental hypothesis for the disease (Olivotto, Cecchi et al. 2009). At the time it was understood that epicardial derived cells were pluripotent and could migrate and differentiate into diverse cells types (Lie-Venema, van den Akker et al. 2007). If we are to target treatment at such an early stage, the complex signaling pathways that determine the fate of these pluripotent cells, must be fully characterised.

2.1.3.8 Is there a unifying hypothesis?

Despite significant advances in structure-function relationships relating to the cardiac sarcomere, there is still limited knowledge of how a mutation leads to clinical HCM. Genetic diversity alone does not explain the variation in HCM phenotypes. The unifying hypotheses of perturbed energy utilization or altered calcium homeostasis are plausible but do not entirely explain features of the disease outside the cardiomyocyte such as fibrosis. The lack of a consistent contractile abnormality refutes hypertrophy as a purely compensatory phenomenon. A developmental hypothesis is attractive because it proposes several cell types contribute to the diverse pathologies of HCM. Although there is no consensus on a unifying hypothesis, critical advances in diagnosis, treatment and prevention mandate a deeper understanding of HCM pathophysiology.

2.1.4 HCM in the post genomic era

For the last half century, we have largely followed the doctrine that there is linear causality from a gene via transcript to protein and then function. Experimental approaches have tended to focus on a single gene, transcription factor, protein or pathway, but when the goal is to explain an organism as diverse as the human body, it is not surprising this approach has had limited success. In both health and disease states, multiple physiological systems work in an integrated way.

2.1.4.1 Systems Biology

The Human Genome Project has shown there are around 20,500 human genes. Since the completed sequence was published in 2004 this information has been used as a reference map to study systems biology (Consortium 2004). It was initially thought that only about 1% of DNA coded for genes that make functional proteins and the remainder was “junk”. However, according to many genome-wide association studies, the majority of single-nucleotide polymorphisms (common variations in genes) that underlie differences in our susceptibility to disease, are found outside of protein-coding regions (Maher 2012). A follow-on project, called ENCODE (Encyclopaedia of DNA Elements) aimed to better understand the role of this non-coding DNA. The consortium concluded that the majority of our DNA is in fact functional and that the fundamental unit of the genome, and basic unit of heredity, is the transcript (Birney, Stamatoyannopoulos et al. 2007). Even this discovery has failed to explain the functional complexity of organisms, which far exceeds its genome sequence alone (Figure 2.6). Clearly one gene does not encode a single protein because of processes such as alternative mRNA splicing, RNA editing, epigenetics and post-translational protein modification.

Increasingly “omic” approaches have been used to study the global products of gene expression. Emerging scientific disciplines such as transcriptomics, proteomics, and lipidomics/metabolomics, are playing a major role in elucidating the functional role of genes and their products and in understanding their involvement in disease. These -omic profiles can be compared with one another to provide insight into disease mechanisms. The transcriptome contains mainly genomic information, whereas the proteome is determined to a much greater extent by environmental influences. The

application of post genomic technology, specifically proteomics, to understand disease mechanisms in HCM is addressed in this thesis.

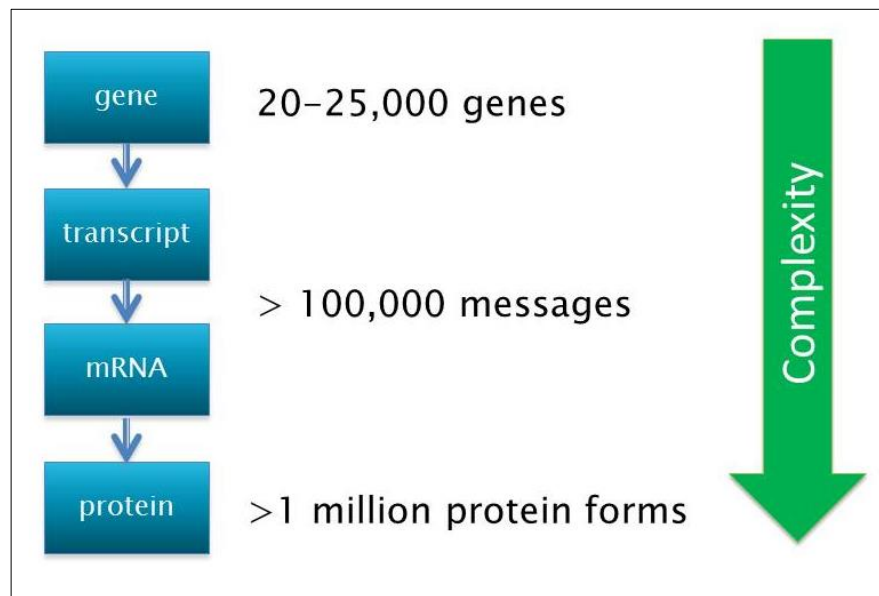


Figure 2.6 Complexity of biological information

2.1.4.2 Transcriptomics

There have been several studies performed at transcript level to try to understand the molecular pathways involved in various heart diseases. This has been made possible by the development of microarray technology (Abdellatif 2000). The standard approach is to extract RNA from normal and diseased hearts and hybridise it against reference complementary DNA (cDNA) libraries (Hwang, Dempsey et al. 1997). This is usually performed using a commercially produced gene chip which consists of a collection of microscopic DNA spots attached to a solid surface, e.g. Affymetrix whole genome arrays (<http://www.affymetrix.com>). Differentially expressed genes are analysed via various software platforms. Validation is usually by quantitative real-time reverse transcription-PCR (RT-qPCR) and the accuracy rate depends on the fold change and statistical significance level. In some studies important gene products are further validated by demonstration of differential expression at protein level using techniques such as Western blotting and immunohistochemistry.

Gene expression in HCM

More studies have been performed in dilated rather than hypertrophic cardiomyopathy because of availability of tissue which often comes from the explanted heart of a transplant recipient. Comparison has been made in both animals and humans; the studies relevant to HCM are summarised in Table 2.3.

Hwang et al showed distinct transcriptional profiles of heart failure in HCM and DCM even with the common end-point of transplantation. Among the upregulated genes, those related to the immune response were typically identified in DCM, whilst genes related to protein synthesis were notable in HCM. Genes related to metabolism were down regulated in DCM whereas those related to cell signalling and structure were down regulated in HCM (Hwang, Allen et al. 2002). Rajan et al studied two transgenic mouse models (α -TM180 and α -TM175) of HCM. Despite the missense mutations being only five amino acids apart, they showed variation in the differentially expressed transcripts consistent with the observed mild and severe disease phenotypes (Rajan, Williams et al. 2006).

Taken together, these studies form the basis for linking the primary defect (i.e. the gene mutation) to the clinical phenotype. While there is considerable overlap in the molecular changes of hypertrophy there are also distinct cardiomyopathy-specific patterns of transcriptional adaptation and progression to heart failure. Gene expression profiling can be used to provide insight into individual disease characteristics and is relevant to interpreting proteomic studies.

The most important limitations of all these studies are the experimental design and bioinformatics analysis. False positive findings (i.e. discrepancy between array and quantitative PCR data) are not uncommon and are frequently discussed by the authors. Errors occur due to poor sample selection, small sample size (including a lack of technical and biological replicates) and emphasis on statistical significance rather than effect size (Schneider and Schwartz 2000). Gene expression studies provide considerable insight into disease pathophysiology but should be accompanied by assessment of protein abundance and post-translational modifications (PTMs) to fully appreciate their relevance.

Author, Year	Objectives and Methods	Main findings
(Yang, Moravec et al. 2000)	Human failing heart explants DCM (n=1), IHD (n=1) and normal controls (n=2) Gene and protein expression validated in normal (n=6) and failing (n=7) hearts.	14 differentially expressed genes in normal and failing hearts with 5 only in failing and 2 only in non-failing hearts. ↓SLIM1 and ↑Gelsolin gene expression and protein in failing hearts.
(Lim, Roberts et al. 2001)	Human LV septal myectomy tissue (n=1) and matched normal donor heart (n=1) Validation by Northern blot in HCM (n=6) and control (n=2) samples.	Diverse genes upregulated in HCM including markers of secondary hypertrophy (e.g. skeletal alpha-actin and isoforms of myosin light chain) suggesting common pathways in HCM and acquired hypertrophy
(Hwang, Allen et al. 2002)	Human failing heart explants HCM (n=2), DCM (n=3) and normal controls (n=3). Validation by qPCR on pooled samples.	192 genes ↑ and 51 ↓ in HCM/DCM versus controls. Several genes (e.g. αB-crystallin, calsequestrin and lumican) had differential expression implying different end-stage HF phenotypes.
(Tan, Moravec et al. 2002)	Human failing heart explants DCM (n=8), normal (n=7) Validation by RT-PCR on pooled samples.	103 genes differentially expressed. Distinct gene clusters identified for different aetiologies (alcoholic and familial) of DCM
(Aronow, Toyokawa et al. 2001)	4 transgenic mouse models of hypertrophy (protein kinase C-ε activation peptide, Gαq, calcineurin and calsequestrin) Validation by Northern blot.	↑ atrial natriuretic peptide in all models, otherwise hypertrophy-regulated genes were specific to individual genetic models e.g. apoptotic gene cluster ↑ in Gαq mouse.
(Kong, Bodyak et al. 2005)	Dahl salt-sensitive rats exposed to physiological (exercise) and pathological (salt diet) hypertrophy that progress to heart failure	↑ expression of genes involved in apoptotic pathway in pathological hypertrophy progressing to HF. Insulin-IGF-PI3-kinase pathway upregulated in physiological hypertrophy
(Rajan, Williams et al. 2006)	α-tropomyosin transgenic mice representing mild (α-TM180) and severe (α-TM175) HCM. Validation by RT-PCR on pooled samples.	266 differentially expressed genes from normal. Significantly ↑ expression of secreted/extracellular matrix genes and ↓ expression of metabolic enzymes.

Table 2.3 Summary of transcriptomic studies relevant to HCM.

α-TM alpha tropomyosin, *RT-PCR* Reverse transcription polymerase chain reaction, *HF* heart failure, *IGF* insulin like growth factor, *PI3K* phosphatidylinositol-3-kinase, *qPCR* quantitative real-time polymerase chain reaction, *DCM* dilated cardiomyopathy, *HCM* hypertrophic cardiomyopathy, *LV* left ventricular, *IHD* ischaemic heart disease.

2.1.4.3 microRNA

Micro RNAs (miRNAs) are small non-coding RNA molecules that regulate gene expression by post-transcriptional regulation of protein levels. Several miRNAs have been identified as important regulators of normal cardiac development and their altered expression has been associated with hypertrophy (e.g. miR-195) and heart failure (e.g. miR-133) (van Rooij, Sutherland et al. 2006, Liu and Olson 2010, Gladka, da Costa Martins et al. 2012). Their role has recently been studied as a possible epigenetic mechanism to explain disease heterogeneity in HCM. Palacin et al compared myocardial tissue from one control to two patients with HCM due to mutations in *MYH7* and three patients with hypertrophy due to valve disease and found a different miRNA profile implying that the mechanism of hypertrophy is different (Palacin, Reguero et al. 2011). Kuster et al found miR-204, was up regulated in myectomy tissue from patients with HCM due to a *MYBPC3* mutation compared with donor myocardium (Kuster, Mulders et al. 2013). Song et al identified that miR-451 was down-regulated in HCM and showed this directly affected levels of tuberous sclerosis complex 1 (TSC1) a known positive regulator of autophagy (Song, Su et al. 2014).

This preliminary work suggests miRNAs could play a role in the pathogenesis of HCM and may even present therapeutic targets. It further emphasizes that genotype alone is insufficient to explain clinical disease. The causes of most disorders are multifactorial and systems-level approaches provide a more comprehensive picture.

2.2 Proteomics

The scientific discipline of proteomics has grown exponentially since the completion of the Human Genome Project. With modern technology, it has become possible to identify and quantify hundreds of proteins from complex samples, rather than analyse one protein at a time. By systematic separation, classification, and study of all of the proteins produced in an organism, the aim is to understand how proteins change structure, interact with other proteins, and ultimately cause disease. Major challenges are that proteomics generates more data, by several orders of magnitude, than genomics and that the biotechnology is still evolving.

2.2.1 Definition

Proteomics is the study of the proteome. The term proteome was first used by Marc Wilkins in 1994, and is defined as the full set of proteins encoded by the genome (Wasinger, Cordwell et al. 1995).

2.2.2 Why is proteomics important?

Despite the vast amount of DNA data available, simply having a complete sequence of the genome is not sufficient to explain biological complexity. The genetic code allows the prediction of the amino acid sequence of a particular protein but it provides little information about the level of protein expression, post translational modifications and correct conformation of the polypeptide sequence. Furthermore, some genes are only expressed during development and never again. The rationale for using proteomics rather than transcriptomics for understanding disease pathophysiology is to provide a more direct measure of gene expression accounting for the contribution of both transcription and translation.

2.2.2.1 Gene and protein expression is not equal

The concept of “one gene, one protein” is an over-simplification. The expression level and function of a protein is modulated at several steps from transcription to post-translation. Increased gene expression, resulting in abundant messenger RNA (mRNA), does not necessarily mean that the corresponding protein is abundant or indeed active in the cell. There is a poor correlation between the amount of mRNA

transcribed from DNA and the amount of protein translated from that mRNA (Gygi, Rochon et al. 1999).

2.2.2.2 Post translational modifications

Transcripts can be spliced in many ways and multiple protein isoforms can be generated from a single gene. Most proteins undergo post-translational modifications (PTM) which are chemical modifications of a protein after its translation. There are several hundred different PTMs e.g. glycosylation, acetylation, phosphorylation and methylation (Khoury, Baliban et al. 2011). The modification of a protein ultimately affects its function, trafficking and half-life. A variety of techniques can be used to detect PTMs experimentally including mass spectrometry and Western blotting.

2.2.2.3 Proteins are functional molecules

Proteins are functional molecules within the cell and are therefore closer to the disease process than genes. When drugs have beneficial effects, they usually do so by interacting with proteins, often enzymes. A powerful use of proteomics is to understand the composition of different intracellular compartments. By identifying sub cellular location, the function of novel proteins and genes can be better understood (Andersen, Wilkinson et al. 2003).

2.2.2.4 Proteins do not work in isolation

Protein interactions are fundamental to understanding biological function which usually results from the interaction of several proteins within a protein complex. Protein-protein interactions are also exploited for therapeutic purposes. For example, Tirofiban, used in the treatment of acute coronary syndromes, is a reversible antagonist of fibrinogen binding to the glycoprotein (GP) IIb/IIIa receptor, the major platelet surface receptor involved in platelet aggregation. Protein complexes are highly dynamic structures and the same protein may have different functions depending on where, when and what it binds to.

2.2.2.5 The genome is static, the proteome is dynamic

Unlike the genome which is considered “static”, a cell’s transcriptome and proteome are “dynamic”, constantly changing over time. The proteome consists of all proteins (including modified proteins) present within the cell at any given time. There are

therefore many proteomes to one genome reflecting the conditions to which a cell is exposed.

2.2.3 Applications of proteomics

Proteomic analysis should be considered complementary to other post genomic technologies such as micro-arrays used in transcriptomic studies. It has wide ranging uses from understanding protein interactions to identifying new biomarkers (Figure 2.7). This thesis focuses on “expression proteomics” where the major purpose is in understanding disease mechanisms and identification of potential disease biomarkers.

2.2.3.1 Understanding disease mechanisms

Since the very beginning, protein profiling has been a fundamental part of proteomics and has been applied to make comparisons between health and disease states. Traditional two-dimensional electrophoresis (2-DE) methods have improved and new non-gel based techniques continue to be developed. Mass spectrometry (MS) plays a central role in proteomics enabling highly accurate identification and quantification proteins as well as assessment of protein-protein interactions and post-translational modifications.

2.2.3.2 Biomarker discovery

The discovery of differentially expressed proteins in health and disease is frequently exploited for biomarker discovery. It is becoming apparent that a single biomarker approach is probably not sufficient to monitor complex diseases and panels of markers are needed. The proteomic approach to discovering biomarkers is a two stage process: firstly, a comparative whole proteome analysis is performed on a small set of samples, followed by a targeted analysis on a small subset of proteins or peptides in a larger cohort of samples.

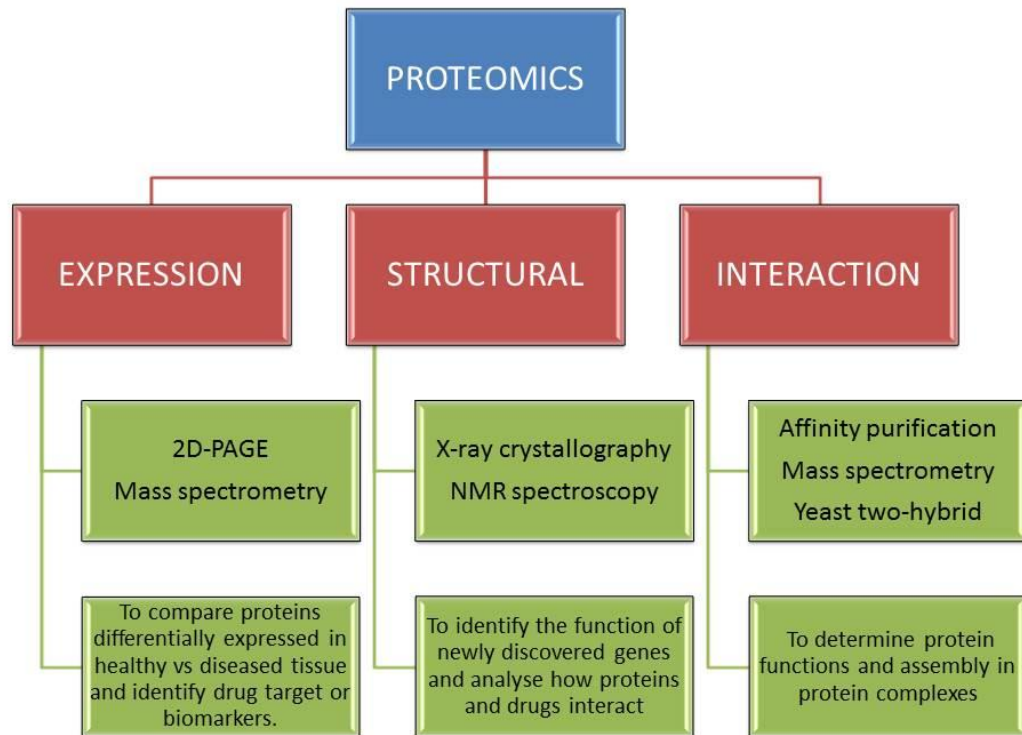


Figure 2.7 Proteomics methods and applications

Each branch of proteomics has specific objectives and uses different techniques. The main methods used in expression proteomics are mass spectrometry and 2D-PAGE (two-dimensional polyacrylamide gel electrophoresis). Structural proteomics uses X-ray crystallography and NMR (nuclear magnetic resonance) spectroscopy. Interaction proteomics uses a various affinity purification techniques and yeast two hybrid experiments.

2.2.3.3 Examples of clinical applications

Proteomics has been applied successfully in oncological medicine where the presence of certain receptors indicate that specific chemotherapeutic agents are more useful e.g. HER2 (human epidermal growth factor) testing is performed in breast cancer patients to assess prognosis and determine suitability for Trastuzumab therapy (Ross, Fletcher et al. 2004). Another example is the use of laser microdissection and mass spectrometry for sub-typing amyloidosis. The identification of important amyloidogenic proteins (i.e. serum amyloid A, transthyretin, and immunoglobulin kappa and lambda light chains) by proteomics is rapidly replacing the need for multiple different antibody-based immunohistochemical tests (Vrana, Gamez et al. 2009). This approach allows targeted treatment for particular disease sub-types on an individual patient basis.

2.2.4 Experimental approaches in proteomics

Over the last decade advances in mass spectrometry technology have contributed to the growing field of expression proteomics. Evolution from pure qualitative (i.e. protein identification) studies to large scale quantitative studies, has been driven by the growing need to analyse large sets of proteins precisely and consistently in biological samples. Experimental approaches in proteomics can be broadly classified as “discovery” or “targeted” proteomic studies.

2.2.4.1 Discovery proteomics

Discovery proteomic studies aim to detect multiple protein changes across different sample groups and can be considered “hypothesis free”. Findings are subject to bioinformatics analyses so that differential expression of groups of proteins can be understood at a pathway or network level. Relative quantitation is possible by describing the fold change in expression and the relative difference in the same protein between different samples. No information about abundance of different proteins relative to one another is possible.

2.2.4.2 Targeted proteomics

Absolute quantitation requires an entirely different approach. A targeted analysis for fewer proteins must be designed to accurately quantify only those of specific interest. Targeted assays are evolving from semi-quantitative antibody-based techniques e.g. Western blotting towards highly accurate sequence-based mass spectrometry techniques e.g. multiple reaction monitoring (MRM). This paradigm shift has been an important driver of biomarker research.

2.2.4.3 Proteomic methodology and technology

Most biological specimens (e.g. cell culture, or tissue isolates) contain thousands of different proteins, which vary in size and concentration by several orders of magnitude. The key steps of any proteomic study include a) separation, b) detection and c) quantitation of proteins. There are numerous ways in which these stages can be achieved. Separation reduces sample complexity. When performed from a mixture of proteins it is known as a “top down” approach, whereas separation after that mixture has been digested into peptides is known as a “bottom up approach”.

Whichever method is used there is a trade-off between how many proteins can be identified and how accurately they can be quantified. Mass spectrometry is integral to all proteomic studies and various instruments and techniques exist to enable detailed quantitative studies to be performed (Figure 2.8)

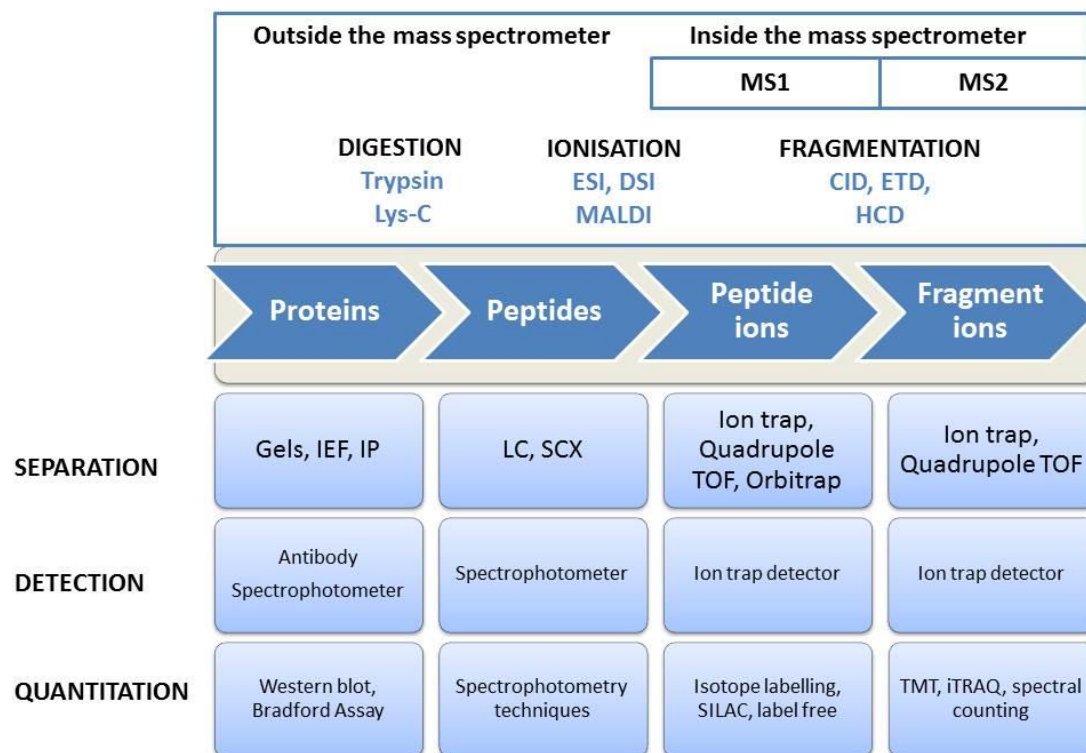


Figure 2.8. Overview of quantitative proteomics methods

The separation, detection and quantitation of proteins may occur outside and/or inside the mass spectrometer. CID Collision-induced dissociation, DSI Desorption electrospray ionization, ESI electrospray ionisation, ETD Electron-transfer dissociation, HCD Higher-energy collisional dissociation, IEF isoelectric focusing, iTRAQ Isobaric tags for relative and absolute quantitation, LC liquid chromatography, MALDI Matrix-assisted laser desorption/ionization, SCX strong cation exchange, SILAC stable isotope labelling by/with amino acids in cell culture, TMT Tandem mass tag, TOF Time of Flight.

2.2.5 Separation of protein mixtures

Historically, proteomic investigations used 2-D gel electrophoresis (2-DE) to separate the proteins in a sample and combined this with mass spectrometry (MS). Over the last decade non-gel based separation techniques have gained popularity.

2.2.5.1 Gel based protein separation

Gel based methods are widely used to separate macromolecules based on their size. By applying an electric field, molecules can be made to move through a gel made of agar or polyacrylamide by a process called electrophoresis. The type of gel used depends on the type and size of the analytes. Agarose gels are best suited to separating small DNA fragments whereas polyacrylamide gels are usually used for proteins ranging in size from 5 to 2,000 kDa. Polyacrylamide gel electrophoresis (PAGE) is the name given to this technique. Resolving gels are typically 6 to 15% depending on the concentrations of acrylamide and bis-acrylamide used in gel mixture. Buffers in gel electrophoresis are used to provide ions that carry a current and to maintain the pH at a relatively constant value.

1D PAGE

The simplest method of separating a protein mixture is one dimensional (1D) PAGE, which applies an electric field across a gel and effectively sieves molecules contained within a mixture. Mobility across the gel is a function of the length, conformation and charge of the molecule.

2D PAGE

Two-dimensional electrophoresis (2-DE), or 2D-PAGE, separates mixtures of proteins according to two properties. The 1st dimension involves separation according to charge or isoelectric point (by isoelectric focusing, IEF) and the 2nd dimension separates according to molecular weight (MW). Despite the development of new gel-free technologies, 2-DE which was first introduced 40 years ago (O'Farrell 1975) remains a powerful technique for quantitative expression profiling of large sets of complex protein mixtures such as whole cell lysates. Developments involving the use of immobilized pH gradients (IPGs) for the 1st dimension IEF separation have resulted in better resolving power and higher reproducibility. 2-DE can now resolve up to 5000 proteins simultaneously, or 2000 routinely, and can

detect <1 ng protein per spot (McGregor and Dunn 2006). It provides a map of intact proteins which not only reflects changes in expression level, but also detects isoforms and post-translational modifications (PTMs). However, 2-DE is technically difficult, labour intensive and has several potential artefacts. Hydrophobic and basic proteins are poorly solubilised in the buffers used for 2-DE, and very high and low molecular weight proteins are poorly resolved. Basic proteins have the tendency to form streaks rather than spots due to electro-endosmotic effects, the migration of reducing agents, and the hydrolysis of acrylamide at basic pH values (McGregor and Dunn 2003). Another problem with 2-DE for global proteomic profiling is its dynamic range is lower ($\sim 10^3$ proteins) than the dynamic range of proteins in cells and tissues ($\sim 10^6$) and plasma ($\sim 10^{12}$) (Corthals, Wasinger et al. 2000). Some of these challenges have been met by a “zoom-in” approach using intermediate (eg, pH 4 to 7 or 6 to 9) and narrow (e.g. pH 4 to 5) range IPG IEF gels to increase resolution in particular regions but at the cost of increased workload if large numbers of samples are investigated (Westbrook, Yan et al. 2001).

2D-DIGE

A more accurate method to quantitate protein expression is 2D-DIGE (2-dimensional fluorescence difference gel electrophoresis). First described by Jon Minden in 1997 (Unlu, Morgan et al. 1997) 2D-DIGE follows the same principles as 2D-PAGE but prior to electrophoresis, proteins are pre-labeled with one of three fluorescent amine-reactive cyanine dyes which attach to lysine, cysteine, phospho- or glyco-groups. The gel is scanned three times at red, blue and green wavelengths (Figure 2.9). The advantages of 2D-DIGE is that up to four samples (one unlabeled) can be run on one gel, which allows increased throughput and eliminates gel to gel variations, making the technique more sensitive and more accurate for quantitation. It also facilitates the use of 2-D analysis software for automated and accurate spot quantitation, gel matching, and statistical analysis (Tonge, Shaw et al. 2001).

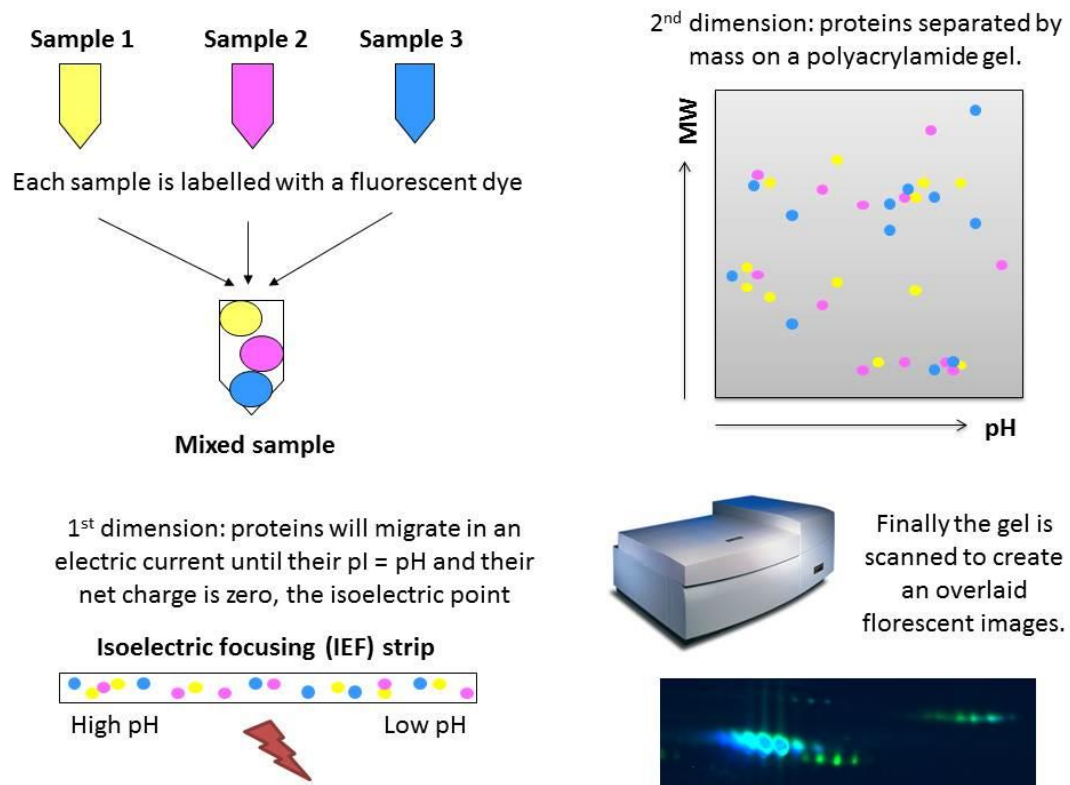


Figure 2.9. 2D-DIGE Method

Samples are labelled with fluorescent dyes prior to electrophoresis. The resulting 2D gel is then scanned at different wavelengths to acquire the separate images.

Identification of protein spots from a gel

After separation on a gel, protein spots need to be identified. This has conventionally been done by silver staining which is highly sensitive and can detect 0.1 ng of protein but its dynamic range is non-linear and PTMs can interfere with staining. Coomassie blue is a commonly used alternative stain which is low cost, detects 8–10 ng of protein and has a linear dynamic range. After visualisation the protein spots are excised for identification by mass spectrometry (Merril, Goldman et al. 1984).

2.2.5.2 Non-gel based protein separation

The limitations inherent to 2-DE (i.e. protein solubility and limited dynamic range) have led to the development of alternative “bottom up” approaches to protein separation. In these so-called shotgun approaches, a protein mixture is initially digested with an enzyme (trypsin or Lys-C). The resulting peptide fragments are then separated into fractions using strong cation exchange (SCX) or liquid

chromatography (LC) before the peptide fractions are introduced into a mass spectrometer for sequence-based identification. Developments, particularly in ionisation techniques, have allowed direct coupling of MS to chromatography systems which enables real-time separation of peptides (Section. 2.2.6.3).

The major challenge is in quantitation which has been addressed by differential isotopic labelling of the two protein mixtures (Figure 2.8). The isotope-coded affinity tag (ICAT) method is well established and based on the labelling of cysteine residues limiting its application to proteins containing these residues (Gygi, Rist et al. 1999). Alternative methods include isobaric tags for relative and absolute quantitation (iTRAQ) and tandem mass tags (TMT) where the N terminus of peptides is chemically tagged (Thompson, Schafer et al. 2003, Ross, Huang et al. 2004).

2.2.6 Mass spectrometry

Mass spectrometry (MS) is an analytical technique that separates ions according to their mass-to-charge (m/z) ratio and measures their abundance. It was first described by JJ Thomson, an English Physicist working in Cambridge who discovered the electron and won the 1906 Nobel Prize for Physics (Thomson 1921). In its early days, MS was used to demonstrate and identify isotopes (i.e. atoms of a single element with different masses due to additional neutrons). Today, mass spectrometry is the most universal and efficient tool for the primary sequence analysis of peptides.

2.2.6.1 General principles

In its simplest form a mass spectrometer can be considered a weighing machine, sensitive enough to weigh a single hydrogen atom. Peptides or proteins are weighed individually in a series of experiments and combined to form a spectrum. A mass spectrum is a Cartesian plot with mass over charge (m/z) on the x-axis and ion intensity on the y-axis. Ion intensity can be in absolute or relative measurements.

Digestion

Most proteins need to be broken down into amino acid chains (i.e. peptides) so they fall within the appropriate mass range for MS analysis. Proteases cleave the protein into peptides at specific regions. The commonest protease used is trypsin, which hydrolyses peptide bonds at the carboxylic groups of arginine and lysine. It creates positive charges at the cleavage site which is beneficial for MS analysis. An alternative protease is Lys-C which cleaves proteins on the C-terminal side of lysine residues.

Ionisation

Biological samples need to be ionized so they can be analysed by MS. Ion sources convert neutral peptides to ions by adding or taking away one or more protons. These ions may be singly or multiply charged and their net charge can be either positive or negative. Ions are separated in the mass spectrometer according to their m/z ratio by accelerating them and subjecting them to an electric or magnetic field.

Fragmentation

The identification, structure and function of a protein, are governed by its amino acid sequence. To determine this sequence, the peptide undergoes fragmentation (into product ions) creating a characteristic pattern which is used to determine structural information about the molecule. Fragmentation occurs to some extent randomly, and some fragmentations are preferred over others, producing a variation in abundance of observed peaks.

2.2.6.2 General mass spectrometer design

All types of mass spectrometer consist of three components: a) an ion source for the creation of charged ions, b) a mass analyser, for differentiation of these ions in a vacuum, and c) a photo /electron multiplier for detection of ions (Figure 2.10).

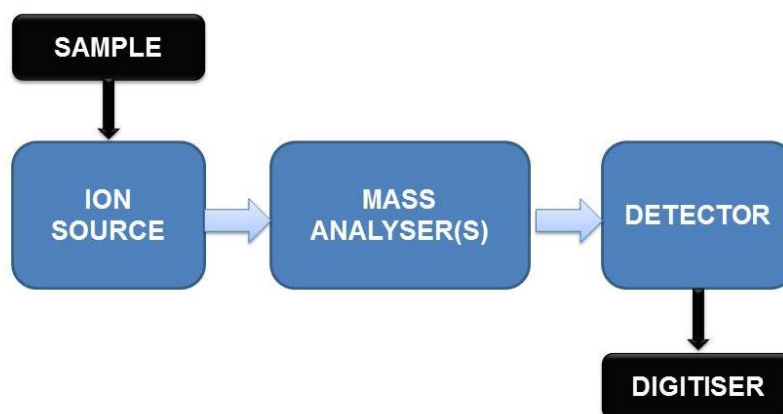


Figure 2.10 Schematic view of a generalised mass spectrometer

Ion source

The ion source makes sure that part of the sample molecules are ionised and brought into the gas phase. Ions are then transported by electric or magnetic fields to the mass analyser. There is a wide variety of ionisation techniques, but the most commonly used in proteomic research include electrospray ionisation (ESI), desorption electrospray ionization (DESI) and matrix assisted laser desorption/ionization (MALDI). The choice depends on the physico-chemical properties of the sample under investigation.

Mass analysers

Mass analysers measure the mass-to-charge (m/z) ratio of ions. They operate using electromagnetic fields under high vacuum which keeps the ions from scattering into the gas molecules. Their key specifications are resolution, mass accuracy, sensitivity and mass range.

Quadrupole (Q), ion trap and Fourier transform (FT) analysers all use frequencies to separate ions. Quadrupoles have lower mass accuracy and resolution than ion trap and FT analysers. All have limited mass range. In contrast Time of flight (ToF) analysers use time and distance to separate ions. These have high resolution and accurate mass capabilities and unlimited mass range. Hybrid analysers combine different types of analysers to achieve specific application requirements e.g. Quadrupole time of flight (Q-ToF) and triple quadrupole (QQQ). A tandem mass spectrometry (MS/MS) is capable of multiple rounds of mass spectrometry and consists of two or more mass analysers separated by some form of molecular fragmentation (Aebersold and Mann 2003). Most proteomic mass spectrometers are now orbitraps (ion trap mass analyzers) which operate by radially trapping ions about a central spindle electrode (Makarov 2000).

Detector

The detector is responsible for recording the relative abundance of each ion type. It is a scanning instrument which records either the charge or current induced when an ion hits its surface.

The specific MS techniques used in this thesis are described in more detail below.

2.2.6.3 High Performance Liquid Chromatography (HPLC)

The mass spectrometer cannot analyse all peptides at once so they must be introduced gradually. Peptides are typically separated by liquid chromatography prior to MS analysis. High-performance liquid chromatography (HPLC) is an analytical technique used to separate components of a mixture through the mass-transfer of analytes between stationary and mobile phases. The liquid mobile phase carrying analytes of interest is pumped at high pressure (500-2000 psi) through the stationary phase, in the form of an HPLC column packed with particles $> 2\mu\text{m}$. The

HPLC column is composed usually of long hydrophobic alkyl chains (C18) which retain hydrophobic compounds over hydrophilic ones. A solvent gradient of increasing organic content (acetonitrile or methanol) is used to elute the peptides in order of their hydrophobicity and is known as reverse phase HPLC (RP-HPLC). Better separation of peptides before they enter the mass spectrometer, means more peptides can be sequenced and sequence coverage of the protein is greater. HPLC enables the simultaneous desalting and separation of analytes including the separation of isomers, which reduces the effects on ion suppression by salts and other co-eluting molecules (Seger, Sturm et al. 2013).

Ultra-High Performance Liquid Chromatography (UPLC)

UPLC™ is a trademark of the Waters Corporation. It is based on the same principles as HPLC but uses columns packed with smaller particles (< 2 µm) which are able to withstand ultra-high pressures (5000 to 14,000 psi). This enables UPLC™ to separate compounds at increased speeds with superior resolution and sensitivity (Zhao, Cheng et al. 2014). As peptides leave the column they enter the static ion source of the mass spectrometer where they are ionised.

2.2.6.4 Electrospray ionisation

The Nobel Prize in Chemistry was awarded to John Bennett Fenn in 2002 for the development of electrospray ionisation (ESI) (Fenn, Mann et al. 1989). This is a soft ionisation technique which imparts little residual energy onto the sample so causes minimal fragmentation.

The ESI process begins after the peptides are eluted from a liquid chromatography column. They enter a static ion source through a narrow bore metal or glass capillary tip. ESI sources typically heat the needle to 40 to 100 °C to facilitate nebulisation, evaporation and the production of multiple charged peptides (e.g. 2⁺, 3⁺, 4⁺). A high electric potential (kV) is applied to the capillary tip, relative to a counter electrode, which causes the liquid surface to become highly charged. This generates a mist of finely charged liquid droplets, which are sprayed from the capillary orifice at atmospheric pressure (Figure 2.11). Once the droplets are airborne, the solvent evaporates, with the aid of a hot bath of nitrogen gas and the charged droplets become smaller in size with an increasing electrical charge density at the surface.

Eventually the surface charge reaches a critical limit called the Rayleigh limit and causes the ions to become so unstable they explode to form even smaller droplets in what is known as the Coulomb explosion. Finally the electrical field becomes large enough to cause desorption of the peptide ions which can then be transported to the mass spectrometer in through a series of focussing lenses by electrical fields and are held in a vacuum (Chapman 1996).

Multiple protonation often occurs with larger peptides, resulting in multiply charged proteins. The mass spectrometer can determine different charge states by analysing the isotope cluster of peaks of the peptide signal. The ion that is sent for sequencing is always the ion that has the highest charge for that particular protein (Steen and Mann 2004).

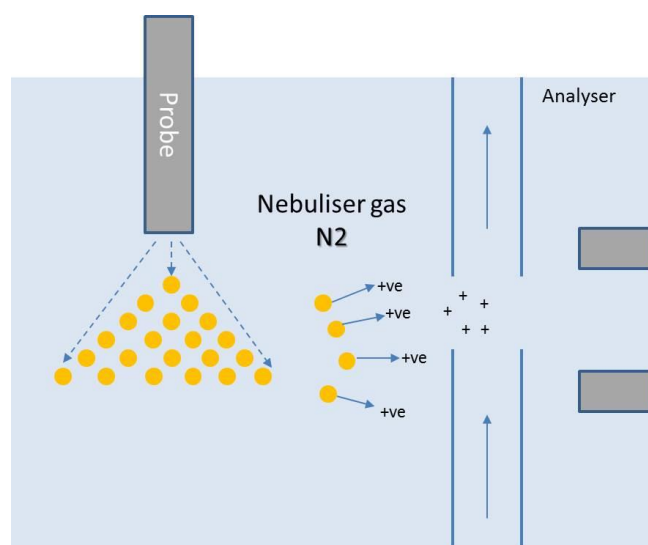


Figure 2.11. Schematic representation of an electrospray ionisation (ESI) source.

Nanospray ESI sources (consisting of a 5-10 μm diameter needle) achieve an even higher sensitivity, probably due to the higher surface to volume ratio. ESI can be coupled with UPLC for in line molecular fractionation prior to mass spectrometric analysis. This enables analysis of both small and large molecules of various polarities in a complex biological sample (Banerjee and Mazumdar 2012).

2.2.6.5 Quadrupole time of flight mass spectrometry

Quadrupole-time of flight mass spectrometry (Q-ToF-MS) is a hybrid “sequencing” technique which combines two mass analyzers within the same instrument: a quadrupole/collision cell and a time of flight analyzer. Q-ToF-MS is widely used for the detection and identification of peptides and metabolites (Figure 2.12).

Quadrupole analyser

Quadrupole mass analysers use oscillating electrical fields to selectively stabilise or destabilise the paths of ions passing through them. Following ESI, ions enter the quadrupole mass analyser (MS1) which filters ions based on their m/z ratio. The instrument consists of four parallel circular metal rods. Each opposing rod pair is connected electrically, and a radio frequency (RF) voltage is applied between one pair of rods and the other. A direct current voltage is then superimposed on the RF voltage. Ions travel down the quadrupole between the rods. The field causes the ions to oscillate in x and y directions and it is alterations in both these frequencies that allow ions with selected m/z to pass through the mass analyser and either into a second mass analyser or a detector in a stable trajectory field. Other ions that have unstable trajectories under that specific voltage collide with the rods and are ejected from the quadrupole.

Collision cell

Ions of a selected m/z (called precursor ions) enter the collision cell, where they are fragmented by collision induced dissociation (CID) using an inert gas such as argon. The fragments (called product ions) are produced mainly by the cleavage of amide bonds between amino acids and can be categorised into b and y ions: b ions are formed when the charge is retained by the N-terminal fragment, whereas y ions are formed when the charge is retained by the C-terminal fragment (Steen and Mann 2004).

Time of flight analyser

A rapidly pulsing electrical field introduces the product ions into the orthogonally situated flight tube of the ToF mass analyser (MS2). These ions are accelerated from the same position, at the same time, with identical kinetic energy. As a result lighter ions fly faster and arrive at the detector earlier than heavier ones. The velocity of an

ion is related to its m/z ratio so by measuring the time taken for an ionised peptide or protein to traverse the flight tube of the mass spectrometer, it is possible to determine precisely its molecular weight.

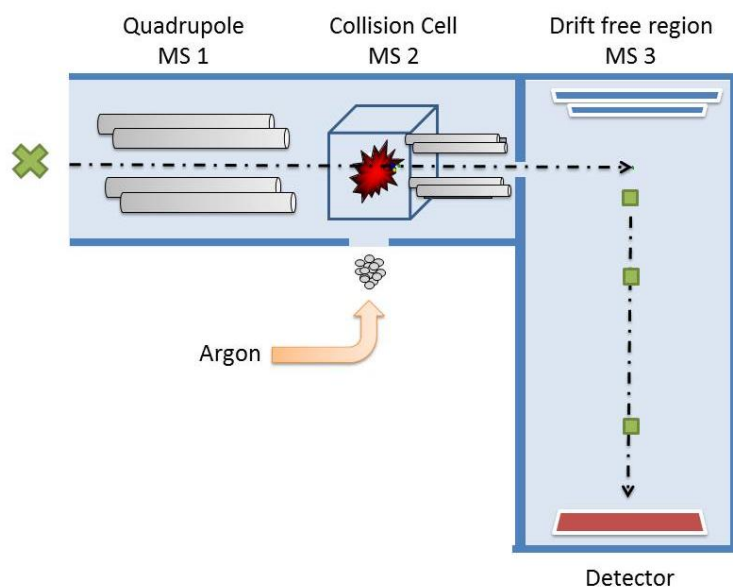


Figure 2.12. Schematic representation of a Q-ToF mass spectrometer

The mass spectrometer runs through a continuous cycle, where an initial mass spectrum is obtained for all peptides eluted from the column at a particular time, followed by tandem MS (MS/MS) analysis for the most abundant peptides. An MS/MS spectrum of the fragments consisting of y - and/or b -ions is generated. Fragment peaks that appear to extend from the amino-terminus are termed " b " ions and those from the carboxy-terminus are termed " y ions". Ideally a MS/MS spectrum for a particular peptide ion would consist of a series of y or b ions where each neighbouring fragment ion differed by one amino acid (Figure 2.13).

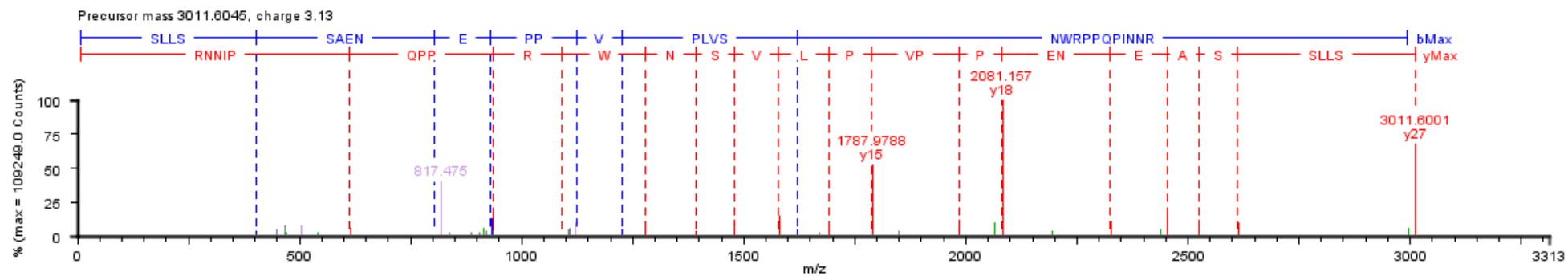


Figure 2.13. Example of an MS/MS spectrum.

The spectrum shows how peptides are sequenced following fragmentation in the QToF-MS. Each letter corresponds to a theoretical amino acid based on its mass.

2.2.6.6 Label free quantitative proteomics

Label free quantitative proteomics is a relatively recent approach to analysing ionised peptides. The process is based on the observation that the concentration of a detected ion is linearly proportional to the intensity of the 3 or 4 most abundant peptides in ESI-MS (Levin, Vouros et al. 2007). As a result protein samples can be spiked with a known concentration of non-human tryptically digested protein (i.e. an internal standard) and comparison made between the intensities of the peptide ions. It provides an alternative to gel based and isotopic labelling methods which only allow a limited number of samples to be compared at a time. The use of HPLC Q-ToF (or nano UPLC Q-ToF) MS for label-free quantitative proteomics allows each sample to be analysed individually and sequentially and puts virtually no constraint on the number of samples being compared. The ability to profile high numbers of proteins from multiple individual samples has huge potential for biomarker discovery research. It is also relatively low cost compared with isotopic labelling methods. However it is essential that reproducibility exists across multiple samples to ensure that detected differences reflect biological and not experimental variation. Therefore results from biomarker discovery experiments always require further validation and correct quality control.

2.2.6.7 Data acquisition and analysis

Following data acquisition, the MS spectra obtained from proteomic studies can be performed automatically. The analysis of ToF-MS spectra can be performed using software programs such as ProteinLynx Global SERVER™ (PLGS, Waters Corporation, Elstree, UK) and Progenesis LC-MS (Non-linear Dynamics, Newcastle upon Tyne, UK). To identify the protein, the peptide sequences are searched against known protein sequences in online databases such as the UniProt Knowledgebase (UniProtKB). The UniProtKB <http://www.ebi.ac.uk/uniprot> is an online open access database of proteins hosted by the European Bioinformatics Institute (EBI) which consists of two sections: UniProtKB/Swiss-Prot, which is manually annotated and reviewed and UniProtKB/TrEMBL which is automatically annotated and not reviewed.

Following the identification of a protein, its applicability as a biomarker requires validation. Traditionally immunoassay platforms such as Western blot or enzyme-linked immunosorbent assay (ELISA) have been used. Whilst this approach is suitable for the development of single biomarkers the process becomes inefficient and costly for multiple biomarkers. Instead a specific “targeted” assay can be developed using tandem mass spectrometry (MS/MS). These assays are rapid, highly specific, quantitative and easily multiplexed, making them suitable for translation for use in clinical laboratories.

2.2.6.8 Tandem Mass Spectrometry

A tandem mass spectrometer (MS/MS) use two quadrupole mass analysers coupled together in series (Figure 2.14). In a typical MS/MS analysis, ions of a selected mass (precursor or parent ions) are isolated by the first mass analyser (MS1) before transmission into a collision cell. In the collision cell ions collide with the neutral atoms of an inert gas usually argon, helium or nitrogen resulting in an increase in internal energy causing the molecule to fragment through CID. The subsequent fragments (product or daughter ions) produced are then analysed by the second mass analyser (MS2). Tandem mass spectrometers are highly specific and sensitive instruments which enable the identification and quantification of extremely low abundant molecules with a high degree of accuracy.

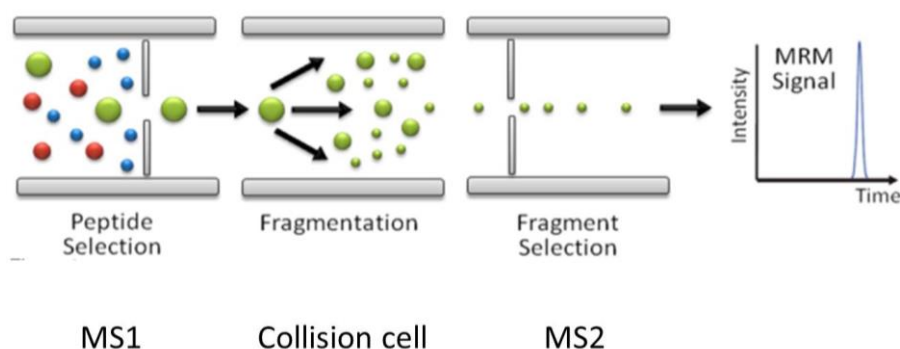


Figure 2.14 Schematic representation of a tandem mass spectrometer

Schematic of a tandem mass spectrometer operating in the multiple reaction monitoring (MRM) mode. The source generated ions are separated in MS1. The target peptide is allowed to pass into the collision cell, where it is fragmented. The characteristic fragment ion is allowed to pass through MS2 into the detector. The output from the detector is an ion intensity signal which can be optimised by tuning the instrument.

Configurations of a tandem mass spectrometer

The 1st and 2nd mass analysers (MS1 and MS2) can be operated in either fixed mass mode, whereby only ions of a certain mass are measured or in scan mode where ions of a range of masses are measured sequentially. By operating MS1 and MS2 in various combinations of fixed mass and scan mode several different types of tandem mass spectrometric analyses are possible.

Straight scan mode

The tandem mass spectrometer can be operated as a simple, single quadrupole mass spectrometer by utilising only the 1st mass analyser in scan mode. In straight scan mode, the mass analyser scans ions sequentially.

Product ion scan

In a product ion scan a precursor ion is selected in the 1st mass analyser (MS1) to be focussed into the collision cell to undergo CID. The 2nd mass analyser (MS2) is set to scan mode to analyse the characteristic fragmentation spectra of the product ions.

Precursor ion scan

In a precursor ion scan, MS2 is fixed for a specific product ion produced by CID in collision cell. Scanning of MS1, while MS2 is fixed allows all precursor ions which fragment to give a specific product ion or class of molecules to be analysed.

2.2.6.9 Multiple Reaction Monitoring

Selected reaction monitoring (SRM) is a quantitative MS technique that has been used for many years to analyse small molecules. Multiple SRM transitions can be measured within the same experiment by rapidly toggling between the different precursor/fragment pairs, in a technique called multiple reaction monitoring (MRM). It is capable of rapid, sensitive, and specific quantitation of hundreds of analytes in highly complex sample and is now being applied to biomarker studies.

Description of technique

Marker proteins are digested to peptides (typically with trypsin) within a sample matrix (tissue, plasma, urine). Tandem mass spectrometry (MS/MS) is then used in fixed mass mode to select a specific precursor peptide ion(s) in MS1 and after CID, select a specific product ion(s) in MS2. The precursor mass to product mass

“transition” is a unique mass signature for a peptide. Peptides from a complex sample can be further separated by using RP-HPLC by providing each peptide with a unique retention time (RT). The high accuracy and sensitivity of this technique makes it well suited to quantitative proteomic and metabolomics studies.

Absolute quantitation

Absolute quantitation of a peptide (and thus protein) can be achieved by MRM-MS/MS including an internal standard (IS). Ideally this would consist of the exact same sequence as the target biomarker peptide but with an incorporated isotopically labelled amino acid that will result in increased mass. These are termed AQUA (Absolute QUAntitation) peptides. Alternative internal standards may be used but they must be unique exogenous peptides or proteins, not present in the sample of interest.

Applications of MRM-MS/MS

Until recently proteomic studies have focused primarily on the global identification of as many proteins as possible. Now the precise, simultaneous quantification of the relative protein abundance of many proteins has become an important goal particularly for biomarker applications. Targeted proteomics experiments using MRM-MS assays are increasingly used in laboratories as it is a more cost effective, robust, reliable, and high throughput method than antibody based quantitation. Tandem MS instruments are also widely available in hospitals so translation into the clinical arena is feasible.

2.3 Cardiovascular Proteomics

Over recent years, proteomics has been applied to understand disease mechanisms, and the molecular basis of cardiac function, find new therapeutic targets and identify potential diagnostic and prognostic biomarkers (Cieniewski-Bernard, Acosta et al. 2008). Advances in MS analysis, better isolation and enrichment techniques and improved separation of organelles and membrane proteins, has enabled a more in depth look into the cardiac proteome (Lindsey, Mayr et al. 2015). The basic strategy for most disease mechanism studies has been to compare the protein complements of affected hearts (or plasma/serum) with normal controls, analogous to gene expression studies.

2.3.1 The human cardiac proteome

There have been progressive efforts to describe the cardiac proteome, i.e. all the proteins in the human heart. Although most studies have been performed using 2-DE, shotgun methods are becoming more widely used. These approaches should really be considered complementary because unless both methods are applied there are substantial subsets of proteins that remain undetected. The number of cardiac proteins identified is still increasing due to advances in proteomic methods and better annotation of the Swissprot database.

2.3.1.1 2D mapping

The first human cardiac proteome mapping exercise was undertaken by Dunn and colleagues in 1992. They used 2-DE and immunoblotting to separate, identify and catalogue 50 proteins from a left ventricular (LV) tissue sample from an unused transplant donor heart (Baker, Corbett et al. 1992). This served as the reference map for several years. Once mass spectrometry and narrow range IPGs became available they were able to improve the separation and resolution of the constituent proteins (Westbrook, Yan et al. 2001). This led to the publication five years later of a significantly enhanced 2D map (pH 3-10) consisting of 110 unique protein identifications (Westbrook, Wheeler et al. 2006). The vast majority of proteins were of mitochondrial origin reflecting the high abundance of mitochondria in the heart due to its high-energy demands. Surprisingly few contractile and regulatory proteins were identified. This was attributed to their tissue lysis method being more favorable

for cytoplasmic proteins. The limitations of 2-DE were acknowledged including the restricted dynamic range of proteins and suboptimal analysis of hydrophobic, extremely basic, acidic and low-abundance proteins. Recognizing that coverage could be substantially increased by studying sub-proteomes, Dunn's group further optimized the 2D reference map by focusing on basic proteins (pH 6-11). A novel reagent was added to prevent streaking in the gel and a paper-bridge loading technique was applied. These modifications led to a total of 151 unique proteins being identified. (Polden, McManus et al. 2011).

2.3.1.2 Gel-free methods

Attempts at gel-free mass spectrometric strategies were pursued as soon as the technology became available to do so. In the largest study of its kind, Kline et al performed shotgun proteomics using Multidimensional Protein Identification Technology (MudPIT) on human heart tissue explanted from two donors, one with a normal cardiac phenotype and the other diagnosed with idiopathic dilated cardiomyopathy (DCM) (Kline, Frewen et al. 2008). A total of 7138 unique proteins were identified (3746 from the non-failing heart and 3818 from the failing heart sample). The overlap of identified proteins between both samples was 3669 (51.4%). One of the most important outcomes of this work was to generate an extensive database of 2558 cardiac proteins and their "proteotypic" peptides i.e. the tryptic peptides most commonly identified by mass spectrometry. These peptides can potentially serve as a measure by proxy for protein abundance within a complex sample and would be ideal for use in targeted MRM-MS assays.

2.3.1.3 Combined approach

The most in depth analysis of the cardiac proteome to date used a multiplexed approach using two separation methods (2D-PAGE and SCX), three different proteases (trypsin, chymotrypsin and Lys-N) and two different peptide fragmentation techniques (CID and electron transfer dissociation, ETD). This combined approach identified 3584 unique proteins in a transmural LV human tissue sample and importantly enabled greater protein sequence coverage thus improved confidence in both identification and quantitative abundance of proteins (Aye, Scholten et al. 2010).

2.3.2 Cardiac sub-proteomes

A variety of methods have been developed to reduce sample complexity. These methods include differential centrifugation, flow cytometry, immune-based isolation, membrane protein enrichment strategies and/or density gradient isolation of organelles such as the nucleus or mitochondria (Sharma, Cosme et al. 2013). Most cardiac sub-proteome studies to date have been performed in animals. This reflects not only the ethical and practical barriers to obtaining human tissue but also the need to optimise analytical methods. These studies are essential to animal based research and also provide some insight into why animal disease models may differ from the human condition.

2.3.2.1 Mitochondrial sub-proteome

In the human heart, the mitochondrial sub-proteome has been best characterised. Studies carried out by Taylor et al. in 2003 used sucrose density centrifugation to enrich for mitochondrial proteins in the presence of detergent to aid solubility. By combining 1D SDS-PAGE with LC-MS/MS they identified 615 mitochondrial proteins (Taylor, Fahy et al. 2003).

2.3.2.2 Membrane proteins

In 2006 Kislinger et al. initially identified 519 murine cardiac proteins (Kislinger, Cox et al. 2006) but when they applied subcellular fractionation techniques 3 years later they identified a total of 4906 proteins (Bousette, Kislinger et al. 2009). Fewer plasma membrane proteins were proportionally identified than expected relative to the predicted proteome. Cell-surface proteins are notoriously difficult to isolate due to their relatively low abundance and hydrophobic nature.

2.3.2.3 Myofilament sub-proteome

The application of MRM-MS, has significantly improved the recognition of myofilament PTMs. This has led to a better understanding of the functional relevance of modification such as phosphorylation to myocyte contraction, and how this might be modified in disease (Peng, Ayaz-Guner et al. 2014, Ramirez-Correa, Martinez-Ferrando et al. 2014)

2.3.2.4 Extra-cellular matrix

More recently the myocardial extra-cellular matrix (ECM) was characterised in a porcine model of ischemia/reperfusion injury. Using a novel proteomic method based on decellularization and sequential extraction to enrich for ECM components, 139 matrix proteins were identified by LC-MS/MS. Differential expression analysis showed several proteins not previously identified in the heart (e.g. cartilage intermediate layer protein 1) that may contribute to cardiac remodeling after myocardial infarction (Barallobre-Barreiro, Didangelos et al. 2012).

2.3.3 Understanding Disease Mechanisms using Proteomics

Proteomics provides a method for understanding disease at a systems biology level. Rather than evaluating individual proteins or pathways, the changes in function and morphology at an organ level can be characterised. For example, the previously described global cardiac proteome studies have provided insight into mechanisms of heart failure (Kline, Frewen et al. 2008) and ischaemic reperfusion injury (Barallobre-Barreiro, Didangelos et al. 2012). Other human protein expression profiling studies have investigated the morphology of atherosclerotic plaques (Mayr, Grainger et al. 2009), changes in the function of mitochondria in atrial fibrillation (Goudarzi, Ross et al. 2011) and alterations in inflammatory dilated cardiomyopathy. (Hammer, Goritzka et al. 2011).

2.3.3.1 Proteomics in HCM

Despite the huge potential of proteomics to provide insight into disease mechanisms there has only been one small study in HCM. Lam et al used two-dimensional gel electrophoresis (2-DE) to study the global cardiac proteome of mice predisposed to developing HCM. Hearts from mice over-expressing a mutant cardiac troponin I gene (Gly203Ser) showed > 2-fold change in 34 proteins compared with controls. These included proteins involved in energy production (e.g. ATP synthases, NADH dehydrogenase), Ca²⁺ handling (e.g. S100 calcium binding protein A10 and A11), and cardiomyocyte structure (e.g. desmin and myosin). These expression level changes may be critical to understanding cardiac dysfunction and progression to heart failure (Lam, Tsoutsman et al. 2010).

2.3.4 Identification of Biomarkers using Proteomics

If proteins expressed differentially in diseased tissue can be detected in the plasma or urine, they may prove to be useful as clinical biomarkers. Mass spectrometry based proteomics has been successfully used to identify and characterise novel cardiovascular biomarkers. For example, secreted frizzled related protein 3 (sFRP3) was identified as being upregulated in a mouse model of dilated cardiomyopathy (Gramolini, Kislinger et al. 2008) and has subsequently been shown to predict poor outcomes in patients with heart failure (Askevold, Aukrust et al. 2014). Additionally myeloperoxidase (MPO), has been investigated as an early marker of coronary artery disease (Tang, Wu et al. 2011) with mass spectrometry being used to identify its mode of action (Geoghegan, Varghese et al. 2012) and potential therapeutic targets (Shao and Heinecke 2011).

It is increasingly apparent that multiple biomarkers are required to stratify complex disease – these may be in the form of multiplex immunoassays that enable detection of numerous different antigens or through multiplexed MRM-MS assays. The advantage of MRM-MS based quantitation is that it can be used for both biomarker discovery and biomarker validation. It provides the sensitivity and accuracy needed in the discovery phase, in addition to the high throughput needed for clinical validation. MRM-MS has been successfully applied to screen patients with coronary artery disease (CAD) using a panel of 44 protein biomarkers, resulting in the identification of at least two proteins not previously known to be correlated with CAD (Cohen Freue and Borchers 2012).

2.4 Cardiovascular Biomarkers

There are various definitions of a biomarker but a widely used form is that proposed by the Biomarkers Definitions Working Group, (Group 2001):

“A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”

Biomarkers may be macro-molecules (DNA, RNA, proteins, lipids, metabolites), cells, or processes that are detected and analysed in tissue, the circulation (e.g. blood or lymph), and in body fluids (e.g. urine or sputum). Biomarkers can be categorised as antecedent (identifying risk of disease), screening (watching for early disease), diagnostic (classifying obvious disease), staging (describing disease severity) and prognostic (predicting disease course including response to treatment).

Biomarkers reflect not only the disease but also the body's response, to having the disease or response to treatment. This response may be adaptive or maladaptive but either way will reflect the biological processes that give rise to sustain the disease phenotype. The principal objective of any clinical biomarker is to help the clinician provide the best possible care for their patient.

2.4.1 General Approach to Biomarker Research

An ideal biomarker should be highly specific for the condition in question (Figure 2.15). In cardiovascular disease, this means the marker is likely to be present at a high concentration in the myocardium, and potentially measurable in the peripheral circulation in proportion to the stage of disease. Analytical assays should be sensitive, standardized and reproducible with clearly defined normal ranges based on age, sex and ethnicity (Emdin, Vittorini et al. 2009). Any novel biomarker(s) must correlate with the clinical or subclinical phenotype, provide incremental value over established markers and be cost-effective to implement. Validation of putative biomarkers requires large, statistically powered sample sets. Which usually means access to high quality biobanks is a prerequisite (Ioannidis and Panagiotou 2011).

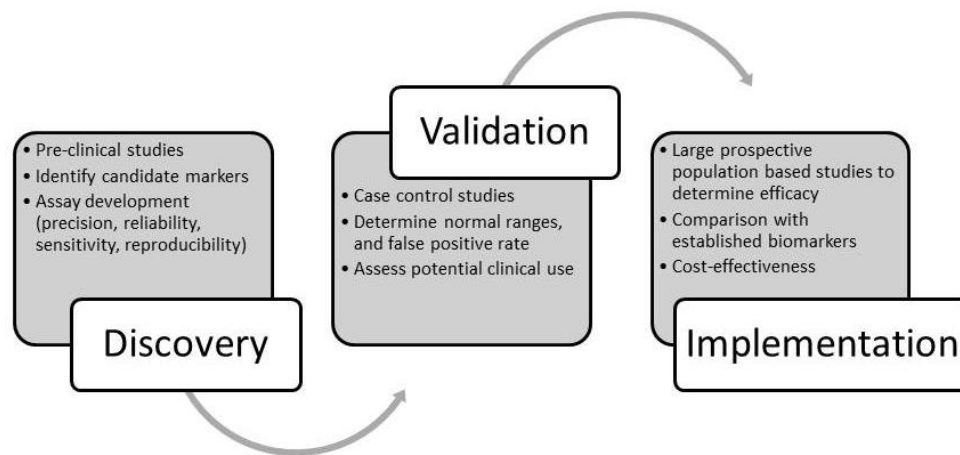


Figure 2.15. The key stages of biomarker development

2.4.2 Evolution of Cardiovascular Biomarker Research

The field of cardiac biomarkers began in 1954 when aspartate aminotransferase (AST) was the first enzyme used to identify myocyte necrosis or early myocardial infarction (Ladue, Wroblewski et al. 1954). Creatine phosphokinase (CPK) and lactate dehydrogenase (LDH) soon followed (Oliver 1955, Wroblewski, Rueggeger et al. 1956). Once the existence of iso-enzymes was understood, these biomarkers became more organ specific, for example CK-MB and LD-1 had greater expression in the myocardium. Many biomarkers explored in the early days were present in both cardiac and skeletal muscle, so their ratio, rather than absolute level, had to be defined. The discovery of cardio-specific isoforms of certain structural proteins, namely Troponin I and T has since revolutionized clinical practice (Jaffe, Ravkilde et al. 2000). These markers are now well established as highly specific indicators of heart muscle necrosis, and form part of the universal definition of myocardial infarction (Thygesen, Alpert et al. 2012). The next step-change in clinical practice was the recognition that B-type natriuretic peptides could be used to screen for congestive heart failure (Burnett, Kao et al. 1986). Subsequent discoveries and technological advances have led to numerous cardiac biomarkers being described in the literature but aside from the troponins and natriuretic peptides, few currently have an established role in clinical practice.

2.4.3 Biomarkers in Hypertrophic Cardiomyopathy

Genetic biomarkers can be helpful to make an accurate diagnosis of HCM (i.e. exclude genocopies (Table 2.2) and to identify relatives at risk of developing the condition. However, DNA sequence variants are of little use in staging disease, determining prognosis or gauging response to treatment. The diverse spectrum of disease, multifaceted pathophysiology, and an inability to link genotype with phenotype make HCM an ideal model for biomarker innovation and disease mechanism research.

2.4.3.1 Diagnostic biomarkers

Patients with HCM usually undergo clinical evaluation with a panel of blood tests to screen for disorders that cause or exacerbate myocardial dysfunction (e.g. thyroid disease, diabetes, and anaemia) and assess secondary organ dysfunction (e.g. renal and liver function). Additional biomarkers can be helpful to diagnose genocopies and should be guided by the presence of particular symptoms and signs (Coats, Heywood et al. 2015). Creatine phosphokinase (CPK) is a muscle enzyme, with predominant skeletal (98% CK-MM) and cardiac muscle (70% CK-MM, 30% CK-MB) isoforms. An elevated CPK in a patient with HCM is typically associated with glycogen storage diseases (Type II, III and IX), Danon disease (*LAMP2* gene) and some mitochondrial disorders. Measurement of α -galactosidase A is appropriate in males over the age of 30 years with HCM to screen for Anderson Fabry disease, an X-linked lysosomal storage disease which accounts for 1-3% adults patients with unexplained left ventricular hypertrophy (LVH) (Linhart and Elliott 2007).

2.4.3.2 Staging and Prognostic biomarkers in HCM

Established cardiovascular biomarkers such as brain natriuretic peptide (BNP), N-terminal pro-brain natriuretic peptide (NT-proBNP) and high sensitivity cardiac troponin T (hs-cTnT) and I (cTnI) have only recently been studied in HCM.

Troponins

Troponins are thin myofilament proteins that regulate the contraction of cardiac and skeletal muscles. They form a complex consisting of three subunits, cardiac Troponin I, C and T (cTnI, cTnC and cTnT). The cTnI subunit is not expressed in skeletal muscle, and cTnT only expressed at low levels, making these isoforms cardio-

specific. Troponins can be detected in plasma and their main application is in early detection of myocardial infarction. Elevated levels of cTnI and cTnT are also associated with higher risk of cardiovascular events and death in heart failure (Sundstrom, Ingelsson et al. 2009, Masson, Anand et al. 2012). High sensitivity (hs) assays can detect chronic sub-clinical myocyte necrosis. Around 50% of patients with HCM have elevated hs-cTnT, and abnormal levels are associated with the severity of LVH, left atrial size, progressive heart failure and embolic events (Moreno, Hernandez-Romero et al. 2010, Kubo, Kitaoka et al. 2013).

Natriuretic peptides

Pro-B-type natriuretic peptide (proBNP) is a cardiac neurohormone synthesized and secreted mainly from the ventricles of the heart in response to ventricular wall stress. After secretion, the 126-amino acid prohormone proBNP is cleaved into the bioactive C-peptide, brain natriuretic peptide (BNP) and an inert amino terminal peptide, N-terminal pro brain natriuretic peptide (NT-proBNP). In addition to their natriuretic action, BNP and NT-proBNP relax vascular smooth muscle, peripherally inhibit the sympathetic nervous and renin–angiotensin–aldosterone systems, and have direct lusitropic effects (de Lemos, McGuire et al. 2003). In patients with chronic heart failure, BNP and NT-proBNP are useful diagnostic and prognostic markers (Maisel, Krishnaswamy et al. 2002). In hypertrophic cardiomyopathy, BNP is expressed in ventricular myocytes with particularly high levels found in patients with left ventricular outflow tract obstruction (LVOTO) and diastolic dysfunction (Maron, Tholakanahalli et al. 2004, Pieroni, Bellocci et al. 2007) but its role in diagnosis and management of this disease has only recently been explored (Coats, Gallagher et al. 2013, Geske, McKie et al. 2013).

The prognostic role of NT-proBNP in HCM

Although natriuretic peptide concentrations correlate with clinical symptoms and markers of systolic and diastolic function in HCM, few studies had examined their relation to prognosis (Binder, Ommen et al. 2007). Therefore an original study was undertaken in the large cohort of patients with HCM managed at University College Hospital NHS Trust, London (Coats, Gallagher et al. 2013).

In a cohort of 847 patients (53 ± 15 years; 67% male) with HCM, 68 (8%) patients reached the primary endpoint of all-cause mortality or cardiac transplantation over a median follow-up of 3.5 years (inter quartile range 2.5 to 4.5 years). A single measurement of NT-proBNP was found to be a strong predictor of overall prognosis, with a serum concentration of ≥ 135 pmol/L associated with an annual event rate of 6.1 % (95% confidence interval 4.4 to 7.7) (Figure 2.16). NT-proBNP concentration correlated with increased left atrial size, LVOTO, atrial fibrillation, age and female gender but remained an independent predictor of the primary outcome.

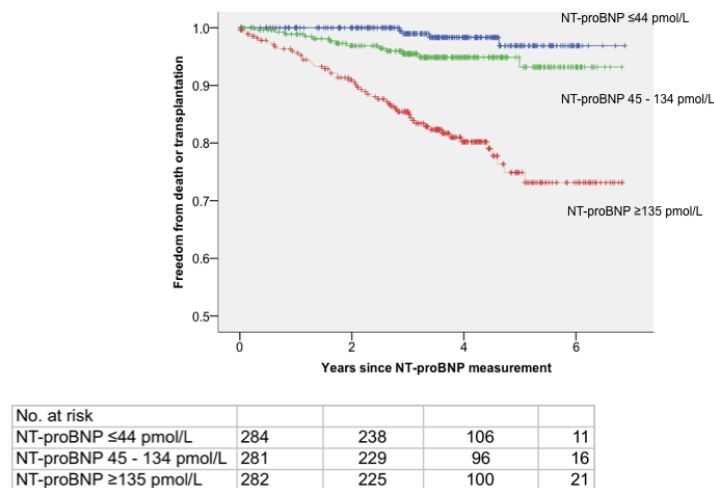


Figure 2.16 Prognostic role of NT-pro BNP in HCM.

Kaplan Meir analysis showing cumulative rates of survival in 847 patients with HCM stratified into tertiles according to serum NT-proBNP concentration.

Whilst NT-proBNP concentration was a significant predictor of heart failure and transplant related deaths, it was not a predictor of sudden death or appropriate implantable cardioverter defibrillator (ICD) shock, defined as therapy delivered for sustained ventricular tachycardia or ventricular fibrillation. This observation is important because most studies suggest heart failure is becoming the predominant mechanism of premature mortality in HCM. There is, therefore, a need for sensitive and specific biomarkers to identify patients at risk of adverse events to target preventative therapy.

2.4.3.3 Emerging Biomarkers in HCM

Several studies have explored biomarkers of inflammation, fibrosis, ischaemia and hemodynamic stress, reflecting the key pathways disturbed in HCM (Section 2.1.3.2 and 2.1.3.3). Studies have measured inflammatory markers including TNF- α , IL-6 and CRP. However, most studies are small and many report contradictory findings (Cambroner, Marin et al. 2009, Kuusisto, Karja et al. 2012). Emerging biomarkers such as Galectin-3 (Gal 3) and ST2 (also known as IL-1 receptor-like 1) directly link fibrosis and inflammation and have a prognostic role in heart failure (Bayes-Genis, de Antonio et al. 2014). A comparable marker, growth differentiation factor-15 (GDF-15) a member of the TGF- β superfamily is elevated in HCM (Montoro-Garcia, Hernandez-Romero et al. 2012).

2.4.3.4 The Need for Further Biomarker Research

There is a clinical need to not only better understand the biology of hypertrophic cardiomyopathy, but also to identify particular patterns of progressive disease such as heart failure, that might be amenable to early intervention. Proteomics offers an unbiased experimental approach to identify potential biomarkers directly related to underlying disease mechanisms.

3 Materials and Methods

This was a single-site cohort study conducted at The Heart Hospital, University College London NHS Trust. The materials and methods used in the pilot, discovery and validation phases of the study are described below. Method development and optimisation experiments are presented in Chapter 4.

3.1 Materials

3.1.1 Clinical samples

3.1.1.1 Heart tissue samples

Heart tissue samples were obtained from patients with HCM undergoing left ventricular septal myectomy. Informed consent was obtained prior to surgery (REC Ref. no. 04/0035 and 11/LO/0913). Portions of heart tissue from the left ventricular basal septum were divided and flash frozen in liquid nitrogen. The remaining tissue was fixed in formalin for paraffin embedding (FFPE). Age matched normal control samples of septal left ventricular tissue were obtained from unused transplant donor organs (through a research collaboration) following Human Research Ethics approval from St. Vincent's Hospital (#H03/118) and from The University of Sydney (#7326).

3.1.1.2 Urine samples

Urine samples were obtained from patients attending cardiomyopathy outpatient clinics at The Heart Hospital. Control samples were obtained from healthy volunteers working in the National Health Service (REC Ref. no. 11/LO/0913).

3.1.2 Chemical reagents

All standard reagents were obtained from Sigma-Aldrich Chemical Company (Dorset, UK). All solvents were LC-MS grade. Trypsin was of sequencing grade. All solutions were prepared using ultrapure water deionised and purified (ddH₂O) to a resistivity of 18.2 M Ω , using a Milli-Q system (Millipore, Massachusetts, USA).

3.1.3 Gels

3.1.3.1 1D gels

Pre-cast mini gels (Mini-PROTEAN® TGX Stain-Free™ Precast Gels) for 1D protein separation were obtained from BioRad, United Kingdom. Protein sample

buffer (Novagen ® 4X SDS) was obtained from EMD Chemicals, USA. A protein molecular weight marker (Amersham ECL Plex Fluorescent Rainbow Markers, GE Healthcare) was used to provide visible marker bands between 12 and 225 kDa.

3.1.3.2 2D gels

Large (16 x 18 cm) 15 % gels were cast in duplicate for 2D separation and analytical DIGE experiments. One litre of PAGE solution was made up from 500 mls acrylamide solution (30 % weight per volume), 250 mls of 1.34 M Tris-HCl, pH 8.8, 2.4 g of piperazine diacrylamide (PDA), 4.9 mls of 5 mM sodium thiosulphate and 0.5ng of ammonium persulphate. The acrylamide gel solution was poured between sialinised plates in the casting chamber, and the gels were over-layered with butan-2-ol, covered with cling film and allowed to polymerise at room temperature. Gels were used immediately or stored at 4 °C for up to a week. Isoelectric focusing strips (18 cm, pH 4-7 and pH 3-10) were obtained from Sigma Aldrich (Dorset, UK).

3.2 Proteomics Methods

3.2.1 Tissue Homogenization

Frozen myocardial tissue was pulverized into a powder using a pestle and mortar, containing liquid nitrogen. The powdered tissue was then homogenized in a 1 ml hand held glass homogenizer by treating with 1 ml of 50 mM ammonium bicarbonate buffer (AmBic), pH 7.4. Homogenates were aliquoted into 1.5 ml Eppendorf tubes separately and sonicated to disrupt the tissue using a Soniprep ultrasonic cell disruptor (5 x 10 second bursts at 10 mA). The samples were then pelleted by centrifugation at 13,000 g for 10 minutes in order to remove insoluble material. Pellets were stored separately at minus 80 °C.

3.2.2 Protein estimation

Protein estimation was performed for all tissue samples prior to analysis. 5 µl of supernatant of each sample was added to 45 µl of 50mM AmBic buffer in order to determine the concentration of protein.

3.2.2.1 Bicinchoninic acid (BCA) protein assay

For the discovery phase, a standard curve with known concentrations of bovine serum albumin (BSA) was prepared. Serial dilutions of the protein samples were also

prepared in the protein storage buffer and 5 µl of each dilution was mixed with 45 µl water. This was followed by the addition of 1 ml of bicinchoninic acid (Sigma-Aldrich, Dorset UK) to each sample which was incubated at 37 °C for 10 minutes. Twenty microlitres of copper sulphate solution was added and incubated at 37 °C for a further 20 minutes. The absorbance of each sample was measured in duplicate with a Cecil CE 2040 spectrophotometer (Cecil Instruments, Cambridge, UK) at a wavelength of 562 nm and the protein concentrations (mg/ml) were calculated from the standard curve generated using Excel 2013 (Microsoft, Washington, USA).

3.2.2.2 Bradford protein assay

In the validation stage, where a greater number of samples were analysed, a plate reader was used. Five microliters (µl) of supernatant from each homogenized tissue sample was pipetted in duplicate into separate wells on 96-well plates. Known concentrations of BSA, ranging from 0 – 2 mg/ml, were used to generate a standard curve for each plate. 200 µl of reagent, made to manufacturer's recommendation (Thermo Scientific, Massachusetts, USA), was added to each well. Each plate was gently agitated for 5 mins, and then warmed at 37 °C for 30 mins. A spectrophotometric plate reader (Tecan Infinite F200, Männedorf, Switzerland), set to a wavelength of 595 nm, analysed each plate. Average absorbance was determined for each duplicate. Excel 2013 (Microsoft, Washington, USA) was used to plot standard curves and determine protein concentrations.

3.2.3 Gel based proteomics

In the discovery phase, several methods to separate the heart homogenate mixture were appraised. Gel based methods employ a “top down” approach i.e. they separate the sample mixture at protein level. Gel spots or bands are stained, then excised, dehydrated and digested in order to extract peptides for mass spectrometry analysis.

3.2.3.1 1D SDS-PAGE

One dimensional protein separation according to molecular weight was performed using 10-well 12 % Mini-PROTEAN® TGX Stain Free Precast Gels (Bio Rad #456-8043) in a gel apparatus. The freeze dried heart tissue protein (50 to 200 µg) was reconstituted with 20 µl of solution containing DTE:H₂O:Blue (1:2:1) then heated at 90 °C for 5 minutes. Either 4 µl of molecular weight marker standard or 20 µl of

sample was placed in each well. Tank buffer was added and the gel run at 200 V until the sample had dispersed.

Preparation for protein identification

Coomassie blue was used for visualising protein bands on 1D gels. The stained gels were washed repeatedly in 300 µl of 50% methanol containing 0.1 % acetic acid and incubated at 37 °C. The 1D gel strip was cut into pieces according to the molecular weight markers. Each excised gel piece was placed in a silanised eppendorf that had been washed with 200 µl of methanol, to remove any contaminants. Each gel piece was washed three times with 200 µl of 50 mM ammonium bicarbonate buffer, pH 7.8 to remove any residual buffer or keratin from the surface of the gel. Gel pieces were partially dehydrated by addition of 500 µl of acetonitrile and mixed on a rotary mixer for 30 minutes. The acetonitrile was discarded and the gel pieces completely dehydrated by centrifugal evaporation for 1 hour at 37 °C. Disulphide bridges in the protein were reduced by the addition of 200 µl of 10 mM dithioerythritol (DTE) in 100 mM ammonium bicarbonate buffer, pH 7.8 and incubated for one hour at 37 °C. The gel pieces were allowed to cool to room temperature and the cysteine residues were carboamidomethylated by the addition of 300 µl of 55 mM iodoacetamide (IAA) in 100 mM ammonium bicarbonate and incubated at room temperature for 45 minutes in the dark. This step prevented the proteins from refolding and allowed the protease to digest the protein more efficiently. The excess solution was discarded then each gel piece was washed three times with 200 µl of 50 mM ammonium bicarbonate buffer, pH 7.8. Gel pieces were again dehydrated by addition of 500 µl of acetonitrile (ACN) and mixed for 30 minutes, prior to centrifugal evaporation for 1 hour at 37 °C.

3.2.3.2 2D-DIGE

Two-dimensional protein separation was performed using in-house protocols and gels (16 x 18 cm) which were cast in duplicate as described in Section 3.1.3.2.

Sample labelling

For DIGE analysis, 50 µg of freeze dried protein was reconstituted in 20 µl of DIGE label buffer consisting of 7M urea, 2M thiourea, 2% CHAPS, 10mM Tris-HCl pH 8.3. CyDyes were reconstituted 5 µl of high quality anhydrous dimethylformamide

(DMF) to make a stock solution of 1 nmol/ μ l. Samples were labelled with one of three CyDyes (Cy2, Cy3, Cy5) using 4 pmol CyDye to 1 μ g of protein then incubated in the dark for 30 minutes on ice. Where only 2 samples/groups were compared, a standard was created consisting of a pooled mix of the 2 samples (which was labelled with Cy2). The reaction was quenched with 1 μ l of 10mM L-lysine and incubated in the dark for a further 10 minutes. All three CyDye samples were combined to make a final volume of 90 μ l. Finally, 10 μ l of 7M urea, 2% CHAPS, 650mM DTE, bromophenol blue was added to make a total final volume of 100 μ l containing 150 μ g protein.

1st dimension:

A 24 cm IPG strip with broad pH range 3 to 10 was used. 350 μ l of rehydration buffer (containing 8M urea, 2M thiourea, CHAPS 4% w/v, resolyte, bromophenol blue and 65M DTE) was combined with the 100 μ l of CyDye labelled samples and incubated at 37°C for 30 minutes. The 450 μ l sample solution was applied evenly to the cassette well and the relevant IEF focusing strip placed face down and covered in mineral oil for at least 12 hours. Once rehydrated, the IEF strip was placed in the IEF manifold, gel-side up. Wicks were placed at either end and the anode and cathode clamped in place. Mineral oil was added so each strip was totally immersed. The strip was focused for between 75 and 100 KVolt hours, using a 0 to 300 Volt gradient for 30 mins then held at 300 V for 4 hours to slowly remove salt from the strips. A 300 to 1000 Volt gradient was then applied for 6 hours, followed by a 1000 to 8000 Volt gradient for 4 hours, prior to being held at 8000 Volts.

2nd Dimension:

The strips were placed in a re-equilibration tube, on a rotating mixer, immersed in a solution of 50mM Tris-HCl, pH6.8, 6M urea, 30% glycerol containing 10% SDS, 0.1% DTE for 15 minutes, to resolubilise the proteins and re-break any disulphide bonds that had reformed during focusing. The solution was then discarded and the strips were immersed in a solution of 50mM Tris-HCl, pH6.8, 6M urea, 30% glycerol containing 10% SDS, 0.125 % IAA with trace bromophenol blue for 15 minutes, to resolubilise the proteins, carboamidomethylate the cysteine groups and correct the pH for SDS PAGE.

Gel scanning and fixing

The gels were immediately scanned after electrophoresis using a Typhoon™ Variable Mode Imager (Amersham Biosciences). Low resolution scans were performed to determine the wavelength required for maximum spot intensity, prior to a final high resolution image acquisition at 100 dpi. After scanning gels were placed in 50 % methanol (7 % acetic acid) fixative at room temperature and gently agitated for at least 1 hour.

Quantitative analysis

The acquired gel images were analysed for quantitative and qualitative differences by Progenesis SameSpots Version 4.1 (Nonlinear Dynamics Ltd., Newcastle upon Tyne, UK) software program. One image was chosen and allocated as the reference gel to align all other images in the experiment. Following this, manual alignment, spot detection, background subtraction and normalisation were performed. To normalise the expression data, the data from one sample were fixed and all other samples were calibrated to this reference. Spots ≤ 250 units area were filtered out. Spots were selected for excision after silver staining to ensure sufficient protein was cut out for MS identification.

Silver staining

The gel was fixed for at least 60 minutes, prior to placing in sensitizing solution (1 litre solution included: 300 mls ethanol, 2g sodium thiosulphate and 68g sodium acetate) for a further 60 minutes. The solution was then removed by washing the gel four times with distilled water. A silver solution containing 2.5 g of silver nitrate was made up to 1 litre and 0.4 mls formaldehyde added prior to applying to the gel. The solution was then removed by washing the gel twice with distilled water. The gel was placed in developing solution (1 litre contained 25 g sodium carbonate and 0.8 mls formaldehyde) for 4 to 6 minutes until the gel spots started to appear. Finally the gel was placed in a stopping solution (1 litre contained 14.6 g EDTA- $\text{Na}_2 \cdot 2\text{H}_2\text{O}$). Gels were stored long-term in fixative solution (Shevchenko, Wilm et al. 1996)

Preparation for protein identification

Each spot of interest was excised using a scalpel and/or pipette tip and placed in a sialinised Eppendorf, washed in 50 mM ammonium bicarbonate pH 7.8 and

dehydrated using acetonitrile and centrifugal evaporation as described in 3.2.3.1. Disulphide bridges were reduced and cysteine residues carboamidomethylated prior to PAGE analysis so this did not need to be repeated.

3.2.3.3 Extraction of peptides from polyacrylamide gels

To enable identification by mass spectrometry, proteins were digested into peptides. Proteolysis was performed by adding 60 µl of a solution of trypsin (15 ng/µl in 40 mM ammonium bicarbonate and 9 % ACN) to a dehydrated gel piece or spot, and incubating for 12 hours or overnight at 37 °C. After digestion 200 µl of 1% formic acid was added to the gel piece, vortexed and mixed for 20 minutes. The wash containing hydrophilic peptides was transferred to a clean silanised Eppendorf. 300 µl of 50% ACN containing 1% formic acid was added to the gel piece, vortexed and mixed for 20 minutes. This wash containing most of the tryptic peptides was transferred to the same Eppendorf containing the previous 1% formic acid wash. Finally a further 300 µl of 50% ACN containing 1% formic acid was added to the gel piece, vortexed and mixed for 20 minutes. This wash was again transferred to the same Eppendorf containing the previous two washes. 200 µl of 1% formic acid was added to the washes, vortexed and freeze dried (Heywood, Madgett et al. 2011)

3.2.3.4 Identification by mass spectrometry

The tryptic digest from each gel band/spot was subject to analysis by mass spectrometry using a nanoAcquity UPLC and Q-ToF Premier mass spectrometer (Waters Corporation, Manchester, UK) and identification using ProteinLynx Global SERVER™ (PLGS) Version 2.4 (Waters UK). See Section 3.2.4.3 and 3.2.4.4.

3.2.4 Label Free proteomics

Label free methods involve separation of a complex mixture at peptide level. A pre-determined amount of protein (between 500 and 1000 µg) for each experiment was aliquoted and freeze dried prior to in-solution digestion. The peptide mixture is then fractionated to reduce sample complexity prior to LC-MS analysis using methods developed in our laboratory (Bostanci, Heywood et al. 2010)

3.2.4.1 In-solution digestion

Freeze dried myocardial protein extracts were reconstituted in 20 μ l of 100 mM Tris, 1% amidosulfobetaine14 (ASB-14) at pH 7.8 containing 6 M urea and vortexed well. 1.5 μ l of a solution of DTE was added, vortexed and left at room temperature for 60 minutes. This was followed by the addition of 6 μ l of iodoacetamide (36 mg of IAA in 1 ml of 100 mM Tris, pH 7.8) then vortexed and left at room temperature for a further 30 minutes. 155 μ l of distilled water was added. Finally, 10 μ l of trypsin solution was added (1 μ l reagent in 10 μ l reaction buffer). Samples were incubated in a water bath at 37 °C overnight.

3.2.4.2 Offline 2D-LC peptide separation

Sample complexity can be reduced in several ways by making use of the variable chemical properties (e.g. hydrophobicity) of the peptide mixture or “digest”. The approach used in this study was reversed-phase (C18) chromatography. Isolute C18 chromatography columns (50 mg/1 ml Biotage, Hengold, UK) were primed with 1 ml of 50 % ACN containing 0.1 % ammonia and 2 mls 0.1 % ammonia. One hundred microliters of tissue digest and 100 μ L of 0.2 % ammonia were combined and applied to the pre-conditioned C18 column and allowed to flow through under gravity. This breakthrough fraction was captured and reapplied to the column and the subsequent elution retained. A 500 μ L solution of 0.1% ammonia was then applied to the column and the eluting wash fraction collected and retained. Thereafter peptides were eluted from the column with 500 μ L of a solution containing an increasing percentage of ACN (3 %, 5%, etc and 100 %) with 3% ammonia and each fraction captured. All fractions were then lyophilised and reconstituted in 30 μ L of 3% ACN containing 0.1% trifluoroacetic acid (TFA) prior to label free quantitative analysis by UPLC-QToF-MS (Section 3.2.4.3).

3.2.4.3 Label Free Quantitation by Nanoacuity UPLC-QToF-MS

Prior to analysis 18 μ l of each digest was spiked with 2 μ l (200 fmols) of a yeast enolase tryptic peptide (*Saccharomyces cerevisiae*). All analyses were performed using a nanoAcquity UPLC and Q-ToF Premier mass spectrometer (Waters Corporation, Manchester, UK). Peptides were trapped and desalted prior to reverse phase separation using a Symmetry C18 5 μ m (5 mm x 300 μ m) pre-column.

Peptides were then separated prior to mass spectral analysis using a 15 cm x 75 μ m C-18 reverse phase analytical column. Peptides were loaded onto the pre-column in 0.1 % formic acid (FA) at a flow rate 4 μ l/min, for a total time of 4 minutes. Peptides were eluted off the pre-column and separated on the analytical column using a gradient of 3 to 40 % acetonitrile (containing 0.1% FA) over a period of 60 minutes and at a flow rate 300 nl/minute. The column was washed and re-generated at flow 300 nl/minute for 10 minutes using a 99% acetonitrile (containing 0.1% FA) rinse. After all non-polar and non-peptide materials were removed, the column was re-equilibrated at the initial starting conditions for 20 minutes. Column temperature was always maintained at 35 °C. Mass accuracy was maintained during the run using a lock spray of the peptide [glu1]-fibrinopeptide B delivered through the auxiliary pump of the NanoAcquity at a concentration 300fmol/l and a flow rate 300 nl/minute (Manwaring, Heywood et al. 2013).

Operating parameters

Peptides were analysed in positive ion mode using a Q-ToF Premier mass spectrometer (Waters Corporation, Manchester, UK) which was operated in v-mode with a typical resolving power of 10,000 FWHM. Prior to analyses, the ToF analyser was calibrated using [glu1]-fibrinopeptide B fragments obtained using collision energy of 25 eV and over the mass range m/z 50 to 2000. Post calibration of data files were corrected using the doubly charged precursor ion of [glu1]-fibrinopeptide B (m/z 785.8426) with a sampling frequency of 30 seconds. Accurate mass LC-MS data were collected in a data independent and alternating low and high collision energy mode. Each low/high acquisition was 1.5 second with 0.1 second inter-scan delay. Low energy data collection were performed at constant collision energy of 4 eV, high collision energy acquisitions were performed using a 15 to 40 eV ramp over a 1.5 second time period and a complete low/high energy acquisition achieved every 3.2 seconds (Manwaring, Heywood et al. 2013).

3.2.4.4 Bioinformatics

Identification of proteins from peptides

The identity of peptides was determined by exporting the MS/MS output files from the matched analysis into the ProteinLynx Global SERVER™ (PLGS) Version 2.4, Waters UK and interrogating it against a human protein database downloaded from the UniProt/Swiss-Prot website (<http://www.uniprot.org>) on 26th March 2013. The sequence of the yeast enolase (P00924) and porcine trypsin (P00761) were added to the database. The following conditions were fixed to carry out a positive identification for a protein:

- More than three fragment ions per peptide
- Five fragment ions per protein
- Two or more peptides per protein
- Maximum protein mass 800,000 Da

These conditions had to be satisfied otherwise a false-positive result with a low matching confidence would be likely to appear when the conditions fall lower than the fixed values. The lock mass window was set at 0.1 Da, low and high energy thresholds set at 150 and 50 counts, respectively. Elution start and end times were 5 and 50 minutes respectively and the chromatographic peak width was set to automatic. Carbamidomethylation of cysteine was a fixed modification and deamidation of asparagine/glutamine and oxidation of methionine were variable modifications. An additional search for post translational modifications included methylation, ubiquitination and carbonylation which were considered as optional modifications. Other parameters used included a < 10 ppm mass accuracy tolerance and up to two missed cleavage sites.

Quantitation of proteins

Two different approaches were used to compare the protein content of myocardial samples from donor control hearts with those from patients with HCM.

Estimate of absolute quantity

The intensities of enolase peptides were employed as a marker for other peptides as the quantity of enolase added to each patient sample was known, thus the

quantitation (fmol) should be absolute. The minimal detectable concentration was set to 0.1 fmol. Results of the identified proteins with fmol concentrations were exported and analysed in Microsoft excel by the formula and VLOOKUP function. These were used to compare presence and absence of proteins in different sample groups and calculate fold change of concentration of proteins between each group.

Relative quantitation

The acquired spectra were also loaded into Progenesis LC-MS software (version 3.0, Nonlinear Dynamics, Waters Corp.) and label free quantification was performed. This commercial software identifies and quantitates peptides and proteins based on ion abundance. For each fraction the MS spectra were transformed using a proprietary algorithm and stored in peak lists comprising m/z and abundance. One sample was set as a reference, and the retention times of all other samples within the experiment were aligned (3 to 5 manual landmarks, followed by automatic alignment) to create maximal overlay of the two-dimensional feature maps. At this point, features with more than 6 charges were masked and excluded from further analyses and all remaining features were used to calculate a normalisation factor for each sample that corrects for experimental variation. Samples were then allocated to their experimental group (e.g. control donor heart or HCM). For quantitation, all unique peptides of an identified protein were included and the total cumulative abundance was calculated by summing the abundances of all peptides allocated to the respective protein. No minimal thresholds were set for the method of peak picking or selection of data to use for quantification. Statistical analysis was performed using the “between subject design” and p-values were calculated by a repeated measures analysis of variance (ANOVA) using the sum of the normalised abundances across all runs. After processing of all samples from the pre-fractionation, the quantification files were merged into a complete data set. Differences between groups were assessed by identifying fold changes > 1.5 with a statistical significance level of $p < 0.05$.

Protein expression profiling and pathway analysis

The identified proteins were submitted for gene ontology (GO) analyses using PANTHER (<http://www.pantherdb.org/>) and DAVD functional annotation tools

(<http://david.abcc.ncifcrf.gov>) to define their cellular compartment, biological process and molecular function (Thomas, Campbell et al. 2003, Huang da, Sherman et al. 2009). Further analysis was performed using ProfCom GO via the BioProfiling.de (<http://bioprofiling.de>) analytical toolkit. ProfCom GO aims to profile experimentally identified genes against a reference set (usually all genes from the organism) in order to understand groups that are over or underrepresented in the sample set. Proteins with statistically significant differences for one or more comparisons were examined further for the presence of overrepresented pathways using Ingenuity Pathway Analysis (<http://www.ingenuity.com/>) software and Reactome (<http://www.reactome.org/>).

3.2.5 Development of targeted proteomic tests using multiple reaction monitoring

A Xevo™ TQ-S triple quadrupole mass spectrometer was used for validation experiments and the development of a clinical assay using multiple reaction monitoring (MRM). The instrument was operated in positive ion mode. The capillary voltage was maintained at 3.7 kV with the source temperature held constant at 150 °C. Nitrogen was used as a nebulizing gas at a flow rate 30 litres/hour. A Waters Acquity UPLC BEH C18 column (1.7 µm, ID 2.1 x 100 mm) was used for separation using two solutions, Solution A (99.9% LC-MS grade water with 0.1% FA) and Solution B (LC-MS grade 99.9% ACN with 0.1% FA). The flow rate was set to 0.8 ml/min and a linear gradient of 0 to 97 % Solution A over 7 minutes (Table 3.1). The total run time was 10 minutes due to loading, cleaning and reconditioning. A 5-10 µL injection volume was used with partial loop and needle overfill mode.

Time (mins)	Flow rate (ml/min)	% A	% B	Gradient
0	0.8	97	3	0
0.2	0.8	97	3	6
7.0	0.8	60	40	6
7.01	0.8	0.1	99.9	6
8.0	0.8	0.1	99.9	6
8.01	0.8	97	3	1
10	0.8	97	3	1

Table 3.1. LC inlet parameters

3.2.5.1 Peptide selection

Candidate peptides were selected using previously acquired QToF mass spectrometry data, supported by online resources such as the Global Proteome Machine (<http://www.thegpm.org/>) and Skyline (V3.1, MacCross Lab, University of Washington). Ideal candidates had high intensity values for the protein of interest, and multiple charged states (James and Jorgensen 2010). Candidate peptides were BLAST searched on UniProt (<http://www.uniprot.org/>) to ensure they were unique for the protein of interest in humans and to identify natural variants and known post translational modifications. Once selected, candidate peptides were custom synthesised by GenScript, USA or Generon, UK for MS tuning, optimisation and generation of standard curves.

3.2.5.2 MS tuning and optimisation of synthetic peptides

Depending on their solubility, peptides were either re-constituted in 0.1% formic acid or 50 % ACN and freeze dried in 1 mg aliquots. If solubility was a problem a small amount of aqueous acetic acid (for basic peptides) or aqueous ammonia (for acidic peptides) was added. Synthetic peptides containing cysteine residues required carboamidomethylation prior to use and were thus treated with DTE followed by IAA, prior to eluting through a C-18 column (5188-2750 Peptide Clean-up C18 Spin Tubes, Agilent Technologies). To determine optimal parameters for peptide detection on the Xevo TQ-S, an initial scan of the peptide was performed at an infusion rate 100 µg/ml to confirm the mass and to determine which charge state gave the best response. The operating conditions are shown in Table 3.2:

MS1	
Low Mass resolution	2.8
High Mass resolution	14.9
Ion energy scan speed	100-2000 amu/second
Desolvation temperature	600 °C
Capillary voltage	2.63 V

Table 3.2. Instrument settings for optimising peptide tuning

The automatic tuning facility “Intellistart” was then used to determine the best product ions from each precursor ion. If this failed to identify product ions, manual

optimisation was performed by performing multiple MS2 scans with various collision and cone energies. Findings were verified by using the online open access tool, Skyline (<https://skyline.gs.washington.edu>). Once transitions were created, the synthetic peptide was run through the LC system to determine the retention time of the peptide to enable identification of the endogenous peptide within the sample matrix. This meant there were three levels of identification of the peptide: precursor and product ion mass and retention time. The optimum transitions were determined by the highest intensity with no interfering signals from adjacent peaks. At least 3 transitions per peptide were selected and in some cases, two peptides per protein. A multiplexed assay was developed by combining MRM files into the MS method, grouped according to retention time. Solvent delays were added to the start and end of each method to reduce non-specific material entering the mass spectrometer and to maintain sensitivity.

3.2.5.3 Sample optimisation for biomarker quantification

For accurate quantitation of peptides in myocardial tissue the previously described digestion protocol was performed (Section 3.2.4.1). A quality control (QC) stock of myocardial protein was prepared for method development using the relevant sample matrix. Various concentrations and amounts of the QC digest were injected into the LC-MS to determine the optimal amount required for detection. Two internal standards (whole yeast enolase protein and a unique non-human custom synthesized peptide) were evaluated to enable an accurate quantitative assay to be developed. Quantitative data for each peptide were taken from one transition with a second transition included for identification purposes only. Where more than one peptide was used per protein, the correlation between peptides was examined. Optimisation of the MRM-MS/MS assay is presented in Chapter 6.

3.2.5.4 Standard Curve

For quantitation, a calibration curve was constructed in water and in the sample matrix to determine the linear range of specific peptide concentration and the ion suppression effect of the biological matrix. This was done by spiking the QC sample with increasing amounts (10, 20, 30, etc. fmol) of the commercially synthesized marker peptides used to quantitate the proteins. Limits of detection (LOD) were

established by reducing amounts of peptide in water. Inter and intra batch variation was determined by repeat injection of the same sample digest within one batch run and weeks apart. The QC sample was also injected at pre-determined intervals on each run to check sensitivity.

3.2.5.5 Data analysis.

All data were processed using Waters Mass Lynx 4.1 and QuantLynx, Manchester, UK). Peak integration (including smoothing and baseline subtraction) was performed according to software instructions. The integrated peak area of the peptide was divided by the area of the internal standard and absolute quantitative amounts were determined from the standard curve. Statistical methods are described in Section 3.5.

3.2.6 Urinary protein quantitation

To determine whether proteins of interest were detectable in peripheral body fluids, a pilot study was conducted applying the multiplexed MRM assay to urine samples collected prospectively. Each urine sample was thawed and well mixed then centrifuged for 10 minutes at 3,000 g to pellet the insoluble fraction. Intact yeast enolase (0.1 µg) was added to 1 ml of urine supernatant and transferred to Amicon Ultra centrifugal filters with a 3 kDa cut-off (Merck Millipore Ltd., Ireland). Samples were centrifuged for 50 minutes at 4,000 rpm. The supernatant was recovered and transferred to 1.8 ml Eppendorf tubes and freeze dried for subsequent tryptic digestion as described in Section 3.2.4.1. Urinary concentrations of compounds of interest were corrected for dilution differences by measuring the urinary creatinine concentration (because creatinine is filtered but not absorbed by the kidneys) by LC-MS using the Biological Mass Spectrometry Unit's previously published method (Mills, Morris et al. 2005).

3.3 Histopathology

Tissue samples were collected and a small portion frozen. The remainder was fixed in formalin before paraffin embedding (FFPE). A clinical histopathological report was produced based on haematoxylin and eosin (H&E) staining and light microscopy. Immunohistochemistry was performed subsequently for selected

proteins to validate the differences identified by the expression proteomics study. A tissue array was constructed to increase sample throughput and reduce costs.

3.3.1 Construction of a tissue array

Two tissue microarray (TMA) blocks were assembled from archival FFPE tissue blocks. All patients (n=38) had undergone septal myectomy between 2009 and 2010 and were selected consecutively. A proportion (n=18, 47 %) had undergone targeted or global proteomic analysis. Haematoxylin and Eosin (H&E) stained histology slides of each case were marked to include an area representative of the myocardium. Areas of disarray were identified according to the criteria of Maron and Roberts (Maron and Roberts 1979). Two cylindrical 2 mm cores were selected from the specimen, representing disarrayed and non-disarrayed myocardium. Vessels or large areas of fibrosis were avoided. Any ambiguity in identifying appropriate sampling areas was resolved at the multi-headed microscope with a consultant pathologist who specialised in cardiovascular pathology. The TMA block was constructed manually following Great Ormond Street Hospital Histopathology department protocol and grid map created with Microsoft Excel. Cylindrical cores of paraffin embedded tissue were removed from the archival donor paraffin block and inserted into a recipient block. Non-myocardial tissue (liver and kidney) were included in the array as control and marker specimens.

3.3.2 Immunohistochemistry

Sections (4 µm thick) from the tissue microarray block were cut and mounted on Leica Xtra® adhesive slides. The slides underwent xylene de-paraffinisation and rehydration through graded alcohols before being incubated with the primary antibody. Sections were stained using commercially produced antibodies including Immunohistochemical staining was performed using the Bond-Max system (Leica Microsystems). Staining was visualised using Leica Bond Refine detection kit and Bond-Max protocol F. Slides were scanned using a Hamamatsu NanoZoomer 2.0-HT slide scanner (Hamamatsu Photonics UK, Hertfordshire, UK) and analysed (blinded) using NanoZoomer Image Analysis Software for quantifiable differences. Each case was scored according to the intensity and proportion of staining in the specimen.

3.4 Clinical Evaluation

3.4.1 Phenotyping

All patients were evaluated in the inherited heart disease clinic at The Heart Hospital. Patient demographics, symptoms, family history and medication were documented. All patients underwent 12-lead electrocardiogram (ECG), exercise testing, 24-hour ECG monitoring, echocardiography, and and/or cardiac magnetic resonance (CMR) imaging. With consent tissue, blood and urine samples were obtained either at outpatient visits or during cardiac surgery. Routine biochemistry and haematology was processed through the University College London Hospitals clinical pathology laboratories. Serum N-terminal pro BNP was measured by a two-site electrochemiluminescence immunoassay on a Roche E170 analyser.

3.4.2 Genotyping

Patients who consented to genetic testing were screened using next generation sequencing for 41 known genes involved in inherited cardiomyopathies. Genomic DNA was extracted from blood using Qiagen DNA extraction kits. For each sample, 3 µg of genomic DNA was prepared for custom capture with the Agilent SureSelectXT Target Enrichment System. The captured DNA was sequenced with 100bp reads on the Illumina HiSeq2000 platform at a minimum of 50x coverage using an optimised multiplexing protocol. A bioinformatics pipeline was developed by UCL genomics for data analysis (Lopes, Zekavati et al. 2013). Any pathogenic mutation was subsequently confirmed by Sanger sequencing in-house at University College London or as part of a research study (A Coruña, ISCIII PI081469) at Instituto de Investigación Biomédica de la Universidad de A Coruña (Nunez, Gimeno-Blanes et al. 2013).

3.5 Statistical Analysis

Statistical analysis was performed using SPSS Statistics version 19 (IBM, Somers, New York) and Graph Pad Prism (Version 5). Checks were made to ensure continuous variables were normally distributed (Shapiro-Wilk). If not they were log-transformed for bivariate testing. Normally distributed data are presented as mean \pm SD, and non-transformed data as median and interquartile range (IQR). Comparisons between groups were performed by 1-way analysis of variance (ANOVA) with post-hoc Bonferroni correction. The Chi-square test or Fisher exact test was used to compare discrete data as appropriate. Receiver-operating characteristic (ROC) curve analysis was performed to define the diagnostic accuracy of specific biomarkers. Correlation between continuous variables was assessed using Pearson's r correlation coefficient. Statistical significance was defined as $p < 0.05$. Principal Component Analysis (PCA) was performed using Progenesis LC-MS software (Version 3.0, Nonlinear Dynamics, Waters Corp.) which uses feature abundance levels across sample fraction runs to determine the principle axes of abundance variation.

4 Analysis of the Cardiac Proteome

4.1 Introduction

Protein separation is a key component of any proteomic study. Sample preparation has an important effect on tissue, protein and peptide separation and the subsequent analysis. The optimal method to separate the cardiac proteome has not been determined and likely requires a combination of several technologies. Most studies on the human heart have used whole tissue (usually left ventricular) lysate. Previous studies defining the human cardiac proteome are summarised in Section 2.3.1.

4.2 Comparison of proteomic profiling techniques

Before undertaking a large scale proteomic study using human samples, different approaches were evaluated using a pooled sample set. The goal was to determine the optimal method to permit maximum protein identifications whilst establishing a reasonable workflow for analysing these samples individually.

4.2.1 Study cohort

For this preliminary work, identical pooled heart samples were used to enable a direct comparison of methods. Myocardial samples from 17 unrelated individuals were pooled into three groups. This enabled an initial comparison between control (n=6) and HCM (n=11) samples, and between genetic sub-groups *MYH7* (HCM Group 1, n=5) and *MYBPC3* (HCM Group 2, n=6) (Figure 4.1). The patient characteristics are shown in Table 4.1. Three out of six control subjects died from a sub-arachnoid hemorrhage – although they were considered to be suitable organ donors, cardiac complications including arrhythmias and myocardial infarction are recognized in 14% of patients with aneurysmal SAH (Ahmadian, Mizzi et al. 2013).

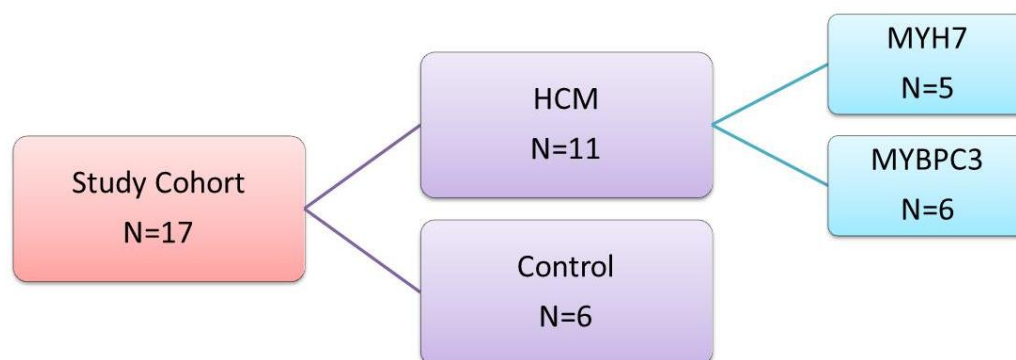


Figure 4.1 Description of study cohort

Subject	Sex	Age	Cause of death			
Control 1	Female	49	Sub arachnoid haemorrhage			
Control 2	Female	48	Sub arachnoid haemorrhage			
Control 3	Female	42	Road traffic accident			
Control 4	Male	45	Unknown			
Control 5	Female	42	Sub arachnoid haemorrhage			
Control 6	Female	31	Cerebrovascular accident			
Subject	Sex	Age	Gene	DNA change	Mutation	Exon/ Intron
HCM 1.1	Female	73	<i>MYH7</i>	c.A2539G	p.K847E	E22
HCM 1.2	Male	37	<i>MYH7</i>	c.A4664G	p.E1555G	E34
HCM 1.3	Male	41	<i>MYH7</i>	c.C2605T	p.R869C	E22
HCM 1.4	Male	20	<i>MYH7</i>	c.G1816A	p.V606M	E16
HCM 1.5	Male	70	<i>MYH7</i>	c.G1063T	p.A355S	E12
HCM 2.1	Male	65	<i>MYBPC3</i>	c.G2497A	p.A833T	E25
HCM 2.2	Female	63	<i>MYBPC3</i>	c.2309-2A>G	IVS23-2A>G	I23
HCM 2.3	Male	61	<i>MYBPC3</i>	c.926+8C>T	IVS11+8C>T	I11
HCM 2.4	Male	36	<i>MYBPC3</i>	c.1928-2A>G	IVS20-2A>G	I20
HCM 2.5	Female	32	<i>MYBPC3</i>	c.C1504T	p.R502W	E17
HCM 2.6	Male	23	<i>MYBPC3</i>	c.C1504T	p.R502W	E17

Table 4.1 Characteristics of study cohort

4.2.2 Methods for comparison

A preliminary characterisation of the human HCM proteome was conducted using three different protein/peptide separation techniques, before one method was chosen for the comprehensive proteomic analysis. The first two methods involved separation at whole protein level, after which gel spots or bands were excised and subjected to in-gel tryptic digestion. The third method involved tryptic digestion prior to separation at peptide level. All peptides were identified using liquid chromatography tandem mass spectrometry (LC-MS/MS). The three analytical methods (Figure 4.2) used for this initial comparison were:

- A. Differential gel electrophoresis (2D-DIGE)
- B. Gel-LC-MS/MS: separation of proteins by 1D-SDS PAGE before LC-MS/MS analysis (a top down approach)
- C. Peptide fractionation using high pH C18 chromatography before LC-MS/MS) (a bottom up approach)

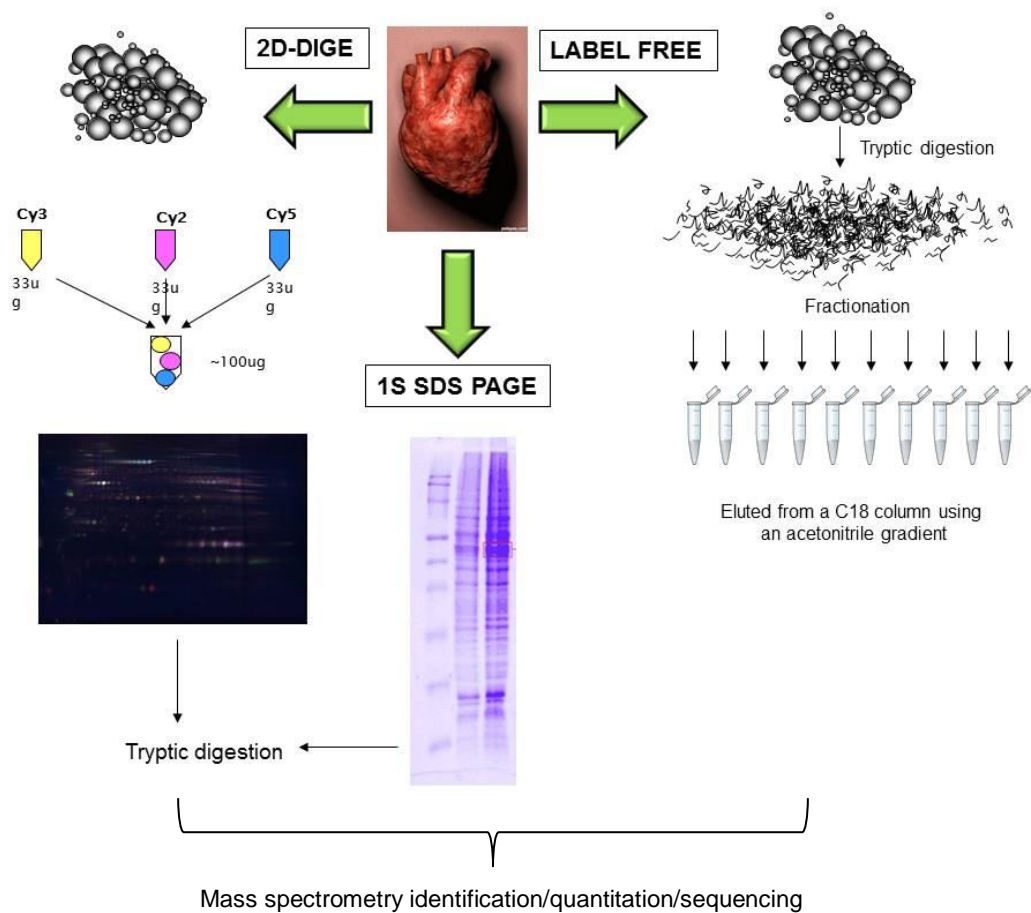


Figure 4.2 Illustration of three techniques to investigate the HCM proteome
Method A: 2D-DIGE, Method B. Gel-LC-MS/MS: separation of proteins, Method C: Peptide fractionation using high pH C18 chromatography.

Method A: 2D-DIGE

50 µg of freeze dried protein from each pooled heart homogenate was reconstituted in 20 µl of DIGE label buffer. Samples were labelled with one of three CyDyes (Cy2, Cy3 and Cy5), added to IPG strip rehydration buffer and applied to a pH 3 to 10 gel as described in Section 3.1.3.2. Image analysis for differentially expressed spots was performed using Progenesis SameSpots software version 3.1 (Nonlinear Dynamics Ltd, Newcastleupon Tyne, U.K.). Normalised protein spots in the Cy5 and Cy3 channels (HCM) were compared to the Cy2 (control) sample to generate a ratio of relative amount.

Results

One hundred and three discrete spots were identified. Four spots were up-regulated ≥ 2 -fold and 17 spots ≥ 1.5 -fold in HCM samples compared with control. Twenty-three spots were down-regulated ≥ 2 -fold and 69 spots ≥ 1.5 -fold in the HCM samples compared with control. Gels were silver stained and spots altered > 2 -fold were excised and digested with trypsin for QTOF analysis according to previously described methods (Section 3.2.3.4). Of the 27 spots submitted for MS analysis, 12 were successfully identified as having one or more proteins present (Table 4.2). The scanned 2D-DIGE gel is shown in Figure 4.3.

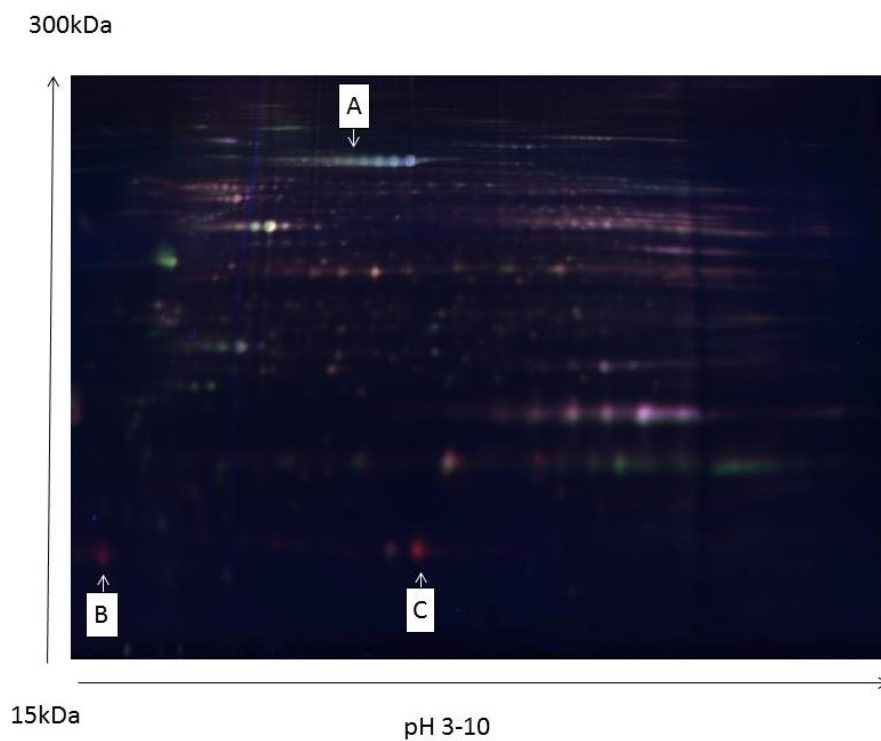
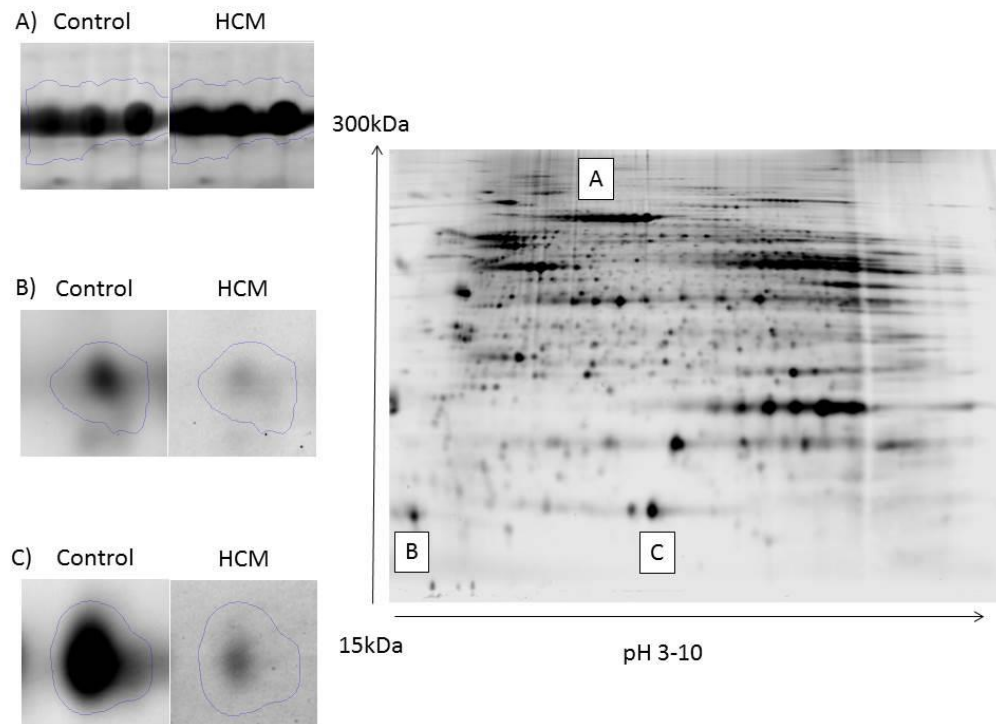


Figure 4.3 2D-DIGE map of the pH 3-10 region of human myocardium

The upper image shown is from the Cy2 signal. Spots labelled are examples of those shown to be significantly differentially expressed in HCM compared with controls. The lower image shows the interacting wavelengths.

Spot No.	Fold	HCM	UniProt ID	Accession number	Protein	pH	MW (Da)
4226	3.2	↓	ANXA5_HUMAN	P08758	Annexin A5	4.7	35.9
1251	2.8	↓	ODO2_HUMAN	P36957	2-oxoglutarate dehydrogenase complex component E2	9.3	48.7
2206	2.7	↓	FABPH_HUMAN	P05413	Fatty acid-binding protein, heart	6.4	14.8
2032	2.3	↓	HSPB6_HUMAN	O14558	Heat shock protein beta-6	5.9	17.1
4230	2.1	↓	LDHB_HUMAN	P07195	L-lactate dehydrogenase B chain	5.6	36.6
4209	2	↑	ACTN2_HUMAN	P35609	Alpha-actinin-2	5.2	103.8
			ACTN1_HUMAN	P12814-2	Isoform 2 of Alpha-actinin-1	5.2	102.6
			ACTN4_HUMAN	O43707	Alpha-actinin-4	5.1	104.8
4231	2	↓	A1AT_HUMAN	P01009	Alpha-1-antitrypsin	5.2	46.7
			TBB4B_HUMAN	P68371	Tubulin beta-4B chain	4.6	49.8
1372	2	↓	MYL3_HUMAN	P08590	Myosin light chain 3	4.8	21.9
1358	2	↓	KCRM_HUMAN	P06732	Creatine kinase M-type	6.8	43.1
			ACADM_HUMAN	P11310	Medium-chain specific acyl-CoA dehydrogenase	8.4	46.6
1614	2	↓	MDHM_HUMAN	P40926	Malate dehydrogenase, mitochondrial	8.8	35.5
4228	2	↓	LDHB_HUMAN	P07195	L-lactate dehydrogenase B chain	5.6	36.6
1879	2	↓	HSPB1_HUMAN	P04792	Heat shock protein beta-1	6.0	22.8

Table 4.2 Protein spots \geq 2-fold difference between HCM and Control

Increased (\uparrow) and decreased (\downarrow) expression of proteins in HCM compared with control with pH and molecular weight (MW) of each protein shown. If more than one protein per gel spot was identified all are listed. The spot number (no.) was assigned by Progenesis SameSpots V4.1 software and is included for reference.

Method B: SDS-PAGE separation

Using a top down approach, 150 µg of protein from each sample group was loaded onto an SDS-PAGE mini-gel and run alongside 12-225 kDa rainbow molecular weight markers (GE Healthcare, Amersham, RPN 600E). A representative gel is shown in Figure 4.4. The gel was cut into 10 bands and proteins were digested with trypsin prior to LC-MS/MS identification. Yeast enolase was added to enable quantitative comparison between groups. 595 proteins were identified across all 3 groups with 242 proteins found to be present in all three samples (Figure 4.5). Yeast enolase was added to enable quantitative comparison between groups. Differential expression analysis was performed using Protein Lynx genesis LC-MS.

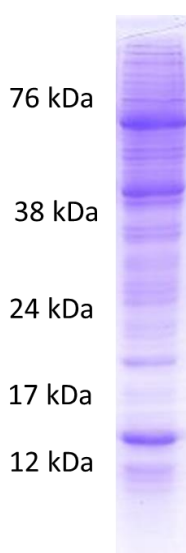


Figure 4.4 Coomassie blue stained 1D SDS-PAGE of heart homogenate.

Method C: Peptide fractionation

Using a bottom up approach, 500 µg of protein from each sample group was freeze dried and reconstituted in 20 µl of 100 mM Tris (1% ASB-14, pH 7.8) and digested with trypsin. Peptides were then fractionated offline using a acetonitrile gradient (Section 3.2.4). Yeast enolase was added to enable quantitative comparison between groups. Peptides were identified using LC-MS as described previously. 598 proteins were identified across all 3 groups. There was some degree of overlap (Figure 4.5) with 205 proteins being identified in all three samples.

4.2.3 Results from different protein separation methods

Fractionation of complex samples at the cellular, subcellular, protein, or peptide level is an indispensable strategy to improve the sensitivity in mass spectrometry-based proteomic profiling. All the techniques used provide complementary protein identification but the “top down” 1D-SDS-PAGE and “bottom up” fractionation had the highest number of “hits” or protein identifications. For subsequent experimental work, non-gel based methods were used to avoid the need for multiple parallel gels and enable analysis of individual rather than pooled patient samples. Table 4.3 summarises the differences between the methods.

	Method A 2D DIGE	Method B 1D SDS separation	Method C Peptide fractionation
Amount protein	50 µg	150 µg	500 µg
Protein spots/IDs	103	595	598
Common IDs	103	242	205
Progenesis			
➤ Total	103	261	320
➤ Fold > 1.5	86	80	54
➤ Fold > 2	27	38	11

Table 4.3 Comparison of proteome separation methods

Although the number of identifications was similar between Method 2 (595 proteins) and Method 3 (598 proteins), there were some important differences in the overlap between the three pooled sample groups. Using the “top down” approach where fractionation occurs at protein level determined by molecular weight, 242 proteins were identified as being common to all groups, compared with the “bottom up” approach where there were 205 proteins common to all 3 groups (Figure 4.5). Comparing across the methods, there were a total of 902 protein identifications of which 52 proteins were only identified through the “top-down” approach and 22 proteins only identified using the “bottom-up” approach.

The advantage of fractionation at protein level is illustrated in Figure 4.6 which considers the 20 most abundant proteins in the sample based on an estimate of absolute quantity using an internal standard of yeast enolase (Section 3.2.4.4). The

top panel illustrates that many of these highly abundant proteins (e.g. haemoglobin and fatty acid binding protein heart) can be depleted from the sample by excising the bands in which they are present. This gel-based approach is particularly useful to examine specific proteins of interest by excising only the band which they localise to. The presence of albumin in all bands implies that there is contamination across lanes, likely due to its very high abundance. The lower panel shows that fractionation at peptide level results in these highly abundant proteins being identified in almost every fraction. The advantage of this is reproducibility of quantitative data, but the disadvantage in these abundant peptides might suppress the less abundant peptides.

When considering the best approach to analyse a larger number of samples, these and other factors were taken into account. It was important to identify differences accurately in a timely manner. The challenge of running many parallel gels makes it unsuitable for analysis of a large number of samples. Given the number of protein identifications is similar between the top down and bottom up approach a gel-free method was selected. Label free quantitative proteomic methodologies have been used in our laboratory previously for the identification of novel biomarkers from a number of biological fluids (Bennett et al 2010, Heywood et al 2012). The methodology has now been developed for the identification of biomarkers from myocardial tissue.

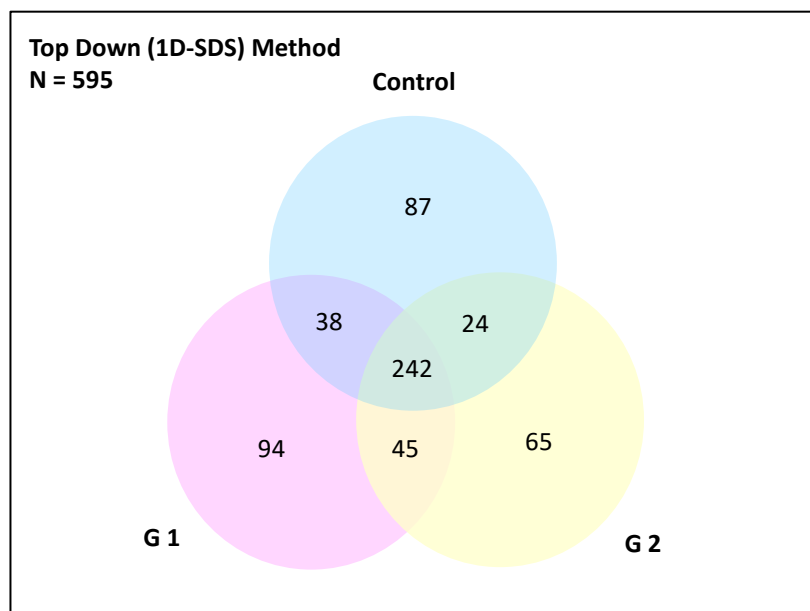
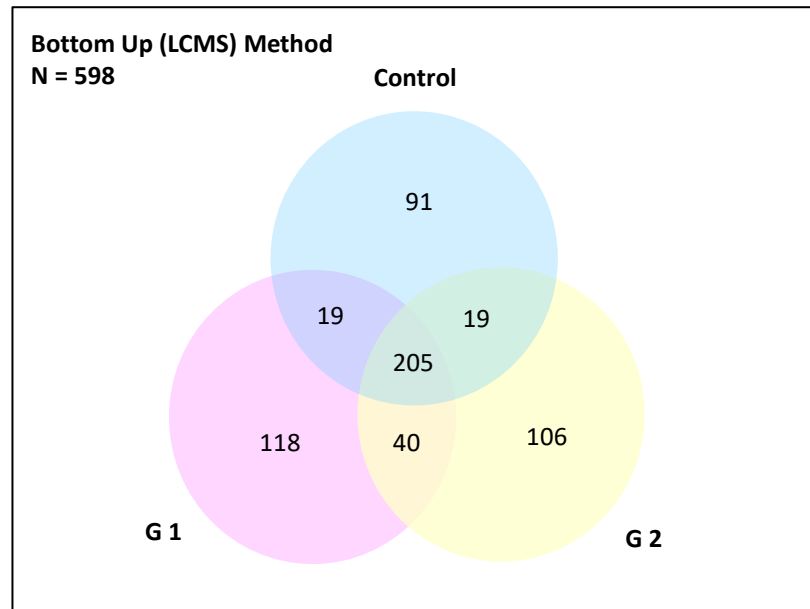


Figure 4.5 Venn diagram representing overlap in protein identifications.
Venn diagrams illustrating overlap in protein identifications between the 3 sample groups, according to the analytical method used.

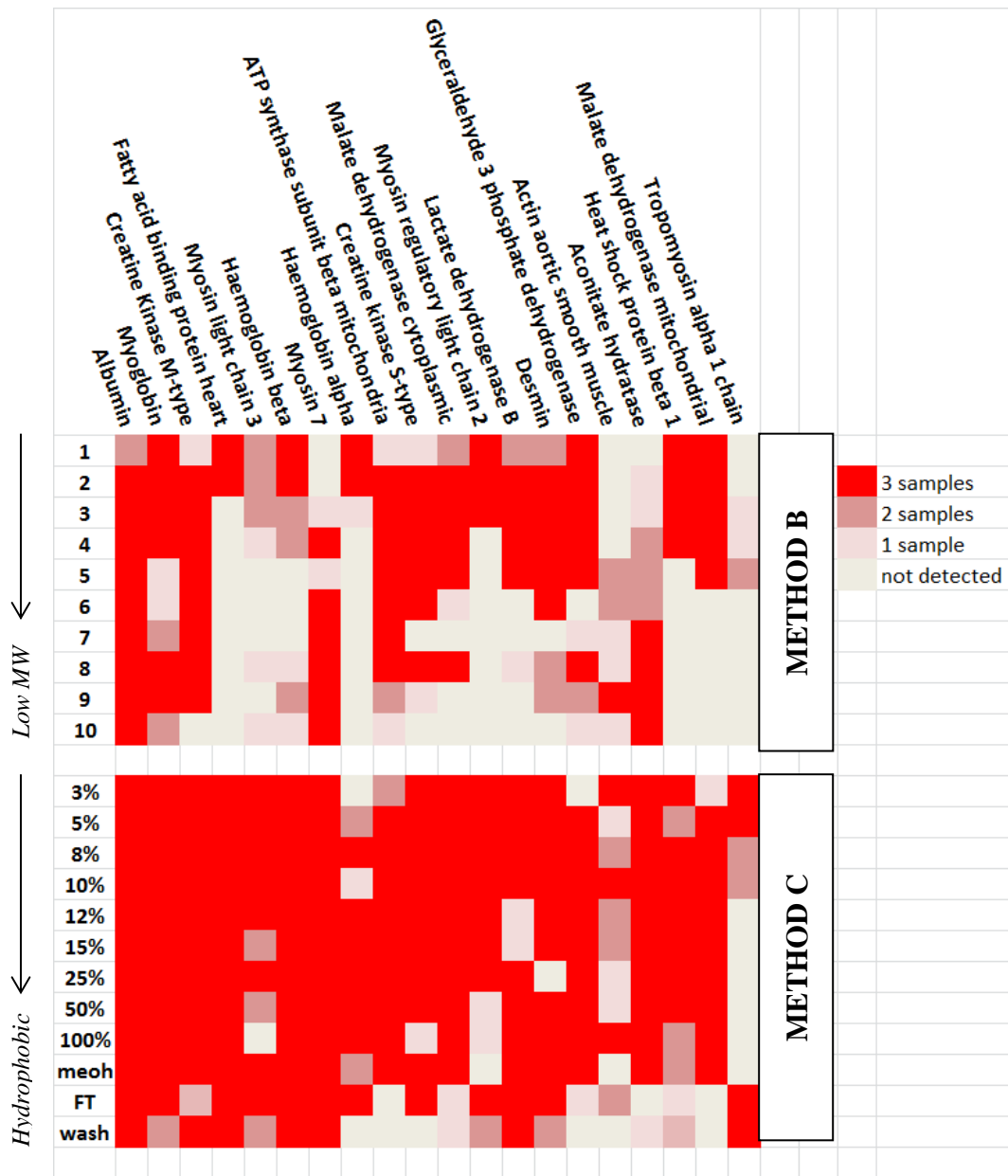


Figure 4.6 The top 20 most abundant proteins in heart

The top 20 most abundant proteins (horizontal axis) in myocardium are displayed according to their presence in each fraction (vertical axis). Shading is determined by whether each protein was identified in 1, 2 or 3 (i.e. all) samples.

Top panel: Fractions 1 to 10 represent high to low molecular weight (MW) gel bands.

Bottom panel: Fractions represent increasingly hydrophobic peptides by applying an offline acetonitrile gradient where % ACN is shown.

4.3 Exploration of potential pre-analytical variation

There is no single standard method for preparing protein samples for analysis by mass spectrometry. Sample preparation is time-consuming but a critical part of any proteomics experiment because the quality and reproducibility of sample extraction can impact results significantly. It can be particularly challenging to collect human tissue samples for scientific purposes. Unlike animal studies, the clinical environment cannot be so readily controlled. For example, time delays may be encountered in the surgical setting due to technical difficulties with removing tissue. The time from cardioplegia-to-bypass and bypass-to-tissue excision, will inevitably vary between patients and operators. Standard laboratory buffers may not be readily available, especially if surgery and tissue excision is performed in the emergency setting. In the post clinical setting, storage conditions of tissue may vary between laboratories and freeze-thaw is unavoidable for multi-stage experiments. There is minimal published literature in this area, so during the method development phase of this study, several aspects related to sample handling were investigated. The aim was to determine which storage conditions influenced protein expression and quantification.

4.3.1 Optimisation of specimen collection

This preliminary work explored the efficiency of normal saline solution (0.9 % NaCl) and phosphate buffered saline (PBS) in the preservation of myocardial specimens for proteomic analysis after different incubation times. Commonly used in biological research, PBS has osmolarity and ion concentrations (1 litre contains 10 mM PO_4^{3-} , 137 mM NaCl, and 2.7 mM KCl adjusted to pH 7.4) that match those of the human body making it isotonic and non-toxic to cells. Normal saline solution (contains 0.9% w/v of NaCl, about 300 mOsm/L or 9 g per litre) has a slightly higher degree of osmolarity than blood; it is readily available in all hospitals, where it is commonly used intravenously for fluid replacement.

4.3.1.1 Methods

Ten grams of myocardial tissue were obtained from one patient during surgery. This was an extensive resection of muscle so provided a large amount of myocardium for experimental work. A portion of this surgical specimen was sub-divided using a

clean scalpel into two pieces and one placed in PBS, and the other in 0.9% NaCl solution. Each piece was further sub-divided into six smaller pieces. The specimens were washed thoroughly to minimise blood contamination then snap-frozen in liquid nitrogen at increasing time intervals (2, 4, 8, 16, 32 and 64 minutes) from excision. Samples were stored at minus 80°C prior to tryptic digestion and analysis using shotgun label free LC-MS. Quantitative analysis was performed using Progenesis LC-MS software (V3.0, Nonlinear) as described previously (Section 3.2.4.4).

4.3.1.2 Results of delayed freezing

The average normalised abundance of most proteins decreased with delayed freezing. For example, the structural protein desmin (P17661, *DES*) identified using 16 unique peptides, was 2.9-fold higher in the earliest specimen compared with the most delayed time-to-freeze specimen, and a downward trend observed across sampling intervals (ANOVA, $p=0.064$). The same pattern was observed for enzymes, e.g. glyceraldehyde-3-phosphate dehydrogenase (P04406, *GAPDH*) identified using 7 unique peptides, was 2.9-fold higher in the earliest specimen compared with the most delayed time-to-freeze specimen (ANOVA, $p=0.081$) (Figure 4.7). To reduce this variability, surgical specimens were subsequently frozen immediately after excision.

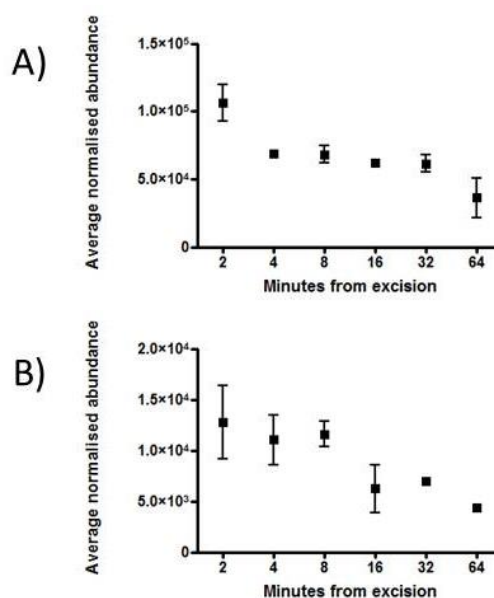


Figure 4.7 Protein abundance after delayed freezing.

Graphs show the average normalised abundance for A) desmin, B) glyceraldehyde-3-phosphate dehydrogenase. Individual values for each paired sample are shown with standard error of the mean displayed.

4.3.1.3 Results of specimen washing buffer

There was no statistically significant difference with the two different buffer collection strategies employed. The PBS washing resulted in the highest number of protein identifications. Although biological replicates were not studied, this was sufficient to support the inclusion of previously banked specimens, collected in both PBS and 0.9% NaCl, for use in later experiments.

4.3.2 Optimal specimen storage

Tissue samples are usually stored at -80 °C but the effects of storage at -20 °C on sample degradation are not known. The effects of sub-optimal sample preservation were examined using two duplicate samples (one disease and one control) that had been stored at -20 °C for 6 months. The loss of protein and range of proteins was calculated.

4.3.2.1 Methods

One specimen was divided and a small portion kept at -20°C for six months, then compared directly with the portion that had been stored at -80°C (standard protocol). The purpose was to explore the effect on protein degradation and subsequent quantification, should this happen inadvertently or appropriate storage facilities were not available.

4.3.2.2 Results

Using a fractionated LC-MS approach, a total of 411 proteins were identified in the -80 °C specimen and 317 proteins in the -20 °C specimen with 222 proteins identified in both samples (Figure 4.8). Metabolic proteins accounted for >50% of the proteins present in both samples, and there was no significant enrichment of a particular protein class in either group.

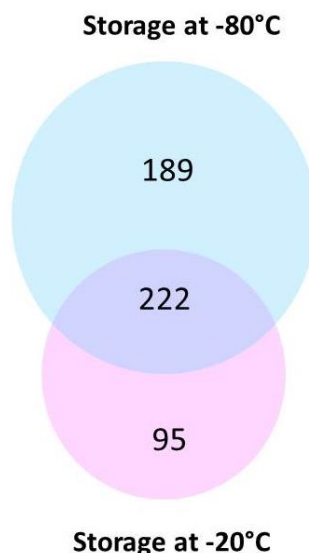


Figure 4.8 Effect of sub-optimal sample storage on number of proteins identified

4.3.3 Conclusion

A 77 % reduction in absolute number and approximate 3-fold change in relative abundance of proteins occurred with delayed freezing and sub-optimal storage conditions. All human myocardial samples used for quantitative proteomic experiments, should be collected bearing in mind, this degree of pre-analytical variability. Thus, a strict protocol was implemented for prospective collection of cardiomyopathy specimens in our unit.

4.4 Summary

Having compared several methods for protein and peptide separation, we decided to pursue a chromatography based separation approach (label-free LC-MS^E) for the subsequent discovery study. This was mainly due to our particular goal to explore disease mechanisms. By identifying as many different proteins as possible we anticipated we would be more likely to link these to pathways and ultimately understand their dysregulation at a network level. Careful attention was paid to specimen collection, storage and wet laboratory work to limit pre-analytical differences between patient samples.

5 The Proteome of Human HCM

Label-free liquid chromatography QTOF mass spectrometry (MS^E) was used to compare protein profiles in individual myocardial samples from eleven patients with hypertrophic cardiomyopathy (HCM) and six healthy donor hearts not used for cardiac transplantation. The samples from patients with HCM had a causative mutation in either the myosin heavy chain, *MYH7* (n=5) or myosin binding protein C, *MYBPC3* (n=6) gene. These were the same samples used in Chapter 4, but instead of pooling them into three groups, they were analysed individually as biological replicates. An in-depth examination of the cardiac proteome was used to explore disease mechanisms and identify potential biomarkers.

5.1 Methods

Myocardial samples were obtained from the basal portion of the left interventricular septum and prepared according to methods described in Section 3.2.1. Five hundred micrograms of protein from each sample was freeze dried and reconstituted in 100mM Tris buffer pH 7.2 prior to tryptic digestion. Sample complexity was reduced using off line high pH/low pH fractionation. Twelve aliquots were obtained using an acetonitrile gradient to separate hydrophilic and hydrophobic peptides (Section 3.2.4.2). Yeast enolase was added as an internal standard prior to analysis by UPLC-QToF-MS to enable quantitative comparison between groups. Samples were run on the Q-ToF mass spectrometer over several days and grouped according to fraction. Calibration and control samples were included between runs to ensure consistent sensitivity and reproducibility. Mass spectrometry data was assessed on the basis of ion intensity of peptide ions with coincident LC retention time and m/z values. These data were displayed as an intensity map by Progenesis LC-MS software (Nonlinear Dynamics, Newcastle upon Tyne, UK) as described in Section 3.2.4.4.

5.2 Results

5.2.1 Protein identification

Applying the previously described filtration criteria (Section 3.2.4.4) a total of 1586 proteins were identified across all heart samples using ProteinLynx Global

SERVER™ (Version 2.4, Waters UK). Of the 1147 proteins identified in the HCM samples, 797 proteins were identified in the MYH7 sub-group (Group A) and 782 identified in the MYBPC3 group (Group B) with an overlap of 432 proteins common to both disease sub-groups (Figure 5.1) The average number of unique proteins identified per patient sample was 394 ± 56 (mean \pm SD) with the highest numbers identified in the 8% fraction (Figure 5.2).

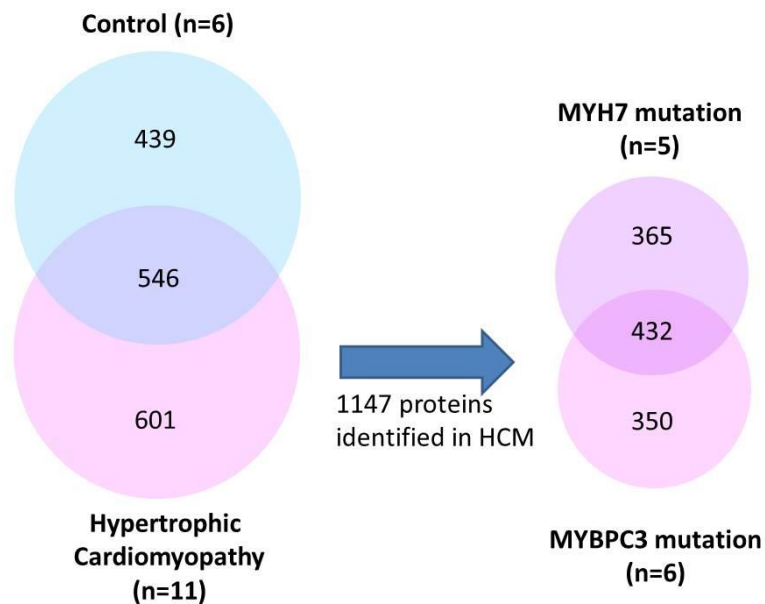


Figure 5.1 Summary of number of proteins identified in heart samples

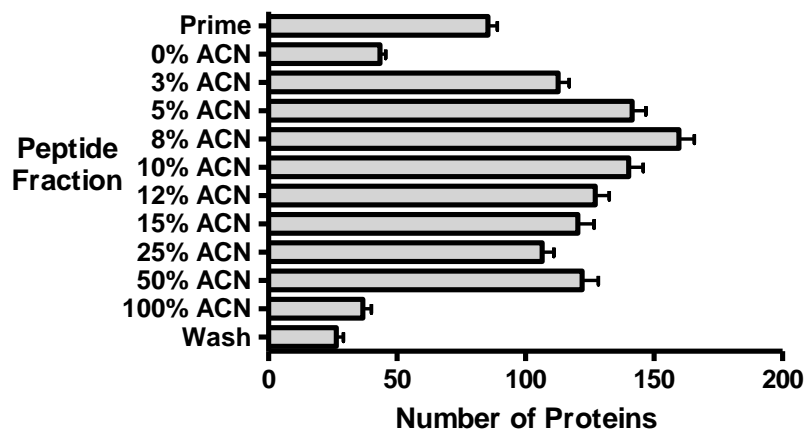


Figure 5.2 Bar chart illustrating proteins identified per fraction
The average number of proteins identified in each fraction is shown (mean and SEM).

5.2.1.1 Proteins common to all samples

Approximately one third (n=546) proteins were identified in at least one sample from the HCM and control groups. Only 137 proteins were found in all 17 samples. The inclusion of principally plasma proteins, such as albumin and α -1 antitrypsin, in this group demonstrates the difficulty in separating myocardium from the plasma. There were 216 proteins common to all control samples and 145 proteins common to all HCM samples. The lower number of common proteins in the HCM samples may reflect either different biochemical pathways being induced or disease heterogeneity.

5.2.1.2 Proteins unique to health and disease

Certain proteins appeared to be unique to either the HCM or control group. The following proteins were absent in the HCM samples but present in four or more (i.e. >50%) of the controls samples: Long chain fatty acid CoA ligase 1 (P33121), Annexin A1 (P04083), Amine oxidase flavin containing B (P27338), Sarcoplasmic ER calcium ATPase 1 (O14983), Mitochondrial inner membrane protein (Q16891) and NADH dehydrogenase ubiquinone 1 beta sub complex (O96000). Only two proteins were absent in the control samples but present in six or more (i.e. >50%) of the HCM samples. These were Gelsolin (P06396) and Carbonic anhydrase 3 (P07451). The presence or absence of a particular protein however is not absolute and it reflects the stringent criteria used to filter the mass spectral data to ensure accurate protein identification and prevent false positives (Section 3.2.4.4). Detailed expression profiling was therefore performed using the commercial software, Progenesis LC-MS (version 3.0, Nonlinear Dynamics, Waters Corp.), based on ion abundance comparisons within the combined sample fractions.

5.2.1.3 Estimate of protein concentration

The spiked internal standard of yeast enolase at a known amount of 100 fmol per 500 μ g tissue, enabled an estimate of protein concentration within each sample. We describe for the first time a relative quantitative protein library of human hypertrophic cardiomyopathy (Figure 5.3). The most abundant proteins were albumin and myoglobin which respectively made up 8% and 9% of the total cardiac proteome.

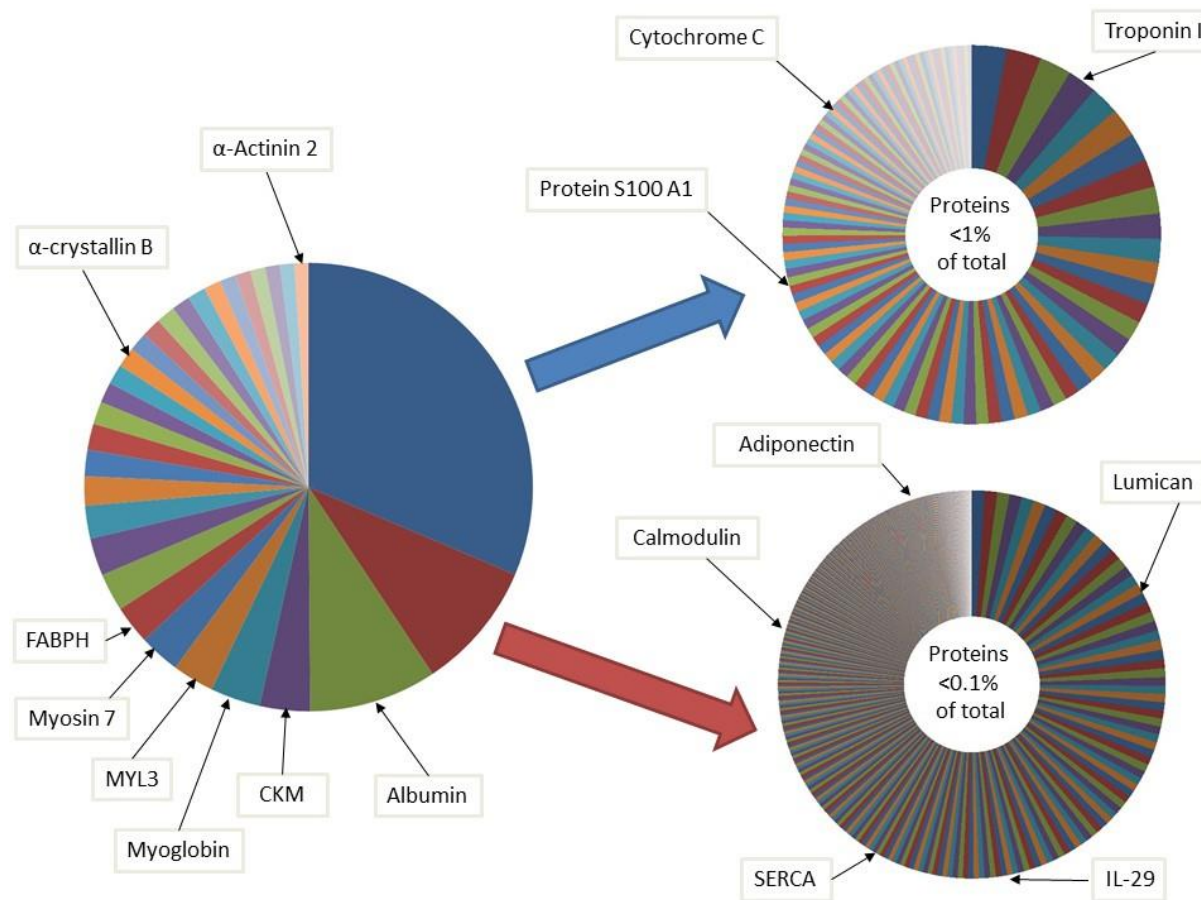


Figure 5.3 The proteome of human HCM.

Schematic representation of a typical proteome of left ventricular myocardium from patients with HCM. Proteins are represented as % fmol of protein of the total protein detected. The expanded charts show examples of the lower abundance proteins which constitute < 1% fmol (upper chart) and < 0.1% fmol (lower chart) of myocardial protein of the total protein detected. Examples of proteins (either known to be important in HCM or discussed later in this thesis) are indicated with arrows.

5.2.2 Gene ontology

The 1586 proteins identified in this study were categorised using the PANTHER bioinformatics tool (<http://www.pantherdb.org>) (Thomas, Kejariwal et al. 2006) and Database for Annotation, Visualization, and Integrated Discovery (DAVID; <http://www.david.niaid.nih.gov>) (Huang da, Sherman et al. 2009). Gene ontology (GO) categories “protein class”, “molecular function”, “cellular component” and “biological process” were analysed, enabling common features between proteins of interest to be grouped and identified.

5.2.2.1 Protein Class

Cytoskeletal proteins were the largest annotated protein class (n=177) however many enzymes were identified (Figure 5.4) including transferases (n=142), oxidoreductases (n=135) and hydrolases (n=121). Few cell junction (n=21) and extracellular matrix (n=9) proteins were identified. An important number (n=90) of annotations were for transcription factors which are essential to elucidating gene regulatory mechanisms.

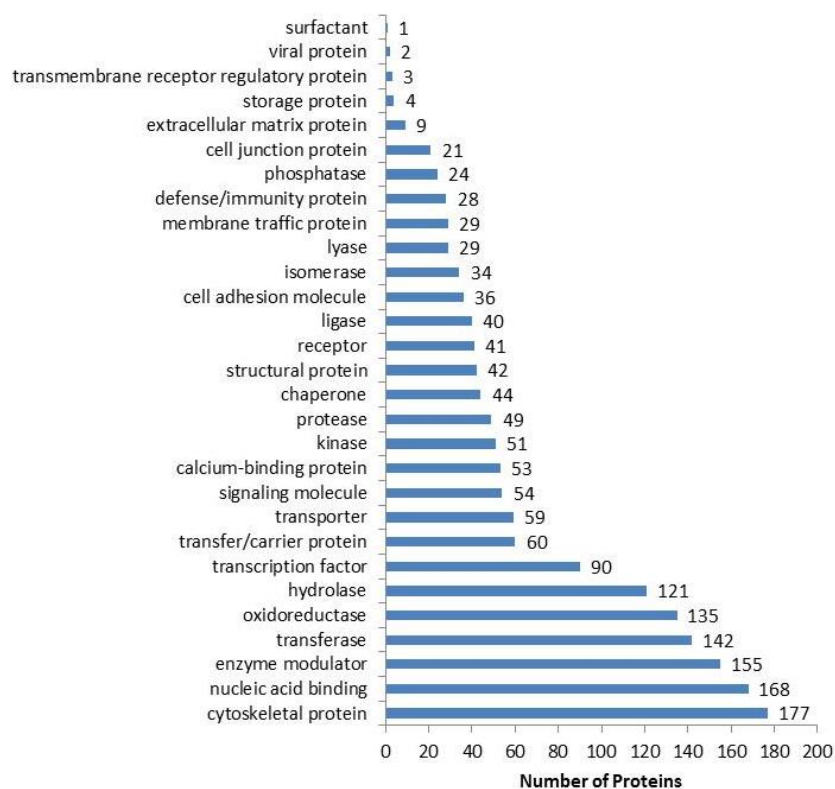


Figure 5.4 Proteins identified according to GO protein class

5.2.2.2 Molecular function

The commonest molecular functions (Figure 5.5) were catalytic activity (n=575) and binding (n=461). Sub-groups within binding, included protein binding (n=255) and nucleic acid binding (n=204).

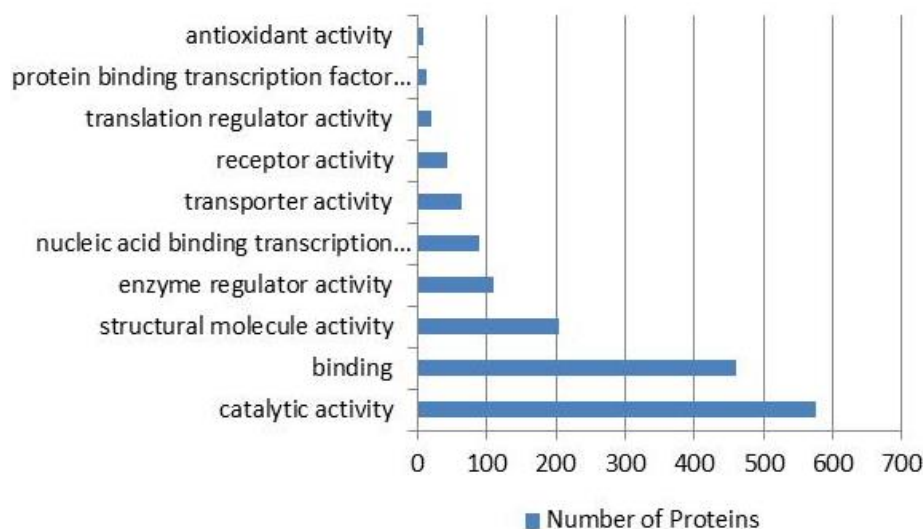


Figure 5.5 Proteins identified according to GO molecular function

5.2.2.3 Sub-cellular compartment

Proteins were commonly annotated to a particular cell part (n=349) of which most were intracellular cytoplasmic proteins (n=321). Only a few plasma membrane proteins (n=41) were identified which is likely to reflect the methods employed for sample preparation. In order to enrich for lipids the myocardial samples could be precipitated in an organic solvent such as acetone or chloroform methanol. The second largest annotation was to particular cell organelles (n=248) of which most were cytoskeletal proteins (n=147). Few extracellular proteins (n=43) were identified compared with intracellular proteins (Figure 5.6).

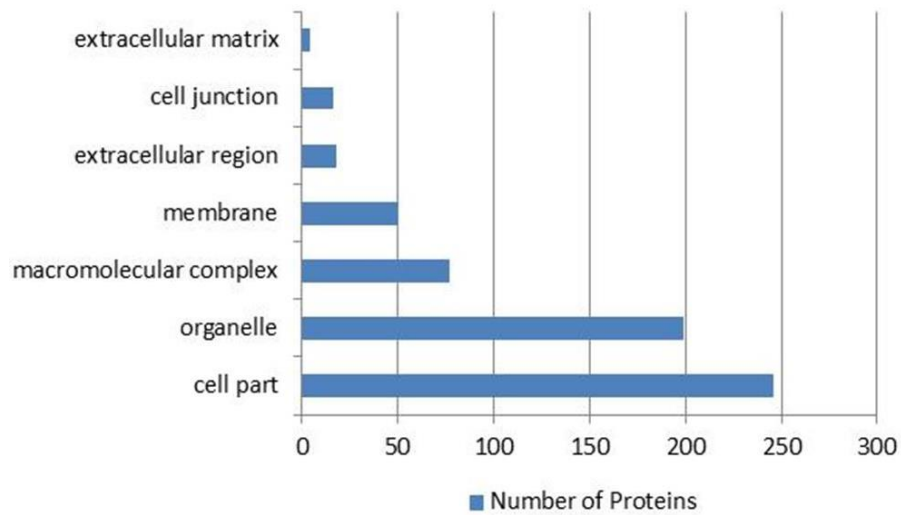


Figure 5.6 Proteins identified according to GO sub-cellular compartment

5.2.2.4 Biological process

Most proteins have several biological functions (Figure 5.7). Many proteins identified were associated with metabolic (n=809) and cellular processes (n=516) such as cell-cell signalling (n=180) and the cell cycle (n=125). The high number of metabolic proteins was expected given the density of mitochondria within the heart.

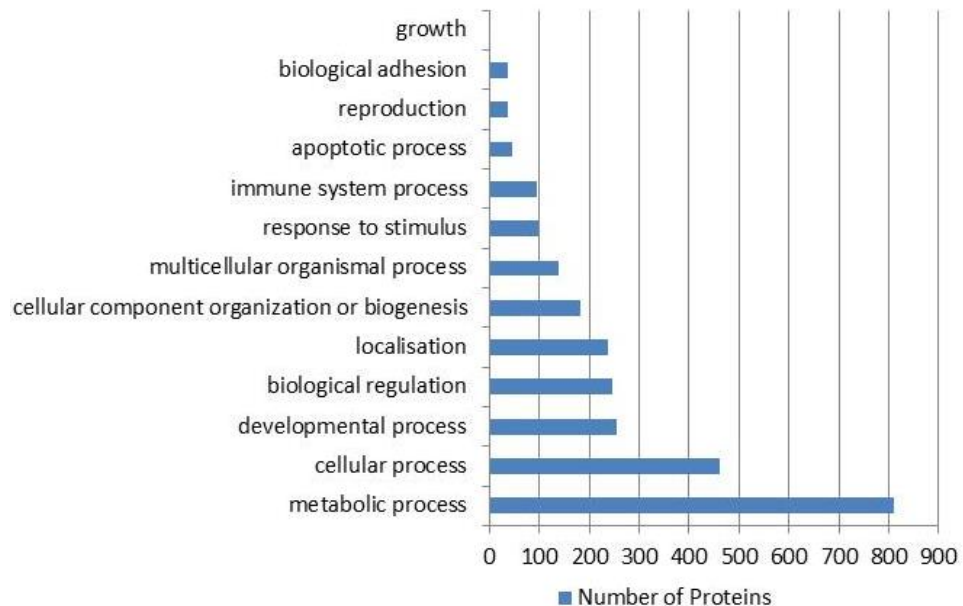


Figure 5.7 Proteins identified according to GO biological function

5.2.3 Differential protein expression

5.2.3.1 Absolute quantification

Absolute quantities of proteins were calculated across all fractions for each sample using ProteinLynx Global SERVER™ (Version 2.4, Waters UK). Average concentration of each protein per sample was calculated and the absolute fold change between the HCM and control groups determined (Appendix 1). Of the total number of proteins (n=1586) identified across all samples, 439 proteins were only identified in control samples and 601 proteins only in HCM samples. Of the 546 proteins identified in both groups, there were 80 proteins > 2-fold higher concentration in HCM group and there were 185 proteins > 2-fold higher concentration in the control group. There was a < 2-fold difference in the remaining 281 proteins.

5.2.3.2 Relative quantification

Extracted peptide intensities were evaluated with Progenesis LC-MS (Version 3.0, Nonlinear Dynamics, Waters Corporation) to obtain label-free relative quantification from the raw data files. The software created a single aggregate run combining all experimental fractions based on LC retention time of each analysis to contain all MS data with distinguished peptide ions. Feature outline maps were generated for detection and quantification of peptide ions from individual analyses using default program parameters. A total of 595 proteins were suitable for comparison between the HCM and control groups. Principal component analysis showed good separation between the two groups (Figure 5.8).

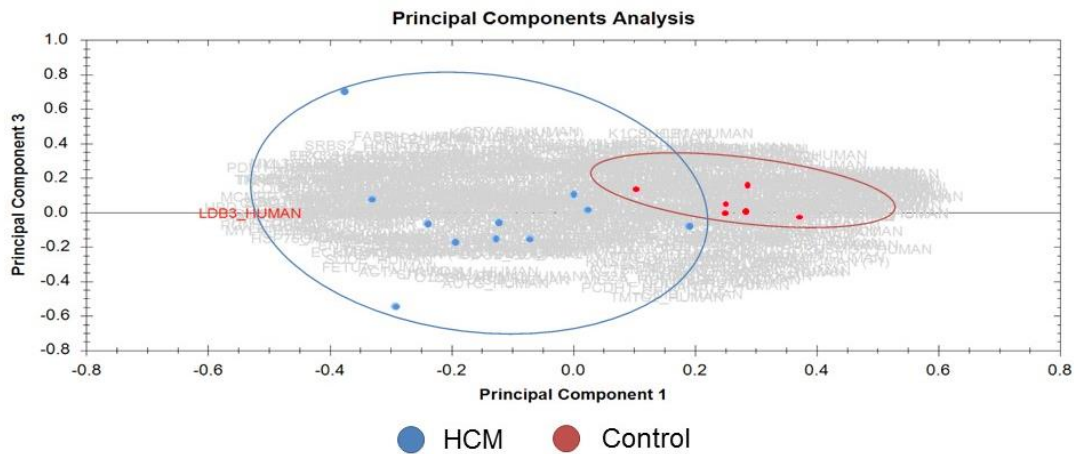


Figure 5.8 Principle component analysis of myocardial protein expression.
Eleven HCM samples are shown by blue circles and six control samples by red circles.

Compared to controls, 158 proteins were significantly ($p < 0.05$) differentially expressed in HCM samples; 62 proteins were upregulated and 96 down regulated (Appendix 1).

Twenty-nine proteins showed more than two-fold difference in expression using Progenesis LC-MS. The difference was statistically significant ($p < 0.05$) by ANOVA for twenty-two proteins, of which seven were up-regulated (Table 5.1) and fifteen were down-regulated (Table 5.2) in the HCM samples.

The relative quantification of protein expression from Progenesis was compared with the absolute quantitative results obtained from Protein Lynx (Figure 5.9). All but three proteins (IGHG3, DJC25 and PTMS) were confirmed to be dysregulated by both analysis methods.

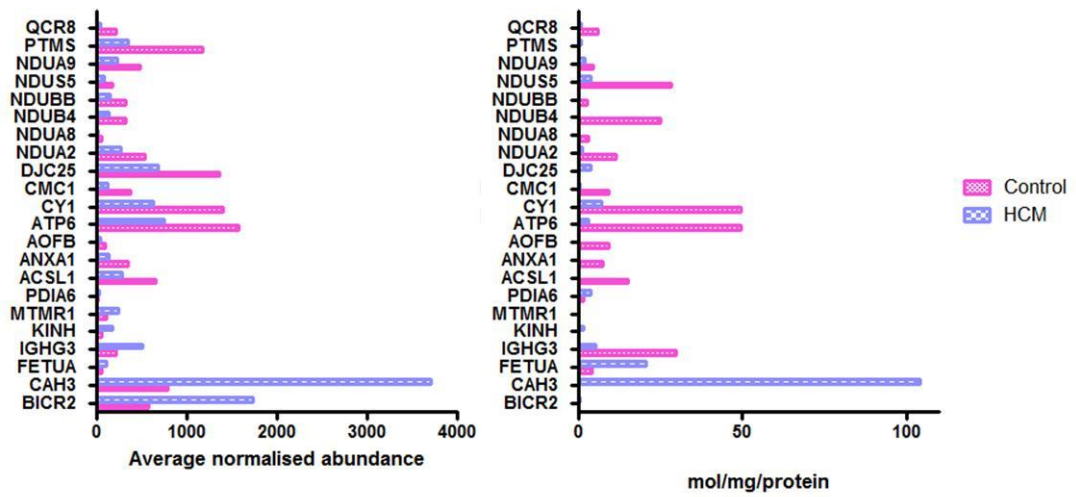


Figure 5.9 Proteins detected as showing >2-fold alteration in heart tissue

Proteins detected as showing a greater than 2-fold alteration in heart tissue from HCM patients (n=11) and control (n=6) analysed with:

a) Progenesis LC-MS (Left panel) where y axis shows the Unit Prot accession ID and x axis, the average normalised abundance

b) ProteinLynx (Right panel) where y axis shows Unit Prot accession ID and x axis, concentration in mol/mg/protein.

Accession	Gene	Peptides	Score	Anova p-value	Fold	Description	Average normalised abundance	
							Control	HCM
BICR2_HUMAN	CCDC64B	1	6.3	0.04	3.05	Bicaudal D related protein 2	569.44	1739.1
CAH3_HUMAN	CA3	7	45.61	1.52E-03	4.69	Carbonic anhydrase 3	789.9	3704.45
FETUA_HUMAN	AHSG	1	6.81	0.01	2.16	Alpha 2 HS glycoprotein	50.29	108.7
IGHG3_HUMAN	IGHG3	22 (1)	102.81	3.70E-03	2.4	Ig gamma 3 chain C region	211.73	507.63
KINH_HUMAN	KIF5B	1	6.43	2.92E-03	3.16	Kinesin 1 heavy chain	53.34	168.76
MTMR1_HUMAN	MTMR1	1	6.18	0.01	2.33	Myotubularin related protein 1	100.4	233.98
PDIA6_HUMAN	PDIA6	2 (1)	12.38	0.04	2.25	Protein disulfide isomerase A6	13.95	31.36

Table 5.1 Proteins up-regulated > 2-fold in HCM

Accession	Gene	Peptides	Score	ANOVA p-value	Fold	Description	Average normalised abundance	
							Control	HCM
ACSL1_HUMAN	ACSL1	1	6.99	2.60E-05	2.31	Long chain fatty acid CoA ligase 1	647.68	279.83
ANXA1_HUMAN	ANXA1	3	18.48	9.71E-05	2.65	Annexin A1	341.42	128.92
AOFB_HUMAN	MAOB	3 (2)	13.11	2.61E-04	2.11	Amine oxidase flavin containing B	98.24	46.64
ATP6_HUMAN	MT ATP6	2	14.42	1.73E-04	2.11	ATP synthase subunit a	1576.31	746.67
CY1_HUMAN	CYC1	3 (2)	24.6	3.36E-04	2.22	Cytochrome c1 heme protein mitochondrial	1395.42	627.56
CMC1_HUMAN	SLC25A12	5 (4)	34.6	3.57E-04	3.25	Calcium binding mitochondrial carrier protein Aralar1	370.6	113.95
DJC25_HUMAN	DNAJC25	2	13.8	0.03	2.00	Dnal homolog subfamily C member 25	1360.67	680.91
NDUA2_HUMAN	NDUFA2	3	23.78	0.01	2.01	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 2	526.81	262.25
NDUA8_HUMAN	NDUFA8	1	7.13	6.17E-03	4.82	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 8	51.38	10.66
NDUB4_HUMAN	NDUFB4	2	21.54	1.15E-05	2.27	NADH dehydrogenase ubiquinone 1 beta subcomplex subunit 4	316.18	139.51
NDUBB_HUMAN	NDUFB	2 (1)	6.3	9.83E-05	2.1	NADH dehydrogenase ubiquinone 1 beta subcomplex subunit 11 mitochondrial	319.19	152.06
NDUS5_HUMAN	NDUFS5	3 (2)	15.79	7.39E-04	2.24	NADH dehydrogenase ubiquinone iron sulfur protein 5	172.51	77.09
NDUA9_HUMAN	NDUFA	4	33.36	6.39E-03	2.11	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 9 mitochondrial	481.54	227.9
PTMS_HUMAN	PTMS	4 (3)	5.98	0.04	3.34	Parathyrosin	1169.68	350.57
QCR8_HUMAN	UQCRQ	2 (1)	14.82	6.18E-05	4.91	Cytochrome b c1 complex subunit 8	215.38	43.83

Table 5.2 Proteins down-regulated > 2-fold in HCM

5.2.4 Altered expression of protein clusters in HCM

Functional annotation clustering using PANTHER (<http://www.pantherdb.org/>) and DAVID (<http://david.abcc.ncifcrf.gov>) tools showed many of the deregulated proteins were associated with contractile and metabolic function of the heart (Figure 5.10). Proteins involved in calcium regulation, cell signalling, apoptosis and oxidative stress were also altered. The increased or decreased expression of a protein could not necessarily determine whether particular pathways were up- or downregulated because protein levels do not necessarily correlate with activity of the protein. Furthermore, a single protein may have multiple functional roles and the functions of many proteins remain unknown.

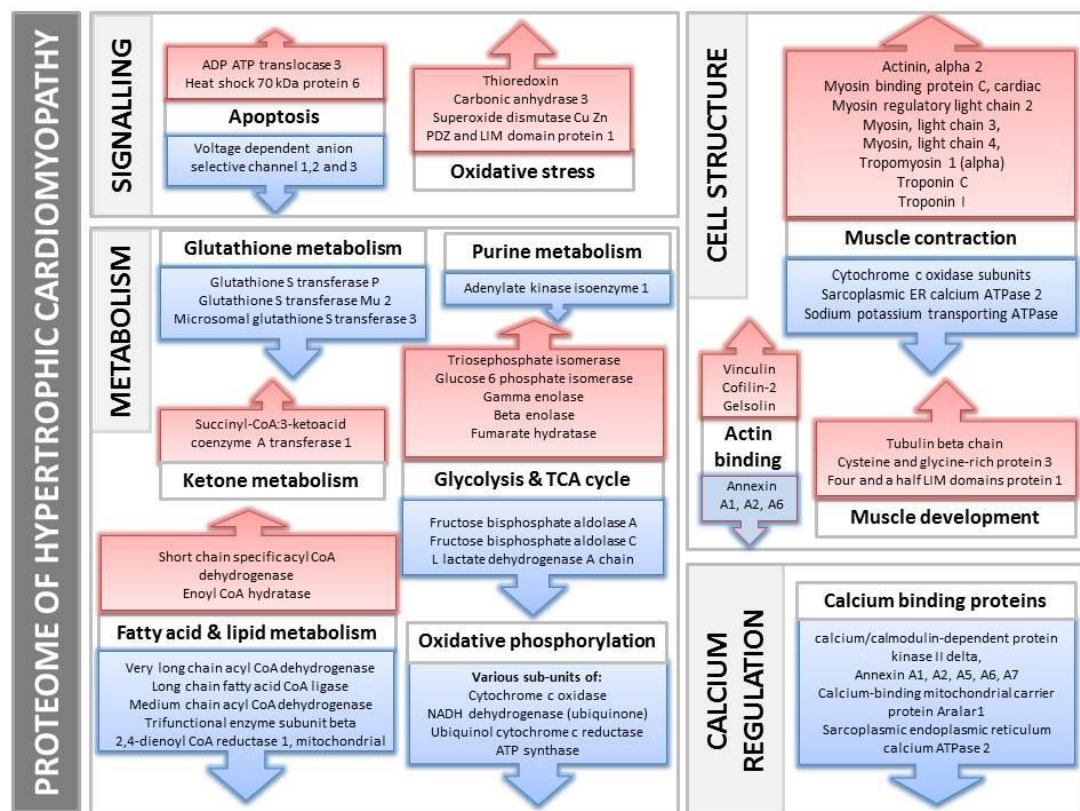


Figure 5.10 Significantly altered protein expression in HCM

Proteins that were significantly upregulated are shown in red boxes and those down regulated in blue boxes. Statistical significance was determined by $p < 0.05$, based on ion abundance ANOVA analysis.

5.2.4.1 Correlation with transcriptome studies

The alterations found in particular protein groups had some similarities to previously published transcriptomic studies in HCM (Lim, Roberts et al. 2001, Rajan, Williams et al. 2006). Markers of secondary hypertrophy e.g. skeletal muscle alpha-actin (*ACTS*, P68133), four and a half LIM domains protein 1 (*FHL1*, Q13642) and isoforms of myosin light chain were increased consistent with the findings of Lim et al (Lim, Roberts et al. 2001). Sarcoplasmic endoplasmic reticulum calcium ATPase 2 or SERCA2 (*ATP2A2*, P16615) was found to be significantly down-regulated, a pattern observed so consistently at mRNA level in failing HCM and DCM hearts that it has been used as a positive control marker in gene expression studies (Hwang, Allen et al. 2002). Proteins associated with energy metabolism, the extracellular matrix and inflammation also overlapped with gene expression studies, e.g. Yang et al showed increased gelsolin (*GELS*, P06396) expression in failing hearts (Yang, Moravec et al. 2000) and Hwang et al found lumican (*LUM*, P51884) was over-expressed in failing DCM hearts (Hwang, Allen et al. 2002).

However not all protein expression was consistent with gene expression. For example, Lim et al found the voltage-dependent ion channel, *VDAC-1* gene was over-expressed in HCM, however our proteomic study found reduced expression of *VDAC-1*, -2 and -3 in HCM compared with controls. These channels play a key role in regulating metabolic and energetic flux across the outer mitochondrial membrane. Several annexin proteins were also identified to be down-regulated, despite gene expression studies showing certain isoforms are upregulated in the failing heart (Hwang, Allen et al. 2002).

Gene expression is often interpreted in terms of protein levels, but when both are measured, the correlations are not very strong. There are many processes between transcription and translation and protein stability is an important factor (Vogel and Marcotte 2012). Furthermore, the presence of a protein does not correlate necessarily with its activity. This is particularly relevant to enzymes, where a functional assay is needed to determine its action within the cell or tissue. This is partly because individual proteins do not function in isolation, but are linked by non-covalent

protein-protein interactions to form protein complexes. These complexes are essential to perform many biological processes e.g. metabolism.

5.2.5 Pathway and Network Analysis

Interacting proteins are more likely to be involved in similar biological functions and processes and therefore are more likely to be co-expressed. In order to better understand the functional relevance of altered protein expression, the proteomic data needs to be analysed in an integrated manner and various platforms exist to do this. In the present study, pathway assignment was performed using the open access Reactome database (<http://www.reactome.org>) (D'Eustachio 2011) managed by the European Bioinformatics Institute (EBI). The analyses of protein-protein interaction and signalling networks was performed through the use of QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity). IPA is a commercially available, manually curated knowledge database generated from peer-reviewed scientific publications that enables discovery of highly represented functions and pathways from large, quantitative data sets.

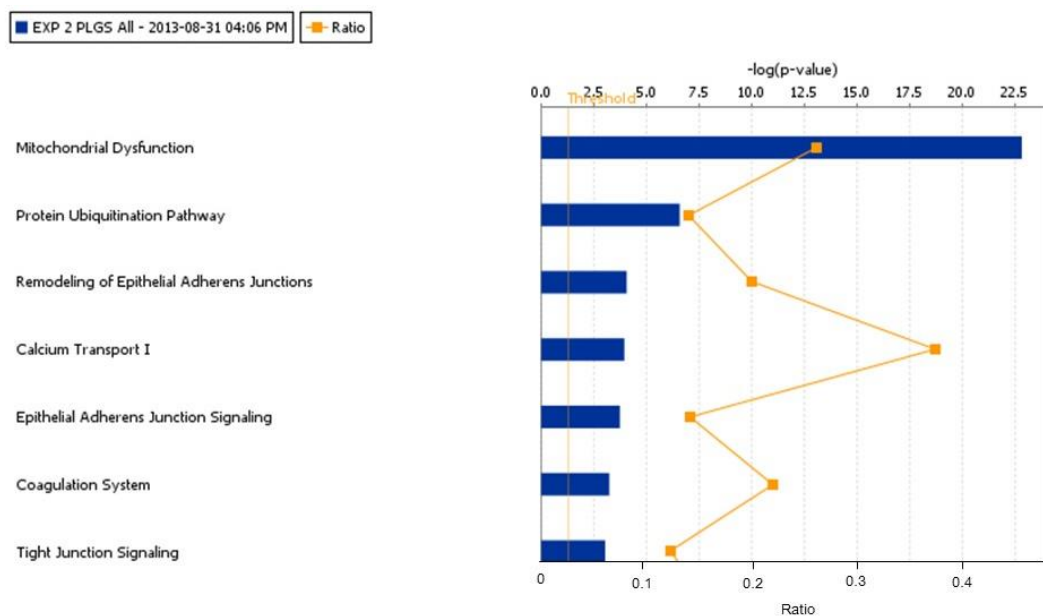


Figure 5.11 Significantly enriched canonical pathways identified by IPA.

The diagram shows significantly overrepresented canonical pathways. A multiple-testing corrected p -value was calculated using the Benjamini-Hochberg method to control the rate of false discoveries in statistical hypothesis testing. The ratio value represents the number of molecules in a given pathway that meet cut criteria (blue bar), divided by the total number of molecules that belong to the function (orange dot).

The most statistically significant enriched categories and networks identified by IPA were associated with mitochondrial dysfunction, protein ubiquitination, epithelial adherens junction signalling and calcium transport (Figure 5.11). Altered expression of individual proteins was visualised by overlaying results onto the relevant canonical pathway (Figure 5.12).

Although pathway analysis tools offer an insight into the underlying biology, there are inherent limitations in using a list of proteins (or genes) expressed differentially to conclude about disease mechanisms. Annotation is usually done at gene level, but multiple transcripts from the same gene may have related, distinct, or even opposing functions. In our dataset twelve proteins were not mapped to an IPA canonical pathway, e.g. calmodulin (CALM, P62158) an intermediate messenger protein which binds calcium and mediates crucial processes such as inflammation, metabolism, apoptosis and muscle contraction. The function of some proteins and links between many smaller pathways remain unknown. As further experimental evidence is generated, the annotation knowledge base will grow and pathway analysis will provide better understanding of the effects downstream of altered protein expression.

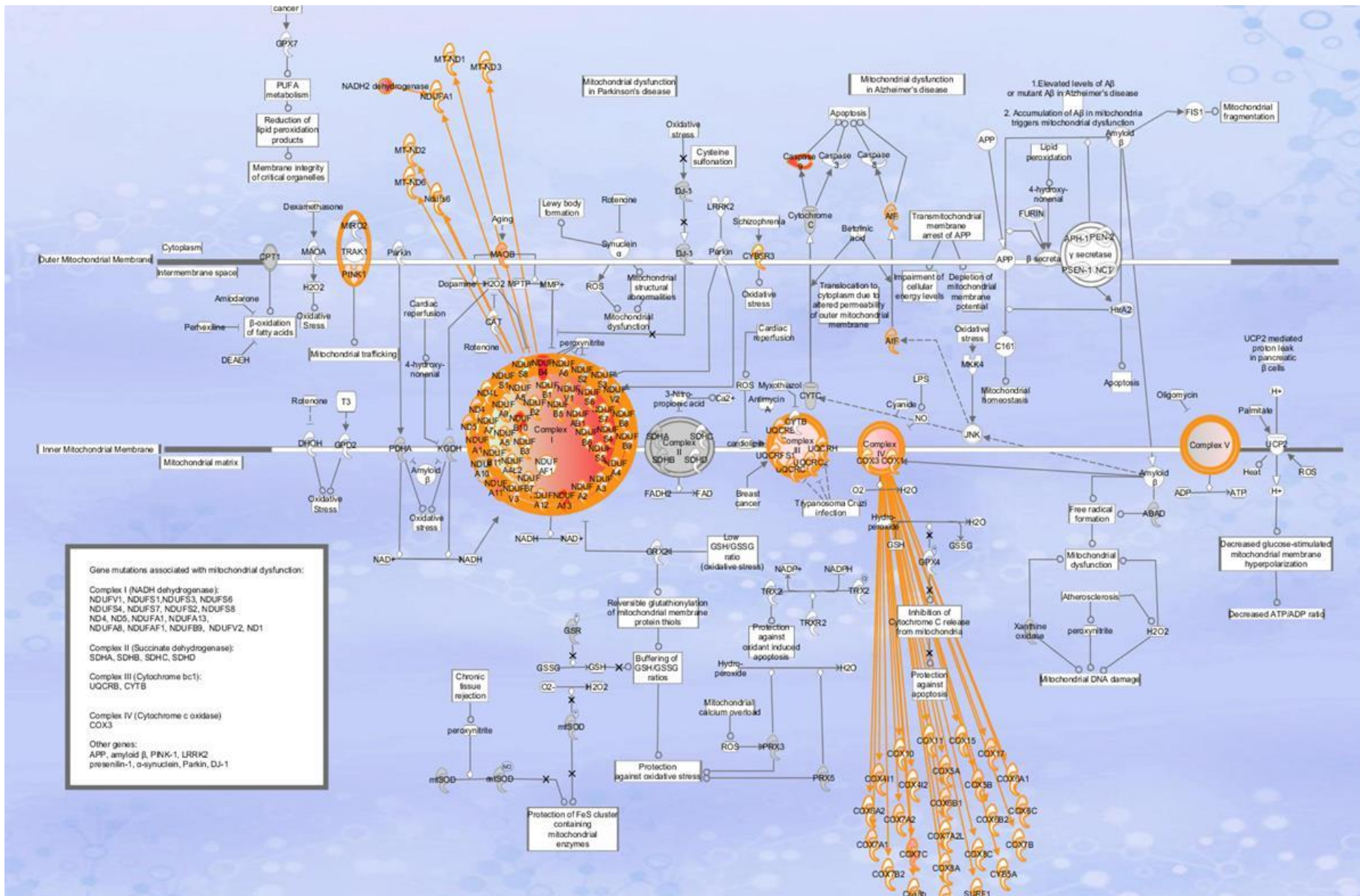


Figure 5.12 Proteins involved in canonical pathway "mitochondrial dysfunction"

Green and red colour indicates proteins significantly increased and decreased in expression, respectively. Grey indicates no significant change in protein expression.

5.3 Discussion

This is the first in-depth characterisation of the proteome of human hypertrophic cardiomyopathy. Several observed changes are compatible with the known pathophysiology of the disease, and others provide new insights into cellular processes that might be involved in cardiac dysfunction.

5.3.1 Evidence for fetal reprogramming

It is well recognised that pathological cardiac hypertrophy activates a group of genes normally expressed in the fetal heart, termed the fetal gene program (FGP). This is understood to be a protective mechanism that enables cells to survive in a relatively hypoxic environment. The FGP is a cardinal feature of the failing adult heart which reverts to a fetal pattern of energy substrate metabolism by down-regulating adult gene transcripts (Razeghi, Young et al. 2001). The proteome of hypertrophic cardiomyopathy was consistent with reactivation of the FGP with typical alterations in both metabolic and structural proteins:

5.3.1.1 Structural proteins

We identified increased levels of skeletal muscle actin, alpha (*ACTA*, P68133), four and a half LIM domains protein 1 (*FHL1*, Q13642), LIM domain binding protein 3, (*LDB3*, O75112) and myosin light chain 3 (*MYL3*, P08590) in HCM, consistent with a switch to the FGP. There was a reduced expression of the calcium cycling protein, SERCA2a (*ATP2A2*, P16615) a finding reported in most studies of hypertrophied or failing hearts (Arai, Matsui et al. 1994). Additionally, an increase in alpha-myosin (MYH6) expression was observed in HCM without significant difference in expression of beta-myosin (MYH7). Usually MYH6 is expressed abundantly in both cardiac atria and cardiac ventricles during embryonic development, but is predominantly in atria after birth, whereas the ventricles express the MYH7 isoform. The increased expression of MYH6 in HCM ventricles further supports reactivation of the FGP and is in keeping with other studies of cardiac hypertrophy (Gupta 2007). It was not possible to determine whether these changes were specific to HCM per se or secondary to cardiac hypertrophy.

5.3.1.2 Metabolic proteins

The normal myocardium depends on oxygen for high-energy phosphate production by oxidative phosphorylation. Adenosine triphosphate (ATP) is produced primarily by the metabolism of free fatty acids (FFAs) and carbohydrates, with FFAs accounting for approximately 70 % of ATP production in the normal adult heart. Importantly, FFAs require approximately 10% more oxygen than glucose in order to produce an equivalent amount of ATP (Stanley, Recchia et al. 2005). The fetal heart, on the other hand metabolises predominantly carbohydrate and the switch in energy substrate preference from carbohydrates to fatty acids is a hallmark of the transition from fetal to adult cardiac metabolism (Lopaschuk, Collins-Nakai et al. 1992).

Glycolytic pathway

In HCM hearts a significant upregulation of glycolytic proteins was observed with 1.3-fold increased expression of triose-phosphate isomerase (TPIS, P60174), gamma enolase (ENOG, P09104) and glucose 6 phosphate isomerase (G6PI, P06744). There was however 1.2-fold decreased expression of fructose bisphosphonate aldolase A (ALDOA, P04075) and C (ALDOC, P09972). There was also 1.3-fold decreased expression of L lactate dehydrogenase A chain (LDHA, P00338). It was difficult to ascertain whether glycolysis was overall upregulated or downregulated.

Beta-oxidation pathway

There was evidence of altered fatty acid metabolism in HCM. Enzymes responsible for metabolism of medium and long chain fatty acids were found to be down regulated. These included long chain fatty acid CoA ligase 1 (ACSL1, P3312; 2.3-fold), very long chain specific acyl CoA dehydrogenase (ACADVL, P49748; 1.2-fold) and medium chain specific acyl CoA dehydrogenase (ACADM, P11310; 1.2-fold). The mitochondrial trifunctional enzyme subunit beta (ECHB, P55084) was also downregulated 1.3-fold. This protein catalyzes the final step of beta-oxidation, in which 3-ketoacyl CoA is cleaved by the thiol group of another molecule of Coenzyme A. Overall these findings suggest beta-oxidation is down-regulated in HCM and confirms a switch to FGP.

5.3.2 Mitochondrial dysfunction and energy production

Mitochondria generate energy in the form of adenosine triphosphate (ATP) to support many intracellular processes including contraction of myofilaments. They are also involved in other tasks, such as signalling, cell differentiation, cell growth and death (McBride, Neuspiel et al. 2006). Mitochondrial dysfunction is a recognised feature of many diseases including HCM but it is unclear whether mitochondrial abnormalities are a primary or secondary event in HCM (Unno, Isobe et al. 2009).

5.3.2.1 The electron transport chain

Oxidative phosphorylation mediated by the electron transport chain (ETC) across the inner mitochondrial membrane accounts for more than 90 % of the ATP synthesized in cardiomyocytes (Chen and Zweier 2014). Energy is derived from the oxidation of nicotinamide adenine dinucleotide (NADH) and flavin adenine dinucleotide (FADH₂) by the transfer of electrons in a series of redox reactions. The ten NADH that enter the ETC at Complex I originate from each of the earlier processes of respiration: two from glycolysis, two from the transformation of pyruvate into acetyl-CoA, and six from the citric acid cycle. The two FADH₂ are produced by the citric acid cycle and enter the ETC at Complex II (Figure 5.13). In the process, they turn back into NAD⁺ and FAD which can be reused in other steps of cellular respiration. The protein complexes (Complex I, II, III, IV and V) are made up of about 100 different proteins, of which 13 are encoded by mitochondrial DNA (mtDNA) and the remainder being encoded by nuclear DNA (nDNA) (Table 5.3). Pathway analysis found decreased expression of multiple protein sub units from Complex I, III, IV and V in HCM compared with normal controls (Figure 5.12). This finding is in keeping with evidence from animal and human studies which suggest that HCM is also characterised by a reduction in the concentration of high-energy phosphates in the myocardium (Spindler, Saupe et al. 1998, Jung, Hoess et al. 2000). The cause is not fully understood, but the demonstration of reduced high phosphate ratios in patients with mutations in sarcomere proteins, and little or no hypertrophy, has led to the hypothesis that energy deficiency is a fundamental driver of the HCM phenotype (Crilley, Boehm et al. 2003).

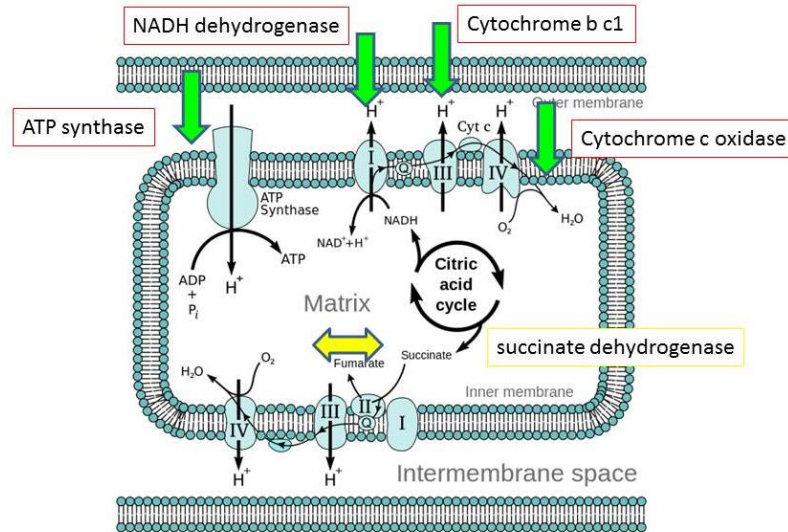


Figure 5.13 The electron transport chain

Complex	Name	mtDNA	nDNA	assembly
Complex I	NADH dehydrogenase	7	~39	~11
Complex II	Succinate dehydrogenase	0	4	~2
Complex III	Cytochrome bc1 complex	1	10	~9
Complex IV	Cytochrome c oxidase	3	10	~30
Complex V	ATP synthase	2	~14	~3

Table 5.3 Table summarising the number of proteins within each complex.

5.3.2.2 Mechanism of mitochondrial dysfunction

Citrate synthase

Mitochondrial citrate synthase (CISY, O75390) is the first enzyme of the citric acid or tricarboxylic acid (TCA) cycle responsible for catalysing the condensation of acetyl-coenzyme A and oxaloacetate to yield citric acid and CoA. Citrate synthase activity is frequently used as a mitochondrial marker and also as a measure of mitochondrial enrichment (Hargreaves, Heales et al. 1999). We found no difference in expression of CISY in HCM compared with normal hearts implying that the observed changes in the electron transport chain were not due to fewer mitochondria overall.

Mitochondrial DNA

Patients with mutations in mitochondrial DNA can have a similar cardiac phenotype to patients with sarcomeric protein mutations, supporting the notion that primary energetic deficits can produce the cardiac phenotype of HCM (Obayashi, Hattori et al. 1992). It is also recognised that mitochondrial dysfunction plays a critical role in the development of heart failure (Neubauer 2007). Given that there was no difference in expression of Complex II, which is encoded entirely by nuclear DNA, another possible hypothesis is that perturbations in mtDNA are contributory to the reduced expression of proteins sub units making up the ETC. This idea has been considered in patients with congenital heart disease where mtDNA depletion was attributed to reduced mtDNA replication in both hypertrophied and failing right ventricles (Karamanlidis, Bautista-Hernandez et al. 2011).

5.3.2.3 Relation to altered metabolism

Since enzymes for fatty acid oxidation are located in mitochondria, it is unclear whether abnormalities in fatty acid oxidation (Section 5.3.1.2) are secondary to mitochondrial dysfunction. Nonetheless, it has led a number of investigators to test the hypothesis that pharmacological modulation of myocardial substrate utilisation might prove therapeutically advantageous and one clinical trial using perhexiline a CPT-1 inhibitor, has shown to be effective in HCM (Abozguia, Elliott et al. 2010).

5.3.2.4 Relation to oxidative stress

Apart from energy production, the ETC is also the main source of reactive oxygen species (ROS) and plays a central role in the regulation of oxidative stress. Chronic increases in ROS production in the mitochondria lead to a vicious cycle of mtDNA damage, further ROS generation, and cellular injury. Superoxide dismutase Cu Zn (SODC, P00441) a cytoplasmic enzyme which catalyses the dismutation of superoxides was significantly higher in HCM than controls, suggesting activation of a defence mechanism against oxidative stress. There were also increased levels of the extracellular form of the enzyme (SODE, P08294). The importance of oxidative stress is increasingly emerging as a mechanism responsible for the development and progression of heart failure (Takimoto and Kass 2007). Specifically, ROS can directly impair contractile function by modifying excitation-contraction coupling,

activating hypertrophy signalling kinases and transcription factors and mediating apoptosis. ROS also stimulate cardiac fibroblast proliferation and activate the matrix metalloproteinases (MMPs), leading to the extracellular matrix remodelling.

5.3.3 Stress response

The proteome of HCM showed an increase in proteins associated with cell stress e.g. heat shock 70 kDa protein 6 (HSP76, P17066) and heat shock protein beta 6 (HSPB6, O14558). This finding is consistent with previous findings by 1D-SDS-PAGE of differential expression of heat shock proteins (HSPs) in heart failure (Knowlton, Kapadia et al. 1998). Several heat shock proteins function as intracellular chaperones for other proteins. They play an important role in protein-protein interactions such as folding and assisting in the establishment of proper protein conformation and prevention of unwanted protein aggregation. IPA revealed significant overrepresentation of the protein ubiquitination pathway within our expression dataset due in part to the upregulation of HSPs (Figure 5.14).

5.3.3.1 Protein Ubiquitination

The protein ubiquitination pathway plays a major role in the degradation of short-lived or regulatory proteins involved in various cell processes (e.g. cell proliferation, apoptosis, antigen presentation, DNA repair, transcription regulation and ion channels regulation). There was an increase in polyubiquitin B (UBB, P0CG47) which may facilitate increased protein ubiquitination in HCM by the downstream 26S proteasome complex leading to protein degradation. There was supporting evidence for activation of the 26S proteasome complex with upregulation of other sub units involved in its pathway e.g. proteasome subunit beta type 2 (PSB2, P49721).

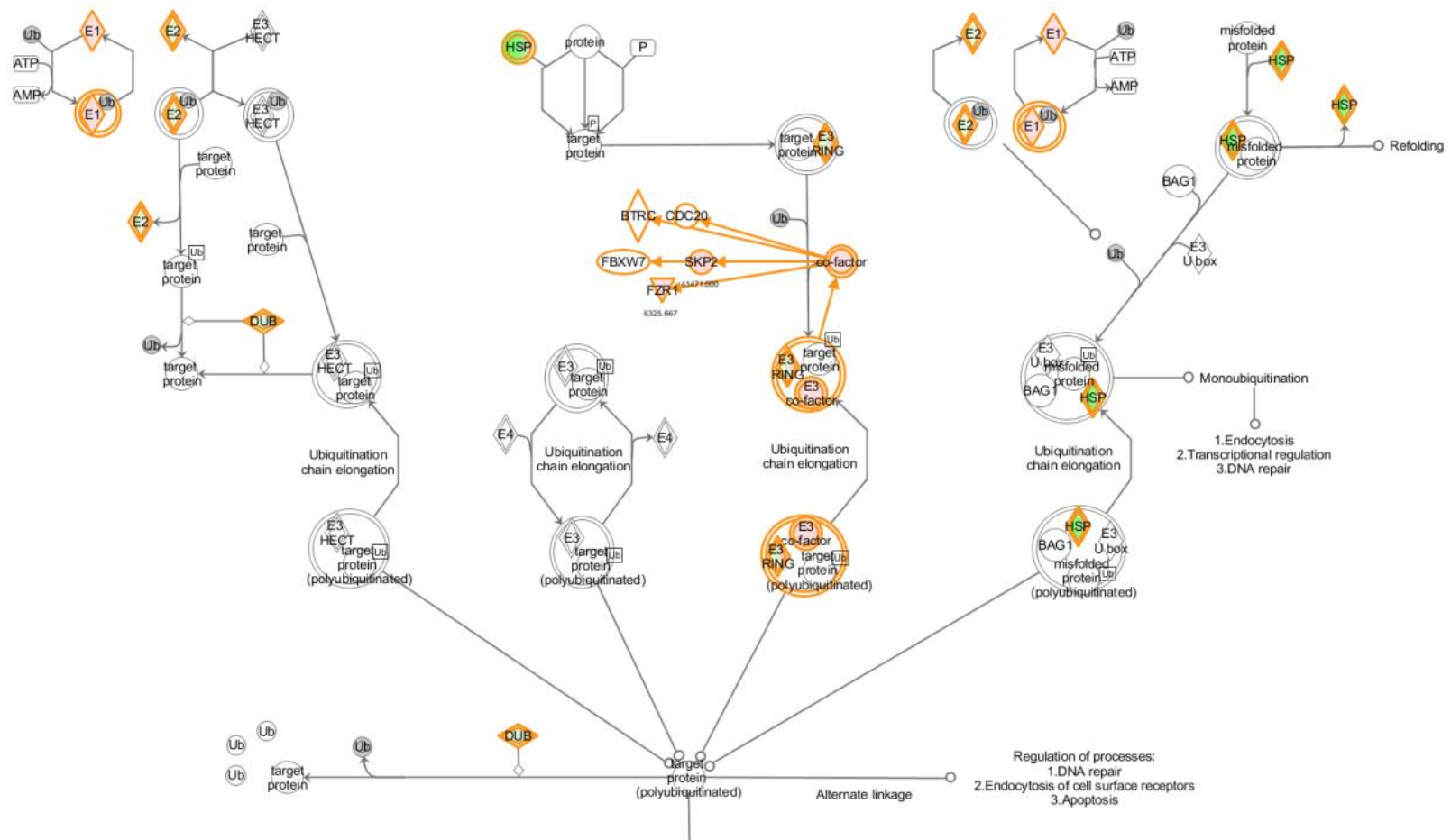


Figure 5.14 Proteins involved in canonical pathway "protein ubiquitination"

Green and red/orange colour indicates proteins significantly increased and decreased in expression, respectively. Grey indicates no significant change in protein expression.

5.4 Summary

Proteomic analysis of myocardium from HCM hearts was compared to normal hearts using label free LC-MS^E technology. Differentially expressed proteins showed enrichment for the GO cell class term “cytoskeletal” and the GO biological process “metabolic process”. There was evidence for reactivation of the fetal gene program with alteration in both structural and metabolic proteins. This was associated with significant mitochondrial dysfunction with reduced expression of several complexes within the electron transport chain. Oxidative stress and protein ubiquitination pathways were also activated in keeping with generalised cardiomyocyte dysfunction, apoptosis and heart failure. Novel proteins that link some of these processes were identified and selected for further validation.

6 Validation and Verification of Proteins

6.1 Study design and rationale

The validation study was performed in a larger group of myocardial samples. The 17 myocardial samples used in the discovery study were all included. Additional prospectively collected samples were obtained from patients with a) HCM and b) aortic stenosis (Figure 6.1). Aortic stenosis (AS) is a narrowing of the aortic valve in the heart in which the restricted blood flow through the valve results in pressure overload of the left ventricle causing cavity remodelling and LV hypertrophy. Once AS has become severe, treatment involves valve replacement surgery. Samples were obtained from the interventricular septum in patients with AS undergoing valve replacement surgery. These patients therefore served as a “disease control” group for HCM, where hypertrophy is a primary feature of the disease. The aim was to understand whether protein levels were altered due to a generalised hypertrophic response of the heart or as a result of the specific genetic disorder.






Sample	Discovery Study	Validation Study
Control		
HCM		
Aortic stenosis		

Figure 6.1 Validation study design

The validation study included 17 samples (bold colours) used in the discovery study (internal validation) and 48 additional samples (lighter colour) which were prospectively obtained (external validation).

6.2 Development of a multiplexed biomarker assay

Targeted MRM-based triple quadrupole mass spectral assays were developed to validate findings from the discovery Q-ToF study.

6.2.1 Selection of candidate protein biomarkers

From the proteins that showed a significant alteration during proteomic profiling by Q-ToF-MS, several had been described previously in hypertrophic heart conditions e.g. increased levels of four and a half LIM domains protein 1 (FHL1, Q13642) and reduced expression of the calcium cycling protein, SERCA2a (ATP2A2, P16615). Following a literature review several potential biomarkers were selected for validation. Candidate markers were selected based on significance, fold change and quality of the label free proteomics analysis data. Established cardiac and muscle biomarkers such as Creatine kinase M-type (KCM, P06732), Troponin I (*TNNI3*, P19429) and Troponin T (TNNT2, P45379) were also included in the included in the final multiplexed assay. Other proteins were included as potential markers of altered metabolism, oxidative stress or inflammation.

6.2.1.1 Selection of peptide markers

Representative tryptic peptides for each protein were determined from the label free proteomics data (ideally using one of the top three most abundant peptides and selecting the optimum daughter spectra for quantitation). Other peptides were selected using the open source online global proteome machine MRM database at www.thegpm.org (Craig, Cortens et al. 2004). Peptides were standardised to a spiked internal standard (Section 6.2.2). Absolute levels were obtained from standard curves. Standard curve linearity of $r^2 > 0.9$ was required for all calibration curves. Data was exported to Microsoft Excel and Graph Pad Prism for statistical analysis. A standard curve of each peptide was analysed at the start and the end of the run for quantitation and performance standardisation.

6.2.2 Optimisation of an internal standard

For validation of multiple biomarkers, internal standards were developed in-house (rather than purchasing separate AQUA peptides for each protein). Both a protein and peptide base internal standard were evaluated and included in the final assay.

6.2.2.1 Peptide based internal standard

A peptide sequence was designed so that it contained one tryptic cleavage site and had a unique sequence not found in other mammals. The peptide was based on the sequence of [glu1]-fibrinopeptide B, making an alteration of a single neutral amino acid (proline) and the addition of a naturally occurring sequence tag, so that it would be cleaved into a unique sequence for quantification. On BLAST search, the highest sequence similarity was 90%, with Fibrinogen beta chain (FIBB, P02675) i.e. the protein it is derived from. By including a theoretical cleavage site, the internal standard could be added prior to digestion to account for any variations in digestion. The peptide (now referred to as GF-IS) was custom synthesised (Generon, UK) and precursor and product ions determined as described in Section 3.2.5.2. The three transitions showing the best linear response in myocardial protein matrix were selected for inclusion in the assay:

Peptide sequence	Precursor m/z	Cone (V)	Product m/z	Collision (V)
EGVNDPEEGFFSARGHRPLD	777.649	24	480.3168	22
	777.649	24	497.4003	20
	777.649	24	1039.816	26

6.2.2.2 Whole protein internal standard

Yeast enolase was used as a second internal standard (YE-IS) because it was able to be introduced as whole protein and it was the internal standard that had been used for the discovery experiments. The three most abundant peptides (spectral count) from the Q-ToF were selected for tuning and inclusion in the MRM assay. The best transitions were selected based on linearity in sample matrix and absence of endogenous peptides in the sample matrix at the same retention time. By including three peptides of yeast enolase in the assay, both high and low abundant peptides could be quantified within the same sample run (Table 6.1). The chromatogram displaying the two different internal standards is shown in Figure 6.2.

Peptide sequence	Precursor m/z	Cone (V)	Product m/z	Collision (V)
SIVPSGASTGVHEALEMR	614.58+++	46	771.73	14
	614.58+++	46	547.88	12
	614.58+++	46	514.88	14
GNPTVEVELTTEK	709.06++	46	948.68	20
	709.06++	46	623.49	18
AVDDFLISLDGTANK	790.16++	46	918.67	22
	790.16++	46	805.57	20

Table 6.1 The internal standard optimised for UPLC-MS/MS assay

The MRM transitions for the yeast enolase internal standard are shown with instrument operating parameters and parent/daughter ions.

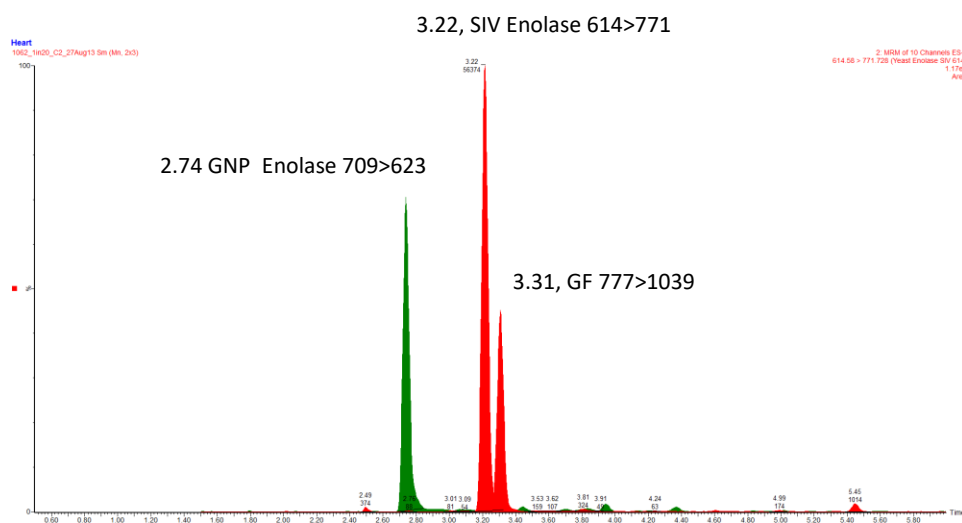


Figure 6.2 Overlaid chromatogram of the internal standard peptides.

Two internal standards were included in the multiplexed targeted proteomic assay.

6.2.3 Multiplexed assay to measure myocardial proteins

Peptides were tuned on the mass spectrometer according to methods described in Section 3.2.5.2. The multiplexed assay was developed to quantitate 35 peptides in three LC runs of ten minutes each. At least two transitions per peptide were included and two peptides per protein were included for specific proteins (actin, transthyretin and lumican). Each LC method contained up to 40 transitions including those of the internal standard. The selected peptides, transitions and tuning parameters for the UPLC-MS/MS analysis are shown in Table 6.2 and a typical chromatogram shown in Figure 6.3.

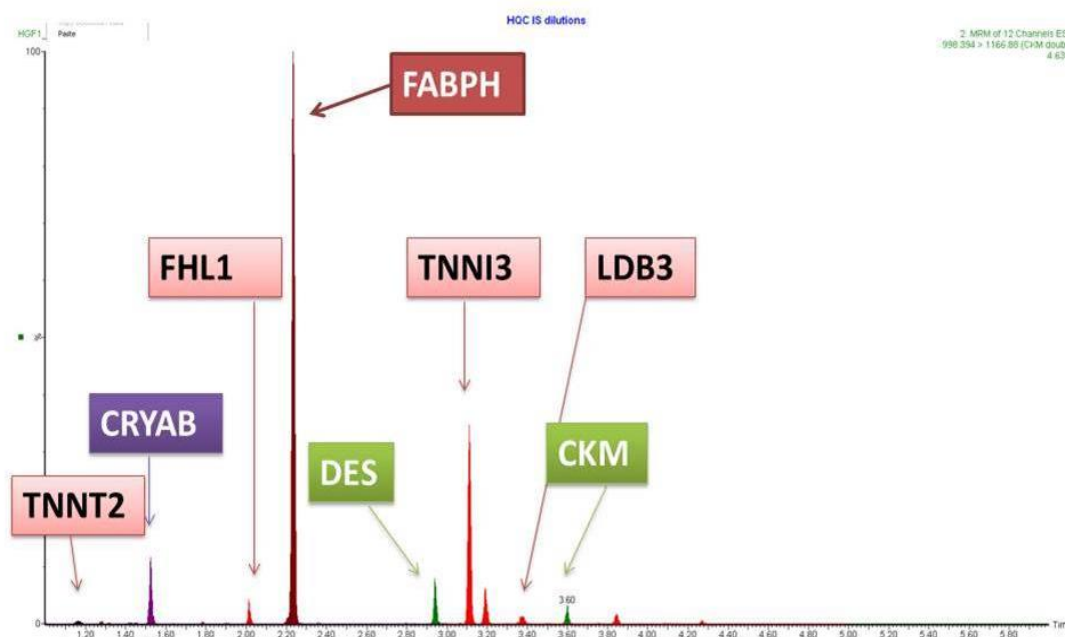


Figure 6.3 Typical UPLC MS/MS chromatogram.

This shows the 10-minute assay developed for the quantitation of individual proteins

Uni Prot ID	Protein name Peptide sequence	Precursor m/z	Cone (V)	Product m/z	Collision (V)
Q13510	Acid ceramidase LTVYTTLIDVTK	683.8681 ⁺⁺	54	890.4025	22
		<u>683.8681⁺⁺</u>	<u>54</u>	<u>1053.4551</u>	<u>22</u>
P68032	Actin, alpha cardiac muscle 1 Ac-DEETTALVCDNGSGLVK	982.42 ⁺⁺	60	<u>273.02</u>	<u>60</u>
		982.4666 ⁺⁺	56	697.1434	38
		<u>982.4666⁺⁺</u>	<u>56</u>	<u>786.1389</u>	<u>26</u>
P68133	Actin, alpha skeletal muscle, DLYANNVMSGGTTMYPGIADR	749.0957 ⁺⁺⁺	36	790.3639	20
		<u>749.0957⁺⁺⁺</u>	<u>36</u>	<u>809.9206</u>	<u>12</u>
	Ac-EDETTALVCDNGSGLVK	982.42 ⁺⁺	60	<u>287.01</u>	<u>60</u>
		982.4666 ⁺⁺	56	697.1434	38
P02511	Alpha-crystallin B chain, QDEHGFISR	<u>363.5851⁺⁺⁺</u>	<u>2</u>	<u>536.5341</u>	<u>8</u>
		363.5851 ⁺⁺⁺	2	579.5365	14
P07451	Carbonic anhydrase 3 SLLSSAENEPVPLVSNWRPPQPINNR	<u>1004.7872⁺⁺⁺</u>	<u>58</u>	<u>1041.2473</u>	<u>30</u>
		1004.7872 ⁺⁺⁺	58	1162.7921	30
P01024	Complement C3 SSLSVPYVIVPLK	<u>701.7766⁺⁺</u>	<u>4</u>	<u>456.3733</u>	<u>16</u>
		701.7766 ⁺⁺	4	928.8845	16
P06732	Creatine kinase M-type GTGGVDTAAVGVSFDVSNADR	998.393 ⁺⁺	44	661.4891	38
		<u>998.3937⁺⁺</u>	<u>44</u>	<u>1166.8797</u>	<u>32</u>
P17661	Desmin FLEQQNAALAAEVNR	<u>837.8404⁺⁺</u>	<u>58</u>	<u>588.5028</u>	<u>24</u>
		837.8404 ⁺⁺	58	843.7887	30
P06396	Gelsolin TPSAAYLWVGTGASEAEK	919.9043 ⁺⁺	8	849.6734	26
		<u>919.9043⁺⁺</u>	<u>8</u>	<u>1134.9157</u>	<u>30</u>
P78417	Glutathione S transferase omega 1 GSAPPGPVPEGSIR	<u>661.0957⁺⁺</u>	<u>46</u>	<u>553.5224</u>	<u>22</u>
		661.0957 ⁺⁺	46	1008.8277	28
P04513	Fatty acid-binding protein, heart SIVTLDGK	445.351 ⁺⁺	38	590.4842	14
		<u>445.3511⁺⁺</u>	<u>38</u>	<u>689.5508</u>	<u>12</u>
Q13642	Four and a half LIM domains protein 1 AIVAGDQNVEYK	<u>654.1596⁺⁺</u>	<u>12</u>	<u>952.6599</u>	<u>20</u>
		654.1596 ⁺⁺	12	1023.7251	22
P04075	Fructose-bisphosphate aldolase A ALQASALK	401.4149 ⁺⁺	36	489.4766	12
		<u>401.4149⁺⁺</u>	<u>36</u>	<u>617.5151</u>	<u>10</u>
O75112	LIM domain binding protein 3 SASYNLSLTLQK	663.1596 ⁺⁺	38	489.4389	16
		<u>663.1596⁺⁺</u>	<u>38</u>	<u>689.6425</u>	<u>20</u>
P51884	Lumican SLEYLDLSFNQIAR	835.1234 ⁺⁺	46	948.5396	32
		<u>835.1234⁺⁺</u>	<u>46</u>	<u>1063.521</u>	<u>22</u>
		NIPTVNENLENYYLEVNQLEK	<u>846.2596⁺⁺⁺</u>	<u>54</u>	<u>1155.1929</u>
P11310	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial, EPGLGFSFEFTEQQK	1268.6957 ⁺⁺	58	1155.0613	36
		872.351 ⁺⁺	54	863.4099	16
		<u>872.351⁺⁺</u>	<u>54</u>	<u>1142.9229</u>	<u>24</u>

P12883	Myosin 7 EDQVMQQNPPK	<u>657.5696⁺⁺</u> 657.5696 ⁺⁺	<u>52</u> 52	<u>355.1985</u> 842.6616	<u>24</u> 20
Q14896	Myosin-binding protein C, cardiac-type VFSQNMVGFSDR	<u>694.095⁺⁺</u> 694.0957 ⁺⁺	<u>28</u> 28	<u>581.458</u> 680.5242	<u>20</u> 22
P36871	Phosphoglucomutase 1 YDYEEVEAEGANK	<u>759.1596⁺⁺</u> 759.1596 ⁺⁺	<u>50</u> 50	<u>718.5202</u> 946.7224	<u>24</u> 24
P15259	Phosphoglycerate mutase 2 VLIAAHGNSLR	<u>576.0319⁺⁺</u> <u>576.0319⁺⁺</u>	<u>12</u> <u>12</u>	<u>413.4226</u> <u>683.5368</u>	<u>20</u> <u>20</u>
P02545	Prelamin-A/C SGAQASSTPLSPTR	<u>680.5852⁺⁺</u> 680.5852 ⁺⁺	<u>14</u> 14	<u>373.2957</u> 945.8009	<u>18</u> 22
Q99497	Protein DJ 1 DVVICPDASLEDACK	<u>554.2234⁺⁺⁺</u> <u>554.2234⁺⁺⁺</u>	<u>18</u> <u>18</u>	<u>674.1549</u> <u>723.7188</u>	<u>10</u> <u>10</u>
P22061	Protein L isoaspartate D aspartate O methyltransferase VFEVMLATDR	<u>591.0319⁺⁺</u> 591.0319 ⁺⁺	<u>28</u> 28	<u>376.3119</u> 462.3853	<u>16</u> 14
P23297	Protein S100 A1 ELLQTELSGFLDAQKDVAVDK	812.5425 <u>812.5425</u>	44 <u>44</u>	861.4399 <u>976.5466</u>	22 <u>24</u>
P16615	Sarcoplasmic reticulum Ca ²⁺ ATPase 2 VDQSILTGESVSVIK	525.351 ⁺⁺⁺ <u>787.4787⁺⁺</u>	26 <u>16</u>	714.9536 <u>313.2295</u>	10 <u>40</u>
Q96199	Succinyl CoA ligase FFVADTANEALEAAK	799.287 ⁺⁺ <u>799.287⁺⁺</u>	58 <u>58</u>	652.1945 <u>1203.9746</u>	18 <u>22</u>
P04179	Superoxide dismutase Mn AIWNVINWENVTER	872.6357 ⁺⁺ <u>872.6357⁺⁺</u>	30 <u>30</u>	1047.4769 <u>1160.5748</u>	26 <u>28</u>
P00441	Superoxide dismutase Cu Zn GPVQGIINFEQK	665.3143 ⁺⁺ <u>665.3143⁺⁺</u>	18 <u>18</u>	1059.3052 <u>1076.3887</u>	24 <u>22</u>
P02766	Transthyretin GSPAINVAVHVFR YTIAALLSPYSYSTTAVVTNPK	683.9043 ⁺⁺ <u>683.9043⁺⁺</u> <u>1180.8872⁺⁺</u> 1180.8872 ⁺⁺	48 <u>48</u> <u>56</u> 56	611.8618 <u>941.399</u> <u>1527.9128</u> 1614.9785	20 <u>26</u> <u>32</u> 38
P19429	Troponin I, cardiac muscle NITEIADLTQK	623.606 ⁺⁺ <u>623.6063⁺⁺</u>	34 <u>34</u>	788.6749 <u>1018.8335</u>	18 <u>18</u>
P45379	Troponin T, cardiac muscle EAEDGPMEESKPKPR	425.9043 ⁺⁺⁺⁺ <u>425.9043⁺⁺⁺⁺</u>	28 <u>28</u>	457.8691 <u>494.8973</u>	10 <u>10</u>
P49748	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial ALEQFATVVEAK	435.7766 ⁺⁺⁺ <u>435.7766⁺⁺⁺</u>	2 <u>2</u>	429.8704 <u>543.2234</u>	6 <u>10</u>

Table 6.2. MRM transitions for the UPLC-MS/MS analysis of heart proteins

Ion charge is indicated by ++ for double charge, +++ for triple charge and ++++ for quadruple charge. Product ions underlined were those used for quantitation whilst those not underlined were used for confirmation only.

6.3 Results

6.3.1 Study cohort

The validation study cohort consisted of a total of sixty-five patients (7 controls, 51 HCM, 7 AS). The clinical characteristics of the study population are shown in Table 6.3. The main differences between patients with HCM and AS were younger age (51 ± 14 versus 69 ± 10 years, $p < 0.05$) and greater maximal left ventricular wall thickness (19 ± 4 versus 11 ± 1 mm, $p < 0.05$). The control and HCM patients were appropriately age and sex matched. Clinical details such as medication or previous medical history were not available for the control subjects (Table 4.1).

6.3.2 Comparison of Targeted MS with Label Free Quantitative Proteomics

Using quantitative MS/MS eight proteins were found to be significantly altered between the HCM and control groups (Table 6.4). Actin, alpha skeletal muscle (ACTS, P68133), Four and a half LIM domains protein 1 (FHL1, Q13642), Desmin (DES, P17661), Carbonic anhydrase 3 (CA3, P07451) and Lumican (LUM, P51884) were significantly elevated in HCM versus control. Fructose-bisphosphate aldolase A (ALDOA, P04075), Phosphoglycerate mutase (PGAM2, P15259 and PGAM1, P18669) and Creatine kinase M-type (KCM, P06732) were significantly reduced. The direction of change was consistent with the pilot and discovery study in all cases; however, the differences were not statistically significant in the discovery study for three of these proteins.

Seven proteins were significantly altered in the discovery study (fold change and p value given) but not validated in the targeted assay: Acid ceramidase (ASAHL1, Q13510; \uparrow 1.28-fold, $p = 0.02$), LIM domain binding protein 3 (LDB3, O75112; \uparrow 1.33-fold, $p < 0.01$), Myosin-binding protein C (MYBPC3, Q14896; \uparrow 1.36-fold, $p < < 0.01$), Sarcoplasmic reticulum calcium ATPase 2 (ATP2A2, P16615; \downarrow 1.69-fold, $p < 0.01$), Transthyretin (TTR, P02766; \uparrow 1.47-fold, $p = 0.04$), Troponin I, cardiac muscle (TNNT3, P19429; \uparrow 1.66-fold, $p = 0.03$), Very long-chain specific acylCoA dehydrogenase mitochondrial (ACADV, P49748; \downarrow -1.22 fold, < 0.01).

	Control	Aortic Stenosis	HCM	P value
Number of subjects	7	7	51	-
Male (%)	3 (43%)	3 (43%)	27 (53%)	NS
Age (years)	47 ± 10	69 ± 10	51 ± 14	<0.05
BMI (kg/m ²)	-	27 ± 5	30 ± 5	NS
Sinus rhythm (%)	-	6 (86 %)	50 (98 %)	NS
Paroxysmal AF (%)	-	0 (0 %)	8 (16 %)	-
VT or VF (%)	-	0 (0 %)	7 (14 %)	-
NT-proBNP (pmol/L)	-	92 ± 132	221 ± 267	NS
eGFR (ml/min)	-	88 ± 27	82 ± 20	NS
CRP (mg/L)	-	2.1 ± 2.7	4.6 ± 6.7	NS
Hb (g/dL)	-	14.3 ± 1.7	13.7 ± 1.6	NS
Hypertension (%)	-	3 (43 %)	17 (33 %)	NS
Coronary disease (%)	-	1 (14 %)	1 (2 %)	NS
Diabetes (%)	-	1 (14 %)	3 (6 %)	NS
Smoking (%)	-	2 (29 %)	9 (18 %)	NS
Family history of HCM (%)	-	0 (0 %)	16 (31 %)	-
Family history of SCD (%)	-	0 (0 %)	6 (12 %)	-
History of syncope (%)	-	0 (0 %)	13 (26 %)	-
<u>Medication</u>				
Beta-blocker	-	4 (57 %)	38 (75 %)	NS
Calcium-blocker	-	1 (14 %)	13 (26 %)	NS
Disopyramide	-	0 (0 %)	24 (47 %)	-
ACE or ARB	-	3 (43 %)	3 (6 %)	<0.05
Warfarin	-	1 (14 %)	12 (24 %)	NS
Amiodarone	-	0 (0 %)	7 (14 %)	NS
Statin	-	4 (57 %)	11 (22 %)	<0.05
<u>Echocardiogram</u>				
Max wall thickness (mm)	-	11 ± 1	19 ± 4	<0.05
LA size (mm)	-	40 ± 6	45 ± 7	NS
LVOT/AV gradient (mm Hg)	-	72 ± 10	91 ± 38	-
Moderate MR (%)	-	0 (0 %)	30 (59 %)	-
RVH (%)	-	0 (0 %)	11 (22 %)	-
<u>Operative details</u>				
Aortic cross-clamp (min)	-	61 ± 17	75 ± 32	NS
Cardiopulmonary bypass (min)	-	90 ± 27	115 ± 63	NS

Table 6.3 Characteristics of patients included in the validation study

BMI body mass index, *VT* ventricular tachycardia, *VF* ventricular fibrillation, *RVH* right ventricular hypertrophy, *NT-proBNP* N-terminal pro brain natriuretic peptide, *eGFR* estimated glomerular filtration rate, *CRP* C-reactive protein, *Hb* haemoglobin, *HCM* hypertrophic cardiomyopathy, *SCD* sudden cardiac death, *ACE* angiotensin converting enzyme inhibitor, *ARB* angiotensin receptor blocker, *LA* left atrium, *LVOT/AV* left ventricular outflow tract/aortic valve, *MR* mitral regurgitation, *RVH* right ventricular hypertrophy. Mean and SD are shown. *P* values are calculated using *t*-test or Chi-squared test to compare means and proportions between subjects in the AS and HCM group.

Phase of study (number of samples)		Pilot (n=3)		Discovery (n=17)		Validation (n=58)	
Chapter of thesis		Ch. 4		Ch. 5		Ch. 6	
ID	Protein name	Fold change		Fold change	p-value	p-value	
Q13510	Acid ceramidase	-	-	↑	1.28	0.02	→ NS
P68032	Actin, alpha cardiac muscle 1	→	-	→	1.01	0.86	→ NS
P68133	Actin, alpha skeletal muscle,	↑	4.3	↑	1.41	0.05	↑ <0.01
P02511	Alpha-crystallin B chain,	→	-	→	1.05	0.83	→ NS
P07451	Carbonic anhydrase 3	↑	3.8	↑	4.69	<0.01	↑ <0.01
P01024	Complement C3	→	-	→	1.01	0.88	→ NS
P06732	Creatine kinase M-type	→	-	→	1.06	0.44	↓ <0.01
P17661	Desmin	↑	1.5	→	1.01	0.86	↑ <0.01
P06396	Gelsolin	↑	1.5	→	1.27	0.36	→ NS
P78417	Glutathione S transferase omega 1	→	-	→	1.04	0.72	→ NS
P04513	Fatty acid-binding protein, heart	→	-	→	1.05	0.81	→ NS
Q13642	Four and a half LIM domains protein 1	↑	1.8	↑	1.49	<0.01	↑ <0.01
P04075	Fructose-bisphosphate aldolase A	↓	1.3	↓	1.17	0.02	↓ <0.01
O75112	LIM domain binding protein 3	↑	1.7	↑	1.33	<0.01	→ NS
P51884	Lumican	↑	1.8	↑	1.76	0.02	↑ <0.01
P11310	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial,	→	-	→	1.13	0.1	→ NS
P12883	Myosin 7	↑	1.5	→	1.3	0.15	→ NS
Q14896	Myosin-binding protein C	↑	1.9	↑	1.36	<0.01	→ NS
P36871	Phosphoglucomutase 1	→	-	→	1.11	0.12	→ NS
P18669	Phosphoglycerate mutase 1	→	1.0	→	1.06	0.34	↓ <0.01
P15259	Phosphoglycerate mutase 2	↓	1.4	→	1.12	0.80	
P02545	Prelamin-A/C	→	-	→	1.29	0.31	→ NS
Q99497	Protein DJ 1	→	-	→	1.03	0.56	→ NS
P22061	Protein L isoaspartate D aspartate O methyltransferase	→	-	→	1.12	0.22	→ NS
P23297	Protein S100 A1	↑	1.5	→	1.14	0.1	→ NS
P16615	Sarcoplasmic reticulum calcium ATPase 2	↓	1.4	↓	1.69	<0.01	→ NS
Q96199	Succinyl CoA ligase GDP forming subunit β mitochondrial	→	-	→	1.06	0.48	→ NS
P04179	Superoxide dismutase Mn	-	-	→	1.09	0.33	→ NS
P00441	Superoxide dismutase Cu Zn	→	-	↑	1.25	0.07	- -
P02766	Transthyretin	-	-	↑	1.47	0.04	→ NS
P19429	Troponin I, cardiac muscle	↑	1.4	↑	1.66	0.03	→ NS
P45379	Troponin T, cardiac muscle	↑	1.4	→	1.48	0.13	→ NS
P49748	Very long-chain specific acylCoA dehydrogenase mitochondrial	↓	1.3	↓	1.22	<0.01	→ NS

Table 6.4 Comparison of targeted MS/MS with LC-MS analysis

The direction and fold change of each protein are shown for HCM compared with control. The findings from the pilot study using pooled samples (n=3) and discovery study using separately profiled samples (n=17) are shown with the rows highlighted where the direction of change was consistent with both the pilot and discovery study (green) or the discovery study only (yellow).

6.4 Validation and Relevance of Specific Proteins

6.4.1 Alpha Actins

Actin is a globular multifunctional protein present in all eukaryotic cells. The alpha-actins, found in muscle tissues, are a major constituent of the contractile apparatus. Mutations in actin are responsible for a range of inherited muscle disorders.

The predominant alpha actin isoform in the heart is alpha-cardiac actin (*ACTC*, P68032; *ACTC1*) but alpha skeletal muscle actin (*ACTS*, P68133; *ACTA1*) also exists in lower levels. In the normal adult heart, *ACTS* positive cardiomyocytes are distributed in a transmural gradient with the highest proportion located sub-endocardially, but in myocardial hypertrophy and cardiomyopathies, the amount of *ACTS* is increased in all layers of the myocardium (Suurmeijer, Clement et al. 2003). Thought to be a marker of secondary hypertrophy, an increase in *ACTS* is recognised as part of the hypertrophic fetal gene program. The isoform switch from cardiac to skeletal alpha actin has been reported to occur in animal models of heart failure, atrial fibrillation and hypertension (Driesen, Verheyen et al. 2009). Gene expression work suggests *ACTA1* mRNA is four fold higher in HCM (Lim, Roberts et al. 2001).

One of the challenges of quantifying actin isoforms at protein level is sequence homology: both proteins are 375 amino acids long but differ in only 4 amino acids. This means the isoforms are difficult to resolve in a gel as they have virtually the same molecular mass (cardiac 41.840 kDa, skeletal 41.873 kDa) and therefore almost identical IEF points. Using a MRM mass spectrometry approach, abundance of each isoform can be determined using a peptide from the N-terminus which differs only in amino acid sequence but not overall mass, therefore having unique and common MRM transitions allowing relative quantification without the need for an internal standard (Ravenscroft, Colley et al. 2008). Actin is irreversibly acetylated in vivo on its N-terminus via the transfer of an acetyl group from acetyl coenzyme A to the termini of alpha-amino groups. This PTM is essential for the function of actin filaments and its interaction with other proteins (Vandekerckhove and Weber 1978).

For quantification of cardiac and skeletal actin the following MRM transitions and collision energies (CE) for the acetylated (Ac), carbamidomethyl modified (*) peptides were selected:

Peptide	Unique transition n	Common transition
Ac-DEDETTALVC*DNGSGLVK	982.9 > 287.1 (65 V) b 2 ion skeletal muscle α -actin	982.9 > 1048.5 (54 V) y 10 ion
Ac-DDEETALVC*DNGSGLVK	982.9 > 273.1 (65 V) b 2 ion cardiac muscle α -actin	

Figure 6.4 illustrates how the high resolution/HPLC mass spectrometer can distinguish between the peptides based on retention time and unique daughter ions. The percentage of cardiac α -actin expression was determined by calculating the ratio of the area under the curve (AUC) of the skeletal muscle α -actin specific b2 daughter ion to the transition of the common striated α -actin N-terminal (standard) and the ratio of the AUC of the cardiac α -actin specific b2 daughter ion to the standard.

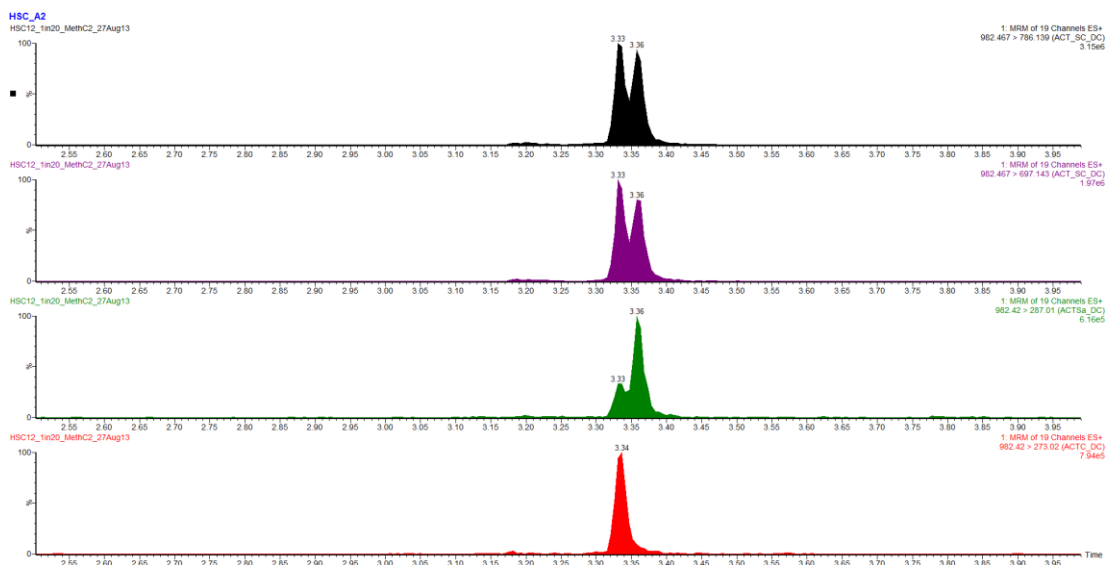


Figure 6.4 MRM transitions of cardiac and skeletal actin.

Chromatograms of the different alpha-actin isoforms. Top two panels (black and purple) show common transitions; 3rd panel (green) shows unique transition for skeletal actin; 4th panel (red) shows unique transition for cardiac actin.

The percentage of cardiac α -actin was significantly lower in HCM ($54 \pm 3 \%$) than control hearts ($65 \pm 7 \%$; $p < 0.05$). This is consistent with gene expression data but has not been shown before at protein level in HCM. This supports the notion that fetal reprogramming occurs in patients with hypertrophic cardiomyopathy. There was a trend towards lower amount of cardiac α -actin in the AS group ($61 \pm 8 \%$), but the difference was not statistically significant (Figure 6.5). This may reflect the more modest hypertrophy compared with HCM.

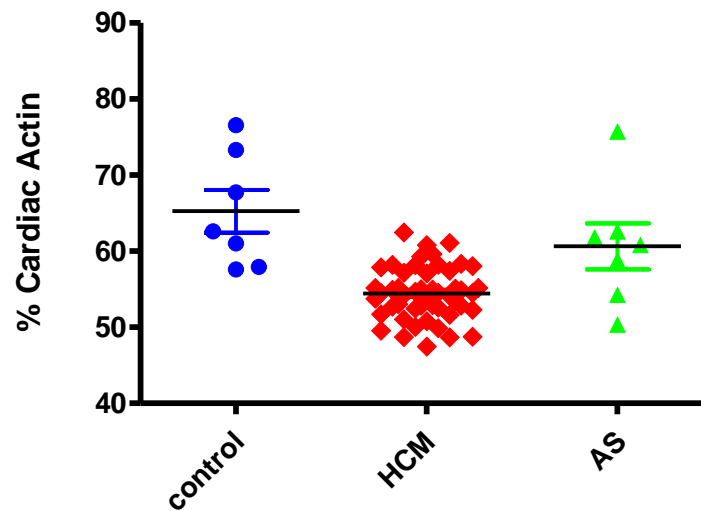


Figure 6.5 Percentage of cardiac actin detected in myocardial samples.

Scatter plot illustrating percentage of cardiac alpha-actin as a proportion of total skeletal and cardiac alpha actin isoforms, as detected by UPLC-MS/MS in controls, hypertrophic cardiomyopathy (HCM) and aortic stenosis (AS).

6.4.2 Four and a Half LIM domains protein 1 (FHL1)

Four and a half LIM domains protein 1 (FHL1, Q13642; *FHL1*) is expressed in cardiac and skeletal muscle where it regulates sarcomere synthesis (Cowling, McGrath et al. 2008). Mutations in the *FHL1* gene, located on Xq26, are responsible for muscle dystrophies and cardiomyopathies, and isolated HCM has been reported in males without obvious skeletal muscle involvement has been reported (Friedrich, Wilding et al. 2012). There are five isoforms of FHL1, each expressed in different amounts in heart and skeletal muscle. The tryptic peptide, AIVAGDQNVEYK, used in the MRM-MS validation assay, is present in all isoforms.

By differential expression analysis in 17 patients, FHL1 was found to be 1.49-fold higher in HCM hearts than controls ($p < 0.05$) with no difference between genetic subtypes (Figure 6.6a). This finding was validated in the targeted MRM-MS assay (Figure 6.6b). Levels were not significantly different between the AS and control groups.

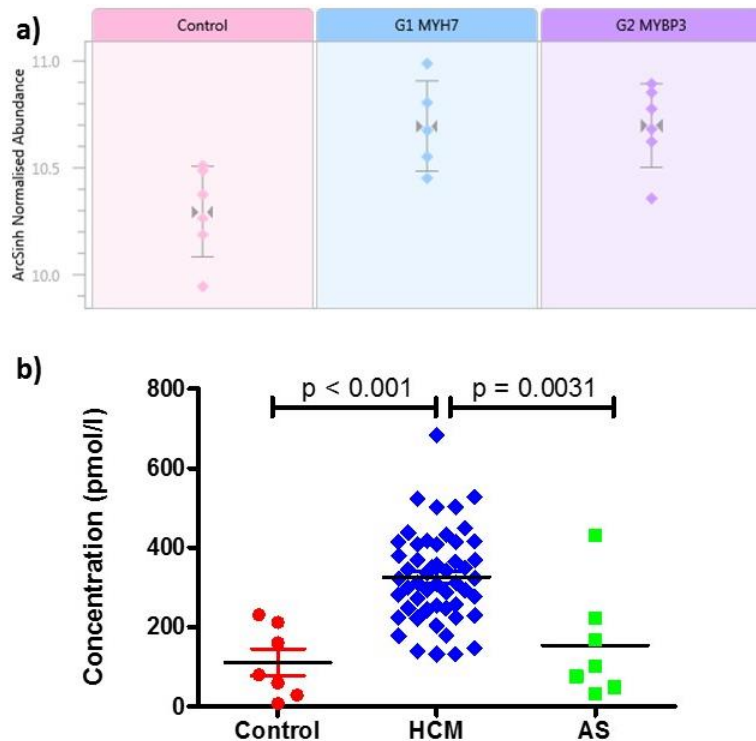


Figure 6.6 FHL1 expression in HCM

a) Average normalised abundances of FHL1 in the discovery proteomic study with Control versus HCM, including genetic subgroups, G1(MYH7) and G2 (MYBPC3), b) Scatter plot illustrating the concentration of FHL1 peptide analysed by UPLC-MS/MS in controls, hypertrophic cardiomyopathy (HCM) and aortic stenosis (AS).

The reason for increased FHL1 expression in HCM is not known. FHL1 is strongly expressed in the developing outflow tract and to a lesser extent in myocardium. In mouse models of dilated and hypertrophic cardiomyopathy, cardiac ventricular expression of FHL1, but not of related proteins FHL2 or FHL3, is upregulated (Chu, Ruiz-Lozano et al. 2000). The FHL1 protein has also been found to be significantly upregulated in human and animal models of pulmonary hypertension where its

expression is dependent on hypoxia-inducible transcription factor and co-immunoprecipitation experiments identified Talin-1 as a novel interacting protein (Kwapiszewska, Wygrecka et al. 2008). Prior to this the only identified interaction was between FHL1 and slow type skeletal myosin-binding protein C, where it was found to contribute to sarcomere assembly (McGrath, Cottle et al. 2006). Recently a transcriptomic study has suggested FHL1 may be a genetic modifier in HCM (Christodoulou, Wakimoto et al. 2014). And because FHL1 is an X chromosome encoded gene, it may explain in part why adverse outcomes are more common in men with HCM and individuals with more severe LV hypertrophy. The mechanism and role of increased FHL1 expression in HCM requires further investigation.

6.4.3 Desmin

Desmin (P17661) is a 53 kDa intermediate filament protein expressed in both cardiac, skeletal and smooth muscle; it was first described in 1976 (Lazarides and Hubbard 1976). It is normally located at Z-bands and interacts with myofibrillar and intercalated disc proteins. Desmin is believed to contribute to the embryogenesis, structural integrity and regeneration of muscle cells. It is prominent at neuromuscular junctions and decreases in amount with fibre maturation. The first knockout mouse was studied in 1996 (Agbulut, Li et al. 1996). Mutations in the desmin gene, *DES*, are associated with variable diseases including isolated skeletal myopathies and pure cardiac phenotypes (including dilated, restrictive and hypertrophic cardiomyopathy), cardiac conduction disease, and combinations of these disorders (Goldfarb, Park et al. 1998)

In HCM, desmin has been shown to abnormally localize to areas of disarray (D'Amati, Kahn et al. 1992) and has been described as a possible immunohistochemical marker for both feline and human hypertrophic cardiomyopathy (Francalanci, Gallo et al. 1995, Hayman, Une et al. 2000). The hypothesis being that, defective sarcomere formation and myofibrillar disarray is a secondary phenomenon regardless of the specific mutation. Altered distribution of desmin filaments seems to be specific to myofibre disarray in HCM, because it was

not found in myocardial samples from transplant recipients or patients with Fallot's tetralogy despite areas of disarray (Francalanci, Gallo et al. 1995).

By differential expression analysis we found desmin was higher in HCM than control during the pooled sample pilot study but not the formal discovery study (Table 6.4). 182 peptides were identified by LC-MS of which 121 were unique to desmin. One abundant peptide (FLEQQNAALAAEVNR, amino acid position 128 to 142) was selected for validation. It is located in the 1st coil of the rod region where no natural variants are currently reported. On targeted analysis, we found the concentration was higher in HCM patients compared with controls and patients with AS (Figure 6.7).

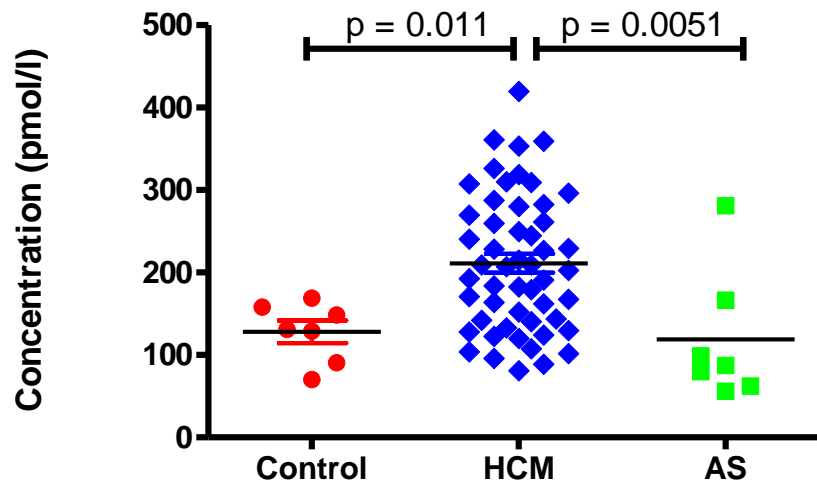


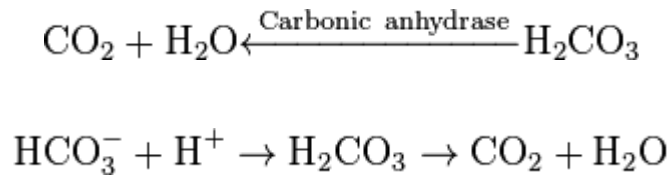
Figure 6.7 Desmin expression in HCM

Scatter plot illustrating the concentration of desmin peptide analysed by UPLC-MS/MS in controls, hypertrophic cardiomyopathy (HCM) and aortic stenosis (AS).

The higher level of desmin peptide identified by the targeted assay was consistent with the initial pooled discovery study (n=3) but not the formal discovery study (n=17). This may be related to the particular peptide selected for targeted analysis. A second and third peptide would be required to be certain that overall protein levels were increased.

6.4.4 Carbonic Anhydrase III

Carbonic anhydrase III (CA III, P07451) is part of a family of enzymes, the carbonic anhydrases whose primary function is to interconvert carbon dioxide and bicarbonate:



The carbonic anhydrases participate in numerous biological processes, such as acid-base balance, respiration, carbon dioxide and ion transport, bone resorption, ureagenesis, gluconeogenesis, body fluid generation, and lipogenesis (Sly and Hu 1995). In addition to their role in normal physiological processes, they are implicated in many pathological processes e.g. carbonic anhydrase activity in tumours contributes to acidification of the extracellular space, which has been linked to the malignant behaviour of cancer cells (Supuran and Scozzafava 2007).

6.4.4.1 Carbonic anhydrase isoforms

The mammalian alpha carbonic anhydrase (CA) gene family is characterized by 16 different isoforms, 13 of which (CA I, II, III, IV, VA, VB, VI, VII, IX, XII, XIII, XIV, and XV) are enzymatically active, while the other three, the CA-related proteins (CARPs) VIII, X, and XI, lack CA activity. CA XV is not expressed in primates, therefore, there are only 15 isoforms expressed in humans. The presence of carbonic anhydrase in heart muscle is well established with cardiac ventricular myocytes expressing different isoforms in the cytosol (CA II), mitochondria (CA V), sarcolemma (CA IV, CA XIV) and sarcoplasmic reticulum membrane (CA IV, CA IX, CA XIV) (Swenson 1997, Scheibe, Gros et al. 2006).

Human skeletal muscle in contrast is known to contain large amounts of CA III in the cytoplasm and serum levels have been evaluated as a potential diagnostic marker for rhabdomyolysis and neuromuscular diseases (Vaananen, Takala et al. 1988, Syrjala, Vuori et al. 1990). Because CA III has high specificity for skeletal muscle, a rise in serum myoglobin/CA III ratio was one of the first biomarkers investigated for acute

myocardial infarction and was shown to be useful in predicting response to thrombolytic therapy (Vaananen, Syrjala et al. 1990, Beuerle, Azzazy et al. 2000, Vuotikka, Uusimaa et al. 2003). Trace levels of CA III have been detected in human smooth muscle, heart, and lung (Jeffery and Carter 1980).

6.4.4.2 CA III expression in hypertrophic cardiomyopathy

Six unique tryptic peptides were identified by LC-MS that correlated with CA III (Table 6.5). One marker peptide (SLLSSAENEPPVPLVSNWRPPQPINNR) was selected for validation using MRM-MS corresponding to amino acid position 227 to 253. The spectrum of this peptide is shown in Figure 6.8.

Sequence	Score	Hits	Mass	Charge	ANA Control	ANA HCM
DGIAVIGIFLK	7.1	1	1144.7187	2	63.7	176.71
FDPSCLFPACR	7.07	3	1368.645	2	50.52	436.5
FDPSCLFPACR	7.81	2	1368.6556	2	82.82	698.27
FDPSCLFPACR	6.38	1	1368.6934	2	102.03	173.57
FDPSCLFPACR	7.41	3	1368.6382	2	223.8	825.14
HDPSLQPWSVSYDGGSAK	7.74	2	1929.91	2	11.5	46.23
IGHENGEFQIFLDALDK	7.67	2	1944.9011	3	6.06	33.78
SLLSSAENEPPVPLVSNWRPPQPINNR	7.76	4	3010.4368	3	3.61	119.99
SLLSSAENEPPVPLVSNWRPPQPINNR	7.81	6	3010.4212	3	38.77	681.62
YAAELHLVHWNP	7.13	2	1576.9366	3	23.84	75.27
YAAELHLVHWNP	7.48	2	1577.0088	3	29.01	134.28

Table 6.5 Unique peptides identified by LC-MS that correlated with CA III
ANA average normalised abundance

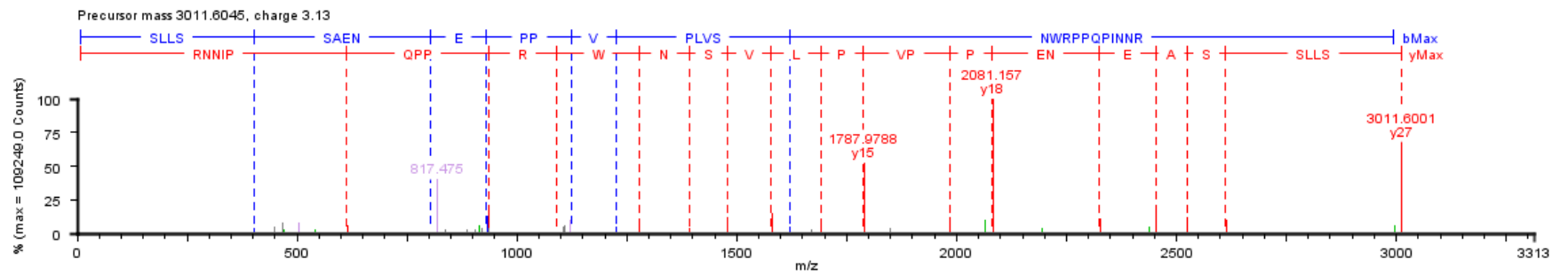


Figure 6.8 Spectrum of peptide marker for carbonic anhydrase III

By global proteomic profiling CA III, was found to be upregulated 3.8-fold in HCM hearts compared with controls ($p < 0.01$). Highest levels were in the *MYH7* sub-group although the difference between the two sub-groups did not reach statistical significance (Figure 6.9a). Using the MRM-MS assay CA III was confirmed to be significantly elevated in heart samples from 51 patients with HCM (108 ± 129 pmol/l) compared with 7 controls (21 ± 7 pmol/l) (Figure 6.9b).

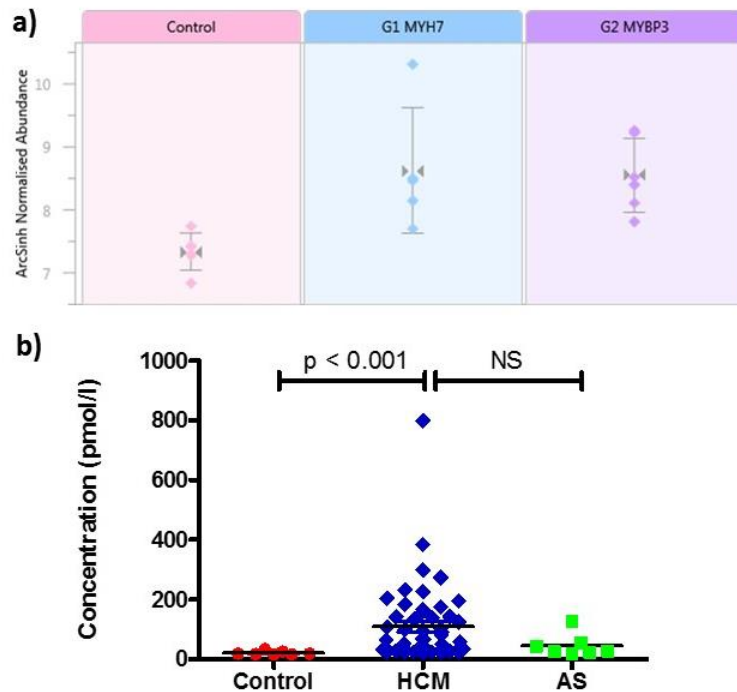


Figure 6.9 Carbonic anhydrase III expression in HCM

a) Average normalised abundances of CA III in the discovery proteomic study with Control versus HCM, including genetic subgroups, G1(*MYH7*) and G2 (*MYBPC3*), b) Scatter plot illustrating the concentration of CA III peptide analysed by UPLC-MS/MS in controls, hypertrophic cardiomyopathy (HCM) and aortic stenosis (AS).

The high expression of CA III in the HCM heart is a novel finding, as this isoform was previously reported to be at low levels in the myocardium (Jeffery, Edwards et al. 1980, Harju, Botorabi et al. 2013). Whether this is specific to HCM or a response to cardiac hypertrophy in general, remains to be determined (the difference in tissue levels of CA III between control and AS samples did not reach statistical

significance). The elevated levels of CA III observed in the present study could represent alterations in circulating (e.g. in red blood cell), rather than myocardial CA III. Immunohistochemistry was therefore performed to determine the tissue distribution of CA III, using skeletal muscle as a control (Chapter 8). The clinical correlates and relevance of CA III levels in HCM are discussed in Chapter 8.

6.4.5 Lumican

Lumican (LUM, P51884) is a 38 kDa protein which belongs to the small leucine-rich proteoglycan family. It either exists as a glycoprotein lacking keratan sulfate (adult cartilage) or as a proteoglycan (immature form). It is most abundant in the cornea, where it regulates collagen fibril assembly. In the discovery study, lumican was observed to be 1.8-fold higher in the hearts of patients with HCM compared with controls ($p < 0.05$). Mean level was higher in the *MYPBC3* sub-group although the range was greater (Figure 6.11a). Because this was a novel finding and was considered to be relevant to disease mechanisms, two peptides were chosen for inclusion in the validation study. Based on ESI-QToF analyses the peptides m/z 835.12 (SLEYLDLSFNQIAR) and 1268.69 (NIPTVNENLENYYLEVNQLEK) produced the largest response and optimum product ion scans and were therefore selected as the “marker” peptides. There was a clear linear correlation between both peptides ($r^2 = 0.96$, $p < 0.01$) indicating they are good markers for whole protein (Figure 6.10).

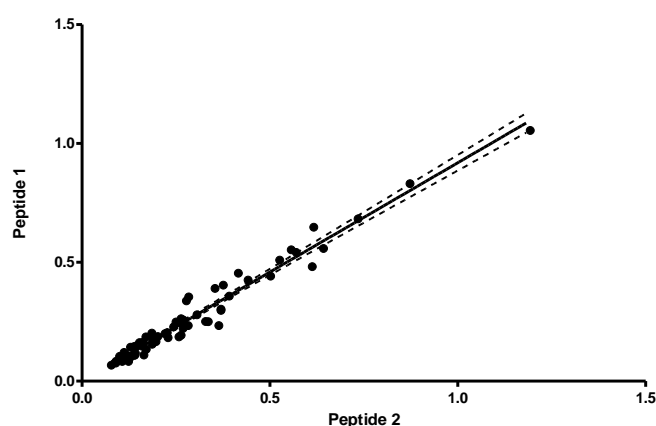


Figure 6.10 Correlation between two unique peptide markers for lumican

Scatter plot showing the response of peptide 1 (NIPTVNENLENYYLEVNQLEK) and peptide 2 (SLEYLDLSFNQIAR), relative to the internal standard (GF-IS) analysed by UPLC-MS/MS in the study cohort. 95% confidence intervals. Equation $y = 0.9171x - 0.0006$, $R^2 = 0.96$.

Using the MRM-MS assay lumican (NIPTVNENLENYYLEVNQLEK) was confirmed to be significantly increased in heart samples from 51 patients with HCM (283 ± 176 pmol/l) compared with 7 controls (133 ± 25 pmol/l, $p < 0.01$) and also increased in the AS samples (544 ± 311 pmol/l, $p < 0.05$) compared with controls (Figure 6.11b).

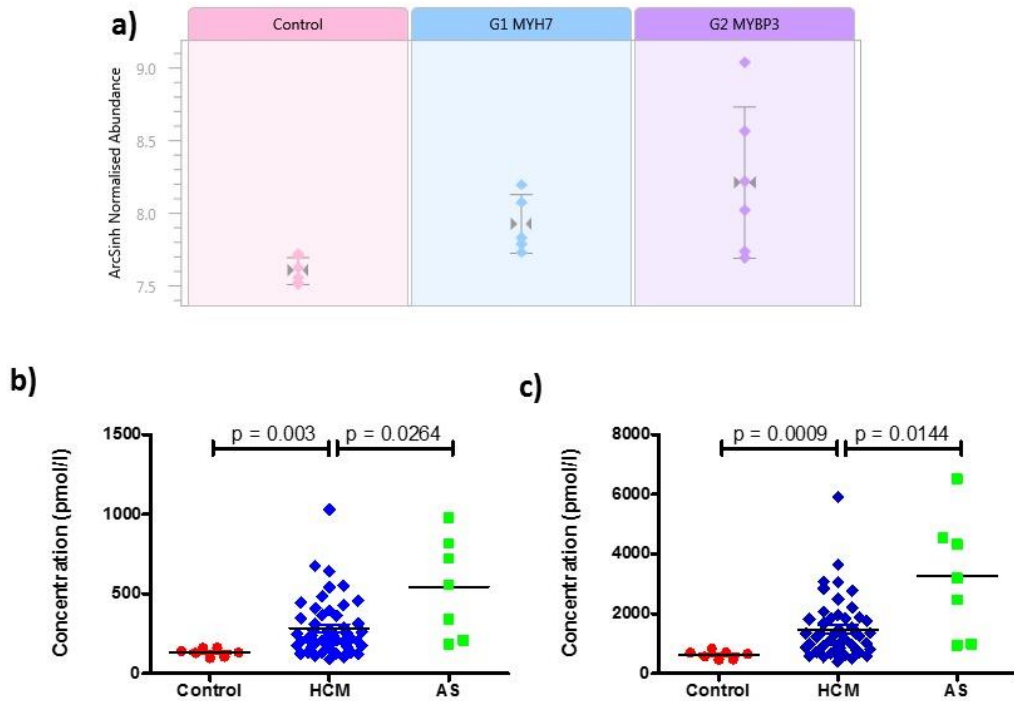
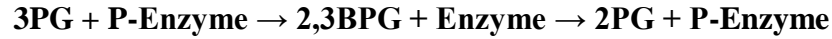


Figure 6.11 Lumican expression in HCM

a) Average normalised abundances of LUM in the discovery proteomic study with Control versus HCM, including genetic subgroups, G1(MYH7) and G2 (MYBPC3). Scatter plots illustrating the concentration of two lumican peptides, b) NIPTVNENLENYYLEVNQLEK and c) SLEYLDLSFNQIAR, analysed by UPLC-MS/MS in controls, hypertrophic cardiomyopathy (HCM) and aortic stenosis (AS).

6.4.6 Phosphoglycerate mutase

Phosphoglycerate mutase (PGM) is an isomerase enzyme that catalyses one of the terminal steps in glycolysis. In a two-step reaction, it transfers a phosphate group (PO_3^{2-}) from the C-3 carbon of 3-phosphoglycerate (3PG) to the C-2 carbon forming 2-phosphoglycerate (2PG) via a 2,3-bisphosphoglycerate (2,3BPG) intermediate:



The enzyme exists primarily as a dimer of two closely related sub-units of 32kDa: m-type (Phosphoglycerate mutase 2, PGAM2, P15259) which is the muscle-derived form and b-type (Phosphoglycerate mutase 1, PGAM1, P18669) which was originally isolated from brain. These create three isozymes: the mm-type found mainly in smooth muscle, the mb-type found in cardiac and skeletal muscle and the bb-type found in the rest of tissues. The peptide chosen for validation (VLIAAHGNSLR) exists in both sub-units.

By differential expression analysis in the pooled samples we found PGAM2 was lower in HCM with no difference in PGAM1 levels (Table 6.4). This was consistent with the absolute quantification data using PLGS but not the relative quantification data using Progenesis in the discovery sample set (Section 5.2.3.1). On targeted analysis, we confirmed the concentration was significantly lower in both the HCM and AS patients compared with controls (Figure 6.12).

Phosphoglycerate mutase has a small positive Gibbs free energy and this reaction proceeds easily in both directions. Since it is a reversible reaction, it is not the site of major regulation of the glycolytic pathway. In clinical practice, PGAM deficiency is a rare disorder that leads to storage of glycogen with prominent tubular aggregates seen on muscle biopsy (Salameh, Goyal et al. 2013). Detection of abnormal enzyme activity is usually performed using a calorimetric assay and this would have enabled further validation in this study.

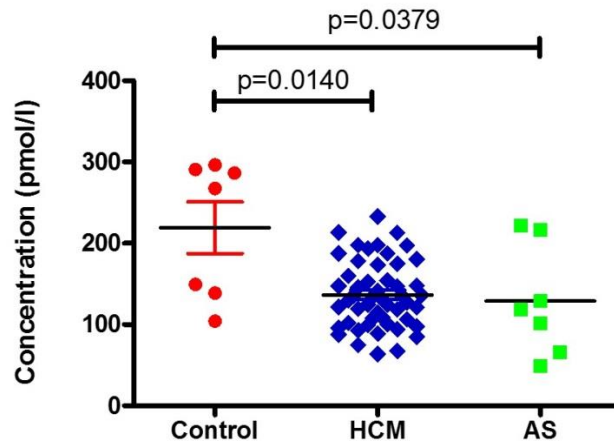


Figure 6.12 Phosphoglycerate mutase expression in HCM

Scatter plot illustrating the concentration of PGAM peptide analysed by UPLC-MS/MS in in controls, hypertrophic cardiomyopathy (HCM) and aortic stenosis (AS).

There are only a few reports in the literature of PGAM activity and cardiomyopathy. In a rapid pacing canine model of DCM, phosphoglycerate mutase was identified by 2D electrophoresis to be increased in heart failure (Heinke, Wheeler et al. 1999). Total PGAM activity has also been shown to be increased in serum following acute myocardial infarction (Durany, Carballo et al. 1996) and the early stages of ischaemia in donor hearts retrieved for transplantation (Li, Li et al. 2012). The relevance of our findings is not clear but supports the notion that energy production is impaired. Corresponding myocardial and serum levels would be informative.

6.4.7 Fructose-bisphosphate aldolase A,

Fructose bisphosphonate, commonly known as aldolase, is an enzyme required for carbohydrate metabolism. There are three isoenzymes in humans: aldolase A is preferentially expressed in muscle and red blood cells; aldolase B is expressed in the liver and aldolase C is expressed in the brain. The enzyme is involved in gluconeogenesis in the liver, and glycolysis in muscle where it catalyses the 4th step of glycolysis pathway:



Aldolase A is a well-characterised key enzyme of the glycolysis pathway (Penhoet, Kochman et al. 1967). It is widely recognised that there is a reappearance of the fetal metabolic pattern in the hypertrophied heart with an increased reliance on glucose and an overall reduced oxidative metabolism (Kolwicz and Tian 2011).

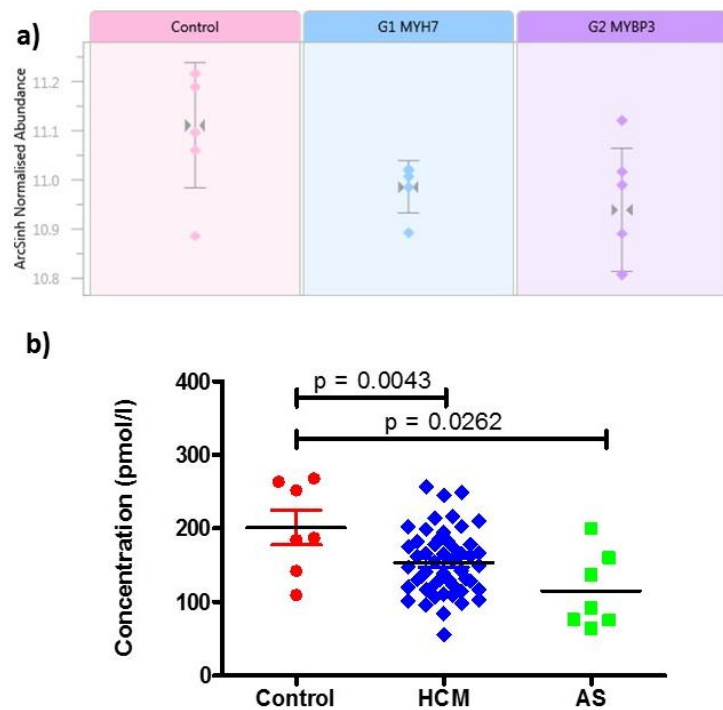


Figure 6.13 Aldolase expression in HCM

a) Average normalised abundances of ALDOA in the discovery proteomic study with Control versus HCM, including genetic subgroups, G1(MYH7) and G2 (MYBPC3), b) Scatter plot illustrating the concentration of ALDOA peptide analysed by UPLC-MS/MS in controls, hypertrophic cardiomyopathy (HCM) and aortic stenosis (AS).

By global proteomic profiling aldolase A (ALDOA, P04075) was found to be downregulated in HCM hearts compared with controls. There was no significant difference between the two genetic sub-groups (Figure 6.13a). Using the MRM-MS assay ALDOA was confirmed to be significantly reduced in heart samples from 51 patients with HCM (154 ± 44 pmol/l) compared with 7 controls (201 ± 62 pmol/l, $p = 0.0043$) and further reduced in the AS samples (Figure 6.13).

Our finding of reduced ALDOA appears contrary to that. However alterations in expression or capacity of glycolytic enzymes do not consistently coincide with increased glycolysis suggesting that the altered glycolytic flux is attributable to altered regulation rather than the expression of the specific glycolytic enzymes (Nascimben, Ingwall et al. 2004). Functional assays would be necessary to confirm this. Supporting evidence for upregulation of glycolysis in HCM is the evidence of downregulation of proteins involved in fatty acid oxidation and mitochondrial oxidative phosphorylation discussed earlier in Section 5.3. We were however unable to validate by targeted MRM-MS the reduction in very long-chain specific acylCoA dehydrogenase mitochondrial (ACDAV, P49748) observed in the discovery study.

6.4.8 Creatine kinase M-type

Creatine kinase (*KCM*, P06732) is a cytoplasmic enzyme involved in cellular energy homeostasis. It has an important role in tissues with high fluctuating energy demands, such as skeletal muscle and the heart. Cytosolic creatinine kinase (CK) enzymes consist of two subunits, B (brain type) and M (muscle type). There are, therefore, three different isoenzymes: CK-MM, CK-BB and CK-MB. In the heart, the homodimer CK-MM and the heterodimer CK-MB are expressed. Creatinine kinase reversibly catalyses the transfer of phosphate between ATP and various phosphogens (e.g. creatine phosphate):



The CK-M chain is made up of 381 amino acids. The most abundant peptide (GTGGVDTAAVGGSVFDVSNADR) identified in the discovery study was used for validation. This corresponded with position 321-341 where no natural variants are

reported. Although the difference was not statistically significant, CK-M appeared to be downregulated in both genetic sub-types of HCM compared with controls (Figure 6.14a). Using the MRM-MS assay KC-M was confirmed to be significantly reduced in heart samples from 51 patients with HCM (458 ± 109 pmol/l) compared with 7 controls (738 ± 154 pmol/l, $p < 0.001$) and also reduced in the AS samples (476 ± 176 pmol/l, $p = 0.018$) compared with controls (Figure 6.14b).

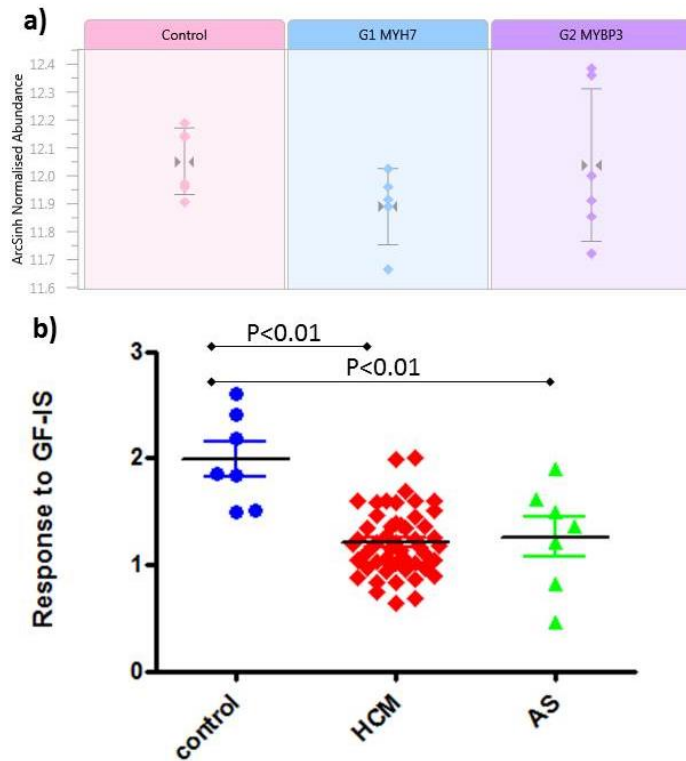


Figure 6.14 Creatinine kinase M-type in HCM

a) Average normalised abundances of CK-M in the discovery proteomic study with Control versus HCM, including genetic subgroups, G1(MYH7) and G2 (MYBPC3), b) Scatter plot illustrating the concentration of CK-M peptide analysed by UPLC-MS/MS in controls, hypertrophic cardiomyopathy (HCM) and aortic stenosis (AS).

Down-regulation of creatinine kinase activity and myocardial creatine levels is a universal finding in chronic heart failure and the degree of impairment has been shown to be an good prognostic indicator in patients (Neubauer, Horn et al. 1997, Neubauer 2007). The reduction of CK-M observed in HCM and AS in this study is consistent with the previous reports of impaired high-energy phosphate metabolism

in these diseases (Neubauer, Horn et al. 1997, Kalsi, Smolenski et al. 1999). In the discovery study there was also a significant reduction of ADP ATP translocase 1 (ADT1, P12235) in HCM compared with controls (1.86 fold lower, $p < 0.001$) and many of the sub-units of ATP synthase (Table 5.2) which supports the notion of an overall decrease in energy flux within the cell, i.e. down-regulation of cytoplasmic creatine kinase causes a decline in ATP transfer into the mitochondria and thus a reduction in energy delivered to the myofibrils. Over-expression of CK in mouse heart has been found to protect against heart failure and improve survival (Akki, Su et al. 2012). The role of metabolic modulation and augmentation of the creatinine kinase system in patients with heart disease is gaining increasing interest as a therapeutic target.

6.4.9 Gelsolin

Gelsolin (GELS, P06396) is a calcium dependent regulator of the actin cytoskeleton. It is 82kDa protein which binds to and regulates actin filament assembly and disassembly (Nag, Larsson et al. 2013). Yang et al found gelsolin gene, *GSN* expression was up-regulated in two end-stage failing hearts (dilated cardiomyopathy) compared with two controls (Yang, Moravec et al. 2000).

In this study, using Protein Lynx analysis, gelsolin was identified in 6 of the 11 HCM samples but was not identified in any of the control samples (Section 5.2.1.2). Using the differential expression analysis tool, Progenesis, gelsolin was found to be 1.27-fold higher in HCM hearts than controls but this did not reach statistical significance ($p = 0.36$). There appeared to be certain individual outliers within the results (Figure 6.15a) hence we sought to include a gelsolin specific peptide in the MRM-MS targeted validation assay.

There was no significant difference in gelsolin levels between 7 controls (38 ± 17 pmol/l) and 51 HCM samples (35 ± 31 pmol/l), however levels were significantly higher in the AS samples (108 ± 58 , $p < 0.005$) than controls. Individual outliers with high gelsolin levels were observed within the HCM group (Figure 6.15b).

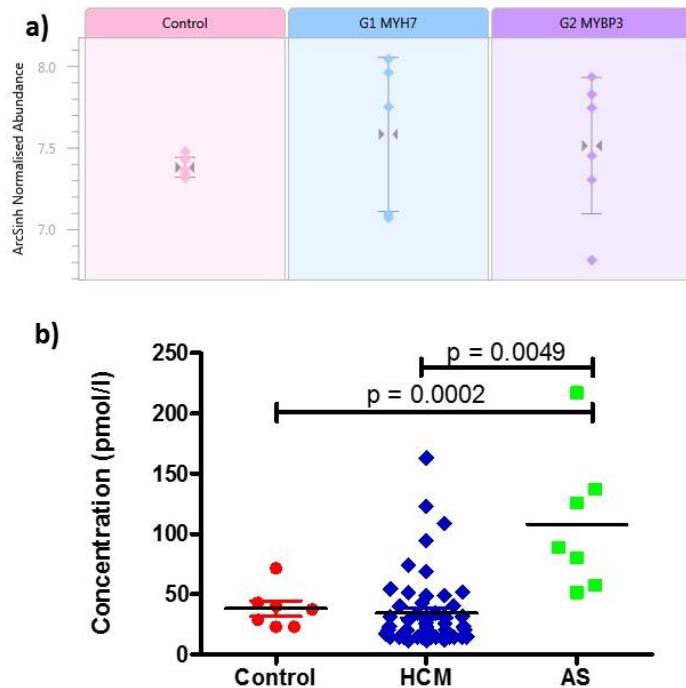


Figure 6.15 Gelsolin expression in HCM

a) Average normalised abundances of GELS in the discovery proteomic study with Control versus HCM, including genetic subgroups, G1 (MYH7) and G2 (MYBPC3), b) Scatter plot illustrating the concentration of GELS peptide analysed by UPLC-MS/MS in controls, hypertrophic cardiomyopathy (HCM) and aortic stenosis (AS).

Gelsolin has previously been shown to be increased during myocardial ischaemia by proteomic analysis on effluents from perfused human hearts of donors at different ischemic times (Li, Li et al. 2012). Given that end-stage heart failure can result from a diverse range of cardiac conditions including cardiomyopathy and valve disease, we postulated that patients with elevated gelsolin levels may be more likely to develop heart failure.

6.4.10 Transthyretin

Transthyretin (TTHY, P02766) formally known as pre-albumin is a 127-amino acid 56kDa transport protein which transports thyroxine (T4) and retinol-binding protein. It is primarily secreted by the liver into the blood, and by the choroid plexus into the cerebrospinal fluid. Senile systemic amyloidosis is a disease which occurs when wild

type TTHY dissociates, misfolds and aggregates; it is estimated to affect 25% of the population over 80 years old (Cornwell, Murdoch et al. 1983). Familial amyloid can also occur and is usually due to point mutations in the transthyretin gene, *TTR* (Ruberg and Berk 2012). Deposition of TTHY amyloid is typically extracellular but deposits are also observed within the cardiomyocytes of the heart. Amyloid cardiomyopathy can result from either senile or genetic TTHY disease; making a clinical diagnosis is challenging because many features overlap with HCM or heart failure with preserved ejection fraction (HFpEF).

By differential expression analysis we found TTHY was 1.47-fold higher in HCM hearts than controls (p=0.04). The highest levels were in the sub group of patients with *MYBPC3* mutation (1.74 fold higher than controls, p=0.02) (Figure 6.16a). Thirteen peptides were identified by LC-MS of which twelve were unique to transthyretin and the two most abundant were selected for inclusion in the validation assay (Table 6.2). On targeted MRM-MS analysis we detected higher levels of the TTHY peptides in some individuals with HCM and AS but overall it did not reach statistical significance (Figure 6.16b).

Transthyretin has previously been shown to be upregulated in failing hearts due to dilated cardiomyopathy (Li, Rong et al. 2012). Heart failure caused by wild-type transthyretin amyloidosis (ATTRwt) is increasingly recognised as a cause of morbidity and mortality in the aging population (Connors, Sam et al. 2016) . The relevance in HCM warrants further investigation. Firstly, it would important to determine whether the TTHY is located within the cardiomyocyte, extracellular matrix or blood pool and secondly to understand whether it is misfolded and associated with amyloid fibril formation. This could be confirmed histologically with Congo red staining and immunohistochemistry.

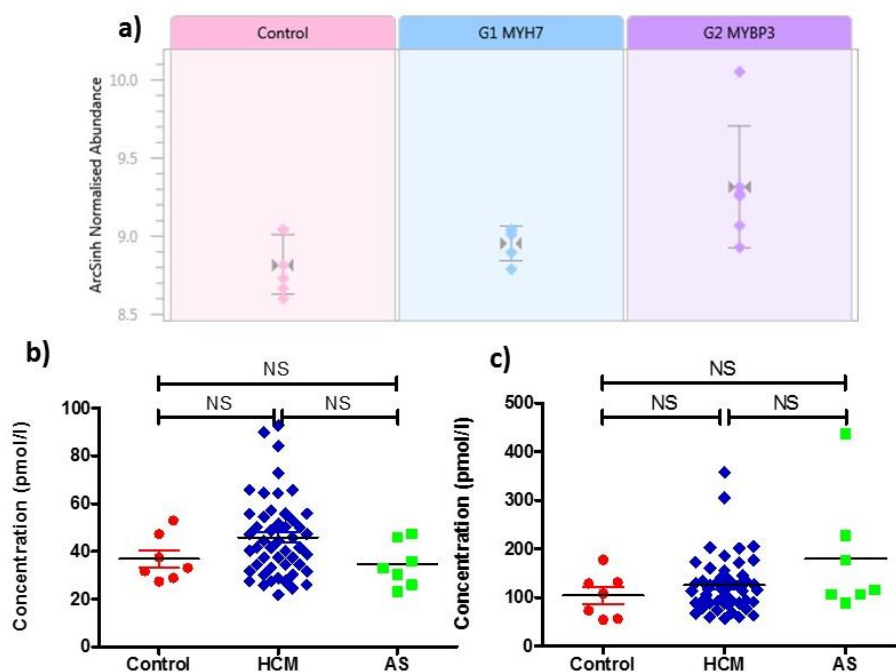


Figure 6.16 Transthyretin expression in HCM

a) Average normalised abundances of transthyretin in the discovery proteomic study with Control versus HCM, including genetic subgroups, G1(MYH7) and G2 (MYBPC3). Scatter plots illustrating the concentration of two peptides for transthyretin:

b) YTIAALLSPYSYSTTAVVTNPK and c) GSPAINVAVHVFR, analysed by UPLC-MS/MS in controls, hypertrophic cardiomyopathy (HCM) and aortic stenosis (AS).

6.5 Conclusion

We have developed a high throughput UPLC-MS/MS assay for confirmation and validation of peptides identified through the ESI-QToF discovery experiment. The targeted assay was run in 10 minutes which included sample loading, cleaning, and reconditioning. This enabled 65 samples to be processed including 51 samples from patients with HCM. This larger sample size enabled us to subsequently explore the potential clinical utility of novel disease biomarkers.

7 Genomics and Proteomics

Genetic testing is recommended in patients fulfilling diagnostic criteria for hypertrophic cardiomyopathy (HCM) primarily to enable cascade screening of their relatives (Authors/Task Force, Elliott et al. 2014). Mutations in Myosin 7, *MYH7* (also known as β -myosin heavy chain) and myosin binding protein C, *MYBPC3* account for nearly half of all HCM cases. There is considerable variation in clinical phenotypes caused by the same or similar mutations and despite genotype–phenotype studies it remains unclear to what extent genetic variants of HCM have a common mechanism of pathogenesis (Lopes, Syrris et al. 2015). This chapter focuses on understanding whether a genotype specific proteome exists and explores to what extent mutant proteins can be detected by mass spectrometry.

7.1 Genotype-phenotype relationship in HCM

7.1.1 Myosin 7

As the original disease-causing gene, *MYH7* has been studied extensively in families where disease is usually highly penetrant and associated with severe hypertrophy (Watkins, Rosenzweig et al. 1992). Over 1,000 mutations have been identified in different functional domains of the protein which may account for the some of the phenotypic diversity. Early on, the concept of ‘benign’ and ‘malignant’ mutations arose: mutations that were considered malignant altered the charge of the amino acid or were located at the head of the protein, particularly around the actin binding site (Woo, Rakowski et al. 2003). However, clinical screening and genotyping of larger cohorts have challenged these hypotheses. Today we recognise that the same *MYH7* mutation can result in variable disease penetrance and phenotypic severity. This suggests there are patient specific factors that modify the genetic substrate and contribute more to the clinical syndrome than the mutation alone. Transgenic murine models may not be representative of human disease because they express alpha myosin, not beta-myosin the predominant isoform in the heart.

7.1.1.1 Mechanism of disease

The majority of *MYH7* mutations are missense and are proposed to cause disease by a ‘poison peptide’ mechanism whereby the mutant protein is expressed in the heart and impairs or overrides the function of the wild-type protein (from the normal allele) leading to disease (Seidman and Seidman 2001). The mutated protein is not destroyed, rather incorporated into the sarcomere exerting a dominant negative effect. The notion of allelic imbalance was subsequently proposed to explain the variation between genotype and phenotype (Tripathi, Schultz et al. 2011). The relative abundance of wild-type and mutated myosin 7 can be accurately determined by mass spectrometry (Becker, Navarro-Lopez et al. 2007).

7.1.2 Myosin binding protein C

Having identified disease loci on chromosome 11, a candidate gene study confirmed that mutations in *MYBPC3* can also cause HCM (Watkins, Conner et al. 1995) and these mutations are now recognised as the commonest cause of familial HCM (Seidman and Seidman 2011). Initial clinical reports suggested *MYBPC3* mutations were associated with a later age of disease onset and a less malignant course (Charron, Dubourg et al. 1998). This is in part supported by contemporary studies that show a higher than expected incidence of *MYBPC3* mutations in the elderly, mild hypertrophy and good prognosis (Niimura, Patton et al. 2002). However, childhood and adolescent disease is still common and overall phenotype–genotype correlations are weak (Page, Kounas et al. 2012).

7.1.2.1 Mechanism of disease

Unlike the missense mutations seen in *MYH7*, the majority of *MYBPC3* mutations are intronic and predicted to cause aberrant splicing, leading to a frameshift and premature truncation of the protein (Marston, Copeland et al. 2009). The truncated peptides have never been identified in human heart tissue and instead of a poison peptide mechanism, haploinsufficiency is observed. A quantitative investigation of protein levels in myectomy samples from HCM patients with known *MYBPC3* mutations, showed that the full-length MyBPC protein level was 20–30% lower than in normal hearts or HCM hearts with mutations in other genes (Marston, Copeland et al. 2009). Although disease mechanisms are not fully elucidated, there is evidence of

increased calcium sensitivity and abnormal phosphorylation of the troponin complex (Marston, Copeland et al. 2012). More recently the concept of allelic imbalance has been applied to truncating mutations in *MYBPC3* related disease (Helms, Davis et al. 2014).

7.2 Exploration of a genotype specific proteome

The myocardial protein samples used in Chapter 5 were subject to further bioinformatics analysis according to the genetic sub-type. Five samples came from unrelated patients carrying a mutation in *MYH7* and six samples from unrelated patients carrying a mutation in *MYBPC3*. The specific mutations are listed in Table 4.1. Extracted peptide intensities were evaluated with Progenesis LC-MS (Version 3.0, Nonlinear Dynamics, Waters Corporation) to obtain label-free relative quantification from the raw data files.

7.2.1 Results of expression analysis

A total of 595 proteins were suitable for comparison between the two HCM groups. Twenty-six proteins showed more than 1.5-fold difference in expression using Progenesis LC-MS; fourteen showed greater expression in the *MYBPC3* group (Table 7.1) and twelve showed greater expression in the *MYH7* group (Table 7.2). The difference was statistically significant ($p < 0.05$) by ANOVA for only two proteins: haemoglobin subunit alpha (HBA, P69905) and haemoglobin subunit beta (HBB, P68871).

7.2.1.1 Haemoglobin concentration

The haemoglobin subunits alpha (HBA) and beta (HBA) were respectively, 1.60 and 1.6-fold higher ($p < 0.05$) in the *MYBPC3* group compared with *MYH7* group (Figure 7.1). Given that these are the two major sub-units of haemoglobin, the iron-containing oxygen-transport metalloprotein in the red blood cells, we can infer that the abundance of red blood cells between the two sample groups was different. There was a trend towards greater abundance of other plasma proteins i.e. alpha 2 HS glycoprotein and transthyretin in the *MYPBC3* group although differences were not statistically significant. It is possible that not only red blood cells but all plasma proteins were more abundant in this group. Whilst blood contamination is an

expected limitation of using whole heart homogenate we would have expected this to have occurred randomly across all samples. We did not anticipate finding differences between the two genetic sub-groups, particularly given that the tissue samples were harvested in PBS, washed thoroughly and vortexed during preparation for MS analysis using a standard protocol. Researchers were blind to genetic subtype at that time. This then raises the question of whether this truly reflects increased vascularity of myocardial tissue in the *MYBPC3* subgroup. Unfortunately, the small sample size makes it difficult to determine whether this might be significant for other plasma proteins.

Measurement of haemoglobin concentration is made routinely in patients prior to surgery. We therefore sought to correlate haemoglobin concentration in whole blood with tissue expression of the haemoglobin subunits. Interestingly haemoglobin concentration was also found to be lower in the *MYH7* group compared with the *MYBPC3* group (Figure 7.2) although it did not reach statistical significance ($p=0.29$). This may account for the differences found in tissue.

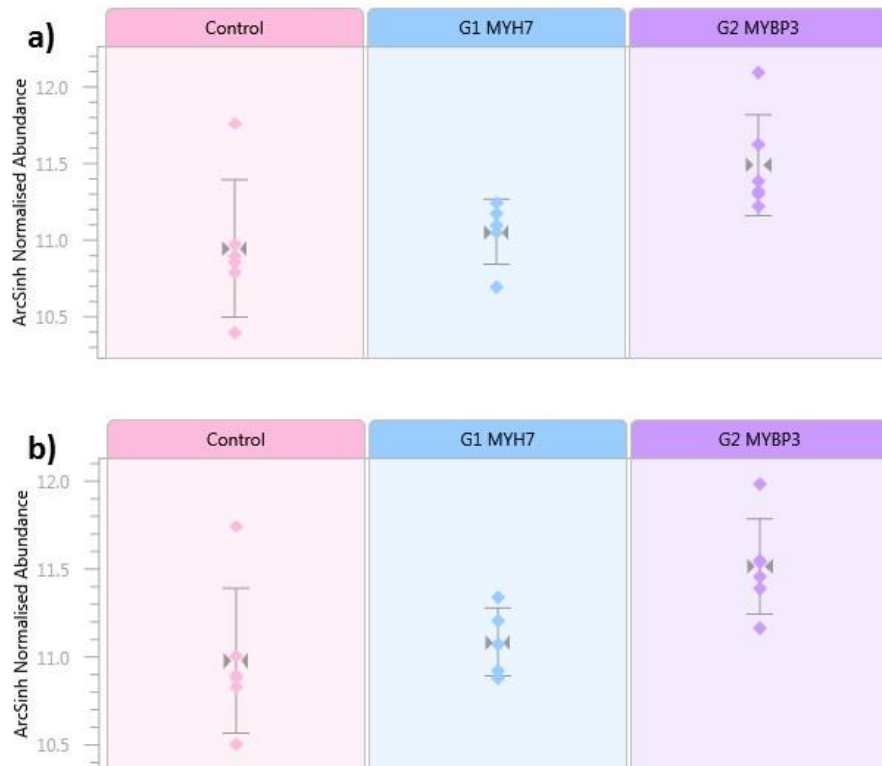


Figure 7.1 Expression of haemoglobin subunits in myocardial samples

Average normalised abundances of a) haemoglobin sub-unit alpha and b) haemoglobin sub-unit beta in the discovery proteomic study. Scatter plots are shown of control versus HCM, including genetic subgroups, G1 (MYH7) and G2 (MYBPC3).

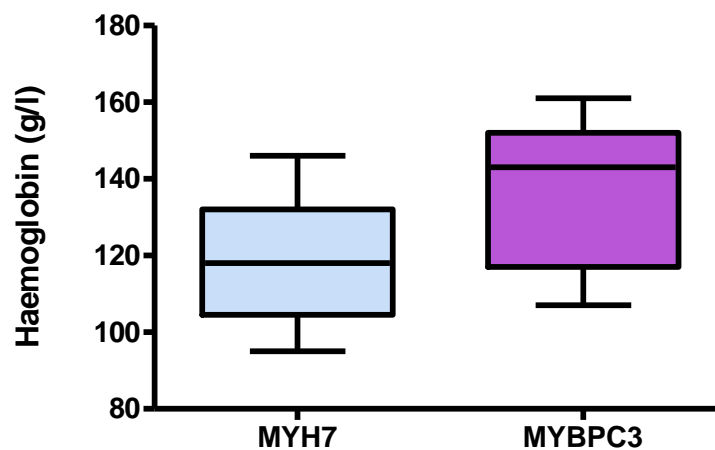


Figure 7.2 Haemoglobin concentration in the discovery cohort.

Box and whisker plot showing median haemoglobin concentration of whole blood with minimum and maximum values (whiskers) in g/L in the two genetic sub-groups of the discovery cohort ($n=11$).

Uniprot Accession	Peptides	Score	Anova (p)*	Fold	Description
PDE4C_HUMAN	1 (1)	5.6	0.28	6.61	cAMP specific 3 5 cyclic phosphodiesterase 4C
A1AG2_HUMAN	10 (2)	66.43	0.38	1.87	Alpha 1 acid glycoprotein 2
TIM9B_HUMAN	1 (1)	5.92	0.92	1.76	Mitochondrial import inner membrane translocase subunit Tim9B
ANGT_HUMAN	3 (2)	20.08	0.08	1.68	Angiotensinogen
K2C75_HUMAN	2 (1)	18.43	0.67	1.65	Keratin type II cytoskeletal 75
FETUA_HUMAN	1 (1)	6.81	0.35	1.64	Alpha 2 HS glycoprotein
HBA_HUMAN	56 (47)	208.08	0.03	1.6	Hemoglobin subunit alpha
TMTC2_HUMAN	1 (1)	6.68	0.68	1.6	Transmembrane and TPR repeat containing protein 2
HV304_HUMAN	1 (1)	8.16	0.36	1.58	Ig heavy chain V III region TIL
HBB_HUMAN	102 (34)	334.14	0.02	1.56	Hemoglobin subunit beta
TTHY_HUMAN	13 (12)	90.43	0.08	1.54	Transthyretin
TBB4A_HUMAN	16 (3)	98.01	0.19	1.53	Tubulin beta 4A chain
CAH1_HUMAN	3 (3)	12.32	0.1	1.52	Carbonic anhydrase 1
RASM_HUMAN	1 (1)	6.19	0.35	1.5	Ras related protein M Ras

Table 7.1 Proteins > 1.5 fold higher in patients with MYBPC3 mutation

The number of tryptic peptides is reported in the column headed “peptides” with the number of unique peptides for that protein shown in parentheses.

Uniprot Accession	Peptides	Score	Anova (p)*	Fold	Description
AFG32_HUMAN	1 (1)	6.9	0.35	2.58	AFG3 like protein 2
CUL1_HUMAN	1 (1)	6.18	0.12	2.48	Cullin 1
QCR8_HUMAN	2 (1)	14.82	0.16	2.15	Cytochrome b c1 complex subunit 8
ISK5_HUMAN	3 (1)	11.64	0.08	1.96	Serine protease inhibitor Kazal type 5
NDUA2_HUMAN	3 (3)	23.78	0.08	1.8	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 2
CFAB_HUMAN	1 (1)	6.14	0.24	1.69	Complement factor B
UBE2N_HUMAN	1 (1)	6.44	0.34	1.67	Ubiquitin conjugating enzyme E2 N
CMC1_HUMAN	5 (4)	34.6	0.2	1.62	Calcium binding mitochondrial carrier protein Aralar1
XIRP1_HUMAN	2 (1)	13.44	0.29	1.56	Xin actin binding repeat containing protein 1
PCDH1_HUMAN	1 (1)	12.86	0.28	1.55	Protocadherin 1
WIT1_HUMAN	1 (1)	11.7	0.26	1.52	Putative Wilms tumor upstream neighbor 1 gene protein
CAH3_HUMAN	7 (7)	45.61	0.89	1.51	Carbonic anhydrase 3

Table 7.2 Proteins > 1.5 fold higher in patients with MYH7 mutation

The number of tryptic peptides is reported in the column headed "peptides" with the number of unique peptides for that protein shown in parentheses.

7.3 Application of next generation sequencing

The past decade has seen notable advances in sequencing technology such that the number of genetic variants identified in patients far exceeds the ability of functional studies or co-segregation analyses to determine their effect. Next generation sequencing (NGS) has revolutionised our approach to genetic testing in clinical practice enabling a dramatic increase in throughput at reduced cost. Multigene testing is now feasible in HCM using these platforms which essentially sequence millions of short-DNA fragments in massively parallel arrays and then realign and map the short reads back to the reference genome. (Ho, Charron et al. 2015). Prediction of in silico pathogenicity is performed using various software platforms e.g. SIFT (Ng and Henikoff 2003) and PolyPhen (Adzhubei, Schmidt et al. 2010). The huge challenge is in understanding how variants of uncertain significance relate to the evolution of the clinical phenotype.

7.3.1.1 Genotyping in the validation cohort

All patients with HCM were counselled and offered genetic testing prior to undergoing myectomy. Thirty nine patients consented to be included in a high-throughput NGS study designed to analyse 41 genes implicated in inherited cardiac conditions (Lopes, Syrris et al. 2015). Six patients underwent clinical genetic testing examining sarcomeric proteins only. Six patients declined genetic testing.

Of the patients (n=45) who underwent testing, twenty-three (51 %) were found to carry a pathogenic mutation in one or more of the sarcomeric proteins. The mutations were in *MYH7* (n=14), *MYBPC3* (n=9), *TNNI3* (n=1) and *MYL2* (n=1). One patient had a mutation in both *MYH7* and *MYBPC3* and one patient had two mutations in the *MYH7* gene. Rare variants (defined as a minor allele frequency < 1%) in non sarcomeric protein genes were identified in 32 (82%) patients enrolled in the NGS study; these were in eleven genes known to be associated with inherited arrhythmias and dilated cardiomyopathies: *MYH6*, *CSRP3*, *RBM20*, *TTN*, *RYR2*, *PKP2*, *DSG2*, *DSP*, *TMEM43*, *TGF β 3* and *PNN*. The relevance of these individual variants was not explored further in the current proteomic study.

The targeted MRM-MS analysis (described in Chapter 6) of specific proteins was evaluated according to genetic sub-groups. Patients (n=45) were assigned to one of the following groups:

- a) mutation in *MYH7* (n=12)
- b) mutation in *MYBPC3* (n=8)
- c) no mutation identified (n=9)
- d) rare variants only identified (n=13).

Three patients were excluded from this analysis (one with double heterozygosity for pathogenic mutations in *MYH7* and *MYBPC3*, and the two patients with mutations, in *TNNI3* and *MYL2*). One way ANOVA was used with Bonferroni's multiple comparison test to determine differences between groups.

7.3.1.2 MRM-MS analysis by genotype

For most proteins, there was no statistically significant difference between the genetic sub-groups. Results for eight peptides are shown in Figure 7.3. Values for the control (normal) hearts are included for reference. The proteins involved in metabolism e.g. creatinine kinase M-type, phosphoglycerate mutase and aldolase, showed minimal difference in mean response between the groups. In contrast, there was a trend towards lower mean levels of four and a half LIM domains protein 1, carbonic anhydrase III and lumican in the sub-group (n=9) where “no mutation” was identified. This only reached statistical significance for lumican with the *MYBPC3* group (p=0.04). Confidence intervals were greater for these proteins implying there are factors more important than genotype which determine their expression. Overall, sample numbers were small and the study was unable to detect significant differences between the groups for most proteins. Nonetheless it supported the notion that final common pathways (e.g. metabolism) are common to all sub-types of HCM, and downstream of the effect of specific sarcomeric mutation.

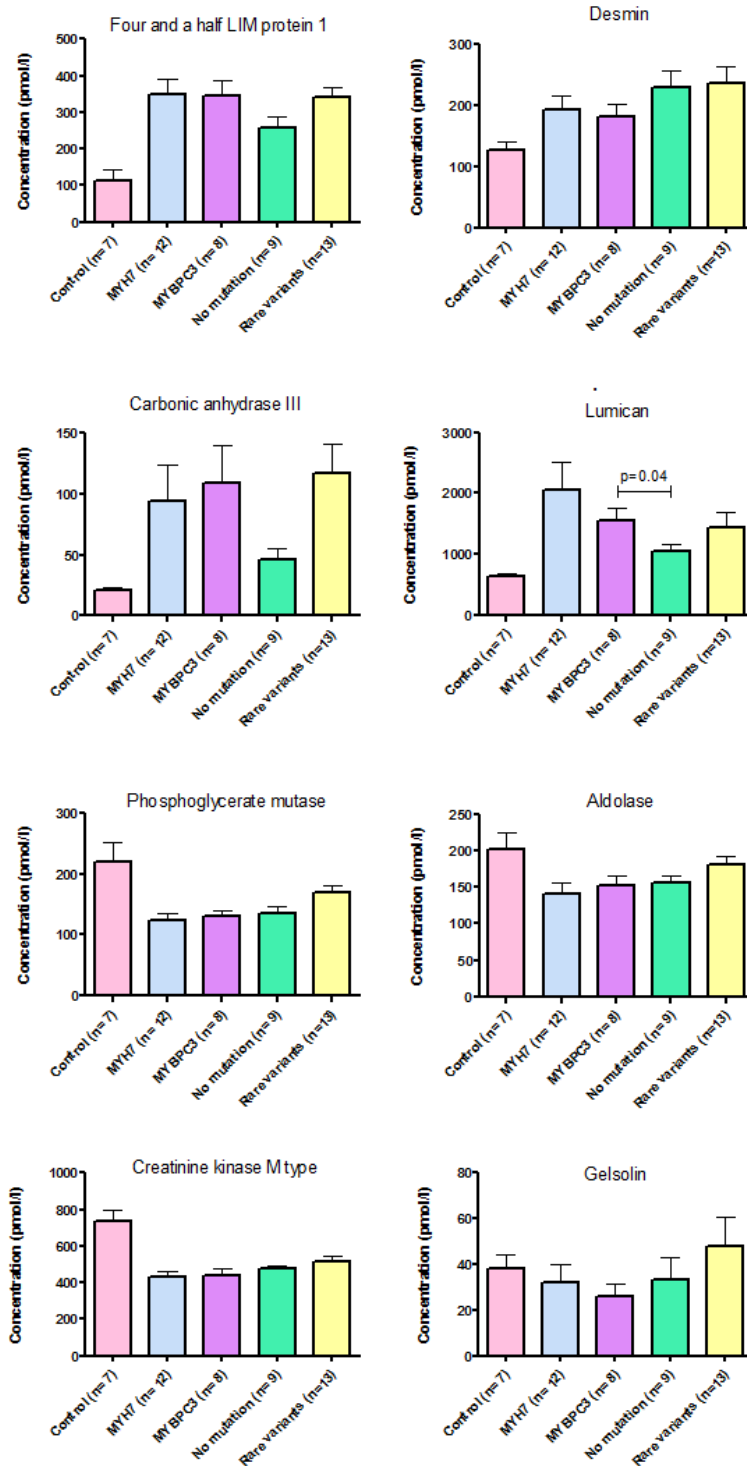


Figure 7.3 Protein expression according to genetic variants

Concentrations of eight peptides analysed by UPLC-MS/MS in myocardial samples from 7 controls and 42 patients with HCM, separated by genetic sub-groups. Bar charts show mean and SEM.

7.3.1.3 Myosin binding protein C levels

In the pilot and discovery studies there was a significant increase in MyBP-C levels between HCM and control (Table 6.4), however there was no significant difference in MyBP-C between the two genetic subgroups (Figure 7.5a). This was contrary to the previous reports of haploinsufficiency in the literature, however numbers were small and there was insufficient power to detect the expected difference of 20-30% between these groups (Marston, Copeland et al. 2012).

We therefore sought to include a unique peptide (VFSQNMVGFSDR) for MyBP-C in the validation study which is described in detail in Chapter 6. This peptide corresponds with amino acid position 1139 to 1150, where no natural variants have been reported. The location of the point mutations in the patient samples are shown alongside this peptide in Figure 7.4.

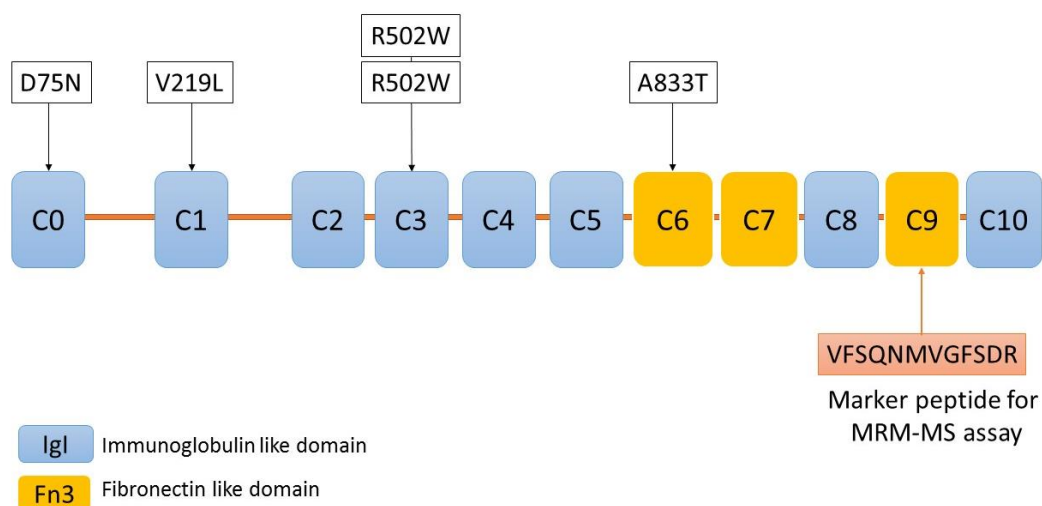


Figure 7.4 Myosin binding protein C secondary structure

Domain structure of MyBP-C showing the location of the point mutations detected in the myectomy samples (n=5). The splice site mutations (n=3) are not shown.

Using a targeted quantitative MRM-MS/MS technique we were able to demonstrate mean response of the peptide relative to the internal standard was 25 % lower in the eight patients (27 ± 8 pmol/l) carrying a known pathogenic mutation in *MYBPC3* compared with the twelve patients (35 ± 10 pmol/l, $p=0.07$) carrying a known

pathogenic mutation in *MYH7* (Figure 7.5b) and also lower than patients not found to carry a mutation (36 ± 12 pmol/l, $p=0.06$) although the difference did not reach statistical significance. Importantly there was no difference in My-BP-C peptide levels in the samples with a *MYBPC3* mutation compared with controls. This is consistent with the study by Helms et al but at variance with the study performed by Marston et al, both experiments conducted using immune blotting (Marston, Copeland et al. 2009, Helms, Davis et al. 2014). This variance in peptide abundance between the genetic sub-groups neither confirm nor refute a mechanism of haploinsufficiency as the primary mechanism of disease. To examine whether reduced amounts of mutant message contribute to the lower total MyBP-C protein content, wild-type/mutant *MYBPC3* mRNA ratios could be measured using real-time PCR with allele specific probes (Helms, Davis et al. 2014).

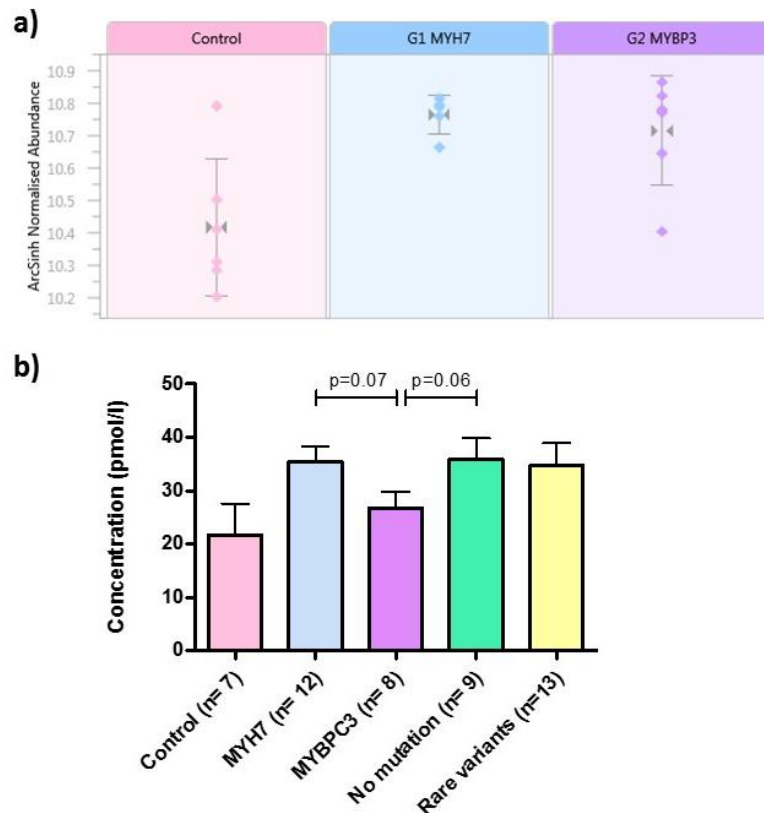


Figure 7.5 Myosin binding protein C expression in HCM

a) average normalised abundances of MyBP-C in the discovery proteomic study with Control versus HCM, including genetic subgroups, G1(*MYH7*) and G2 (*MYBPC3*). *b)* Bar chart showing concentration of MyBP-C peptide analysed by UPLC-MS/MS in 7 controls and 42 patients with HCM, separated by genetic sub-groups. Bar charts show mean and SEM.

7.4 Identification of mutant peptides by MS

Mass spectrometry (MS) allows the identification of protein sequences with a high degree of confidence from a very small quantity of sample. A point mutation within a protein usually leads to a change in molecular mass of the protein. Eighteen of the twenty natural amino acid residues have distinctive molecular masses (except for the isomers, Leucine, Leu and Isoleucine, Ile). The mass difference ranges from 0.0364 Da (Glutamine/Lysine, Q/K) to 129.0578 Da (Glycine/Tryptophan, G/W) and all substitutions between these amino acids can theoretically be detected by mass spectrometry. The accuracy of molecular mass determination is critical for successful characterisation of the mutation.

7.4.1 Methods

To explore the application of LC-MS to detect specific mutations in proteins the human protein database UniProt/Swiss-Prot website (<http://www.uniprot.org>) was supplemented with the predicted mutant protein sequences in addition to the wildtype protein sequence. The identity of peptides was determined by exporting the MS/MS output files from the matched analysis into the ProteinLynx Global SERVER™ (PLGS) Version 2.4 as described in Section 3.2.4.4.

7.4.2 Identification and relevance of mutant peptides in MYH7

7.4.2.1 V606M, an established pathogenic mutation in HCM

The substitution of valine (V) for methionine (M) in myosin 7 (*MYH7*) is a widely published mutation (p.V606M, c.G1816A) in families with HCM. It was originally reported to be associated with normal life expectancy (Watkins, Rosenzweig et al. 1992) but was later identified in a family with multiple sudden deaths (Fanapanazir and Epstein 1994) challenging the concept of “benign” mutations in HCM. The mutation affects a highly conserved amino acid located close to the actin binding interface (S1 domain) and the insertion of methionine into this region has been shown to destabilise it and reduce the ATPase activity of myosin (Sata and Ikebe 1996). Functional studies using a myectomy sample with the V606M mutation found a hyper contractile molecular phenotype with greater myosin sliding speed and

reduced relaxation compared with donor heart (Jacques, Briceno et al. 2008). We were able to confirm the presence of the mutant tryptic peptide LNETVMGLYQ (1167.6 Da) alongside the wildtype peptide LNETVVGLYQ (1135.6 Da) offering support for a poison peptide mechanism in *MYH7* related HCM (Figure 7.6).

7.4.2.2 E1555G, a novel mutation in HCM

The substitution of glutamate (E) for glycine (G) in myosin 7 (*MYH7*) is a novel mutation that had not previously been reported in patients with HCM. The patient (male, age 37) harbouring this genetic variant had no evidence of familial disease and was included in a collaborative project examining somatic mutations in HCM (Nunez, Gimeno-Blanes et al. 2013). In silico prediction tools were used to ascertain the pathogenicity of this variant (Table 7.3). As expected the change from a positive polar amino acid, E to a non-polar one, G is predicted to alter the physico-chemical properties of the protein such as hydrophobicity, mass, polarity and charge. There was agreement between all bioinformatics tools to enable this mutation to be classified as pathogenic.

Tool	Website	Score or reliability index	Interpretation
SIFT	http://sift.jcvi.org/www/SIFT_seq_submit2.html	Score = 0	Damaging
Pmut	http://mmb.pcb.ub.es/PMut/	Index = 7	Pathological
PolyPhen2	http://genetics.bwh.harvard.edu/pph2/	Score = 0.833	Probably damaging
SNAP	http://www.rostlab.org/services/snap/	Index = 0	Non-neutral
PhDSNP	http://snps.uib.es/phs-snp.html	Score = 6	Disease

Table 7.3 Prediction of pathogenicity for the E1555G mutation in MYH7.

Using LC-MS five mutant peptides were identified alongside six wildtype peptides. The mutant peptides were identified in only two of the ten fractions analysed in this experiment. Modified peptides (oxidation of methionine and deamidation of glutamine) were also detected (Table 7.4). This illustrates the sensitivity of LC-MS and importance of pre-processing with fractionation. An example spectra is shown in Figure 7.7. The detection of mutant tryptic peptides within the myocardial sample further supports the pathogenicity of this novel mutation, and suggests it causes disease by a poison peptide mechanism

Fraction	Precursor MH+ (Da)	z	Sequence	Peptide Type	Products
pH 4	2230.022	3	(K)MELQSALEEEAEASLEHEEGK(I)	POM	13
pH 4	2246.017	3	(K)MELQSALEEEAEASLEHEEGK(I) * oxidation of methionine	VM	7
pH 4	2158.003	2.79	(K)MELQSALEEEAEASLEHEGGK(I)	POM	22
pH 4	2174.001	2.87	(K)MELQSALEEEAEASLEHEGGK(I) * oxidation of methionine	VM	7
pH 6	656.2982	1	(L)EHEGGK(I)	ISF	1
pH 6	927.447	1	(E)ASLEHEGGK(I)	ISF	0
pH 6	2158.012	2.75	(K)MELQSALEEEAEASLEHEGGK(I)	POM	34
pH 6	910.4196	1	(A)SLEHEEGK(I)	ISF	0
pH 6	2230.035	2.84	(K)MELQSALEEEAEASLEHEEGK(I)	POM	30
pH 6	2231.021	2.82	(K)MELQSALEEEAEASLEHEEGK(I) * deamidation of Q	VM	4
pH 6	2928.383	3.39	(K)QLEAEKMELQSALEEEAEASLEHEEGK(I)	MC	10

Table 7.4 Tryptic peptides related to E1555G mutation in MYH7

POM: Pass One Match, MC: Missed Cleavage, VM: Variable Modification, ISF: In Source Fragment. The mutant peptides are indicated by highlighting the altered amino acid in red.

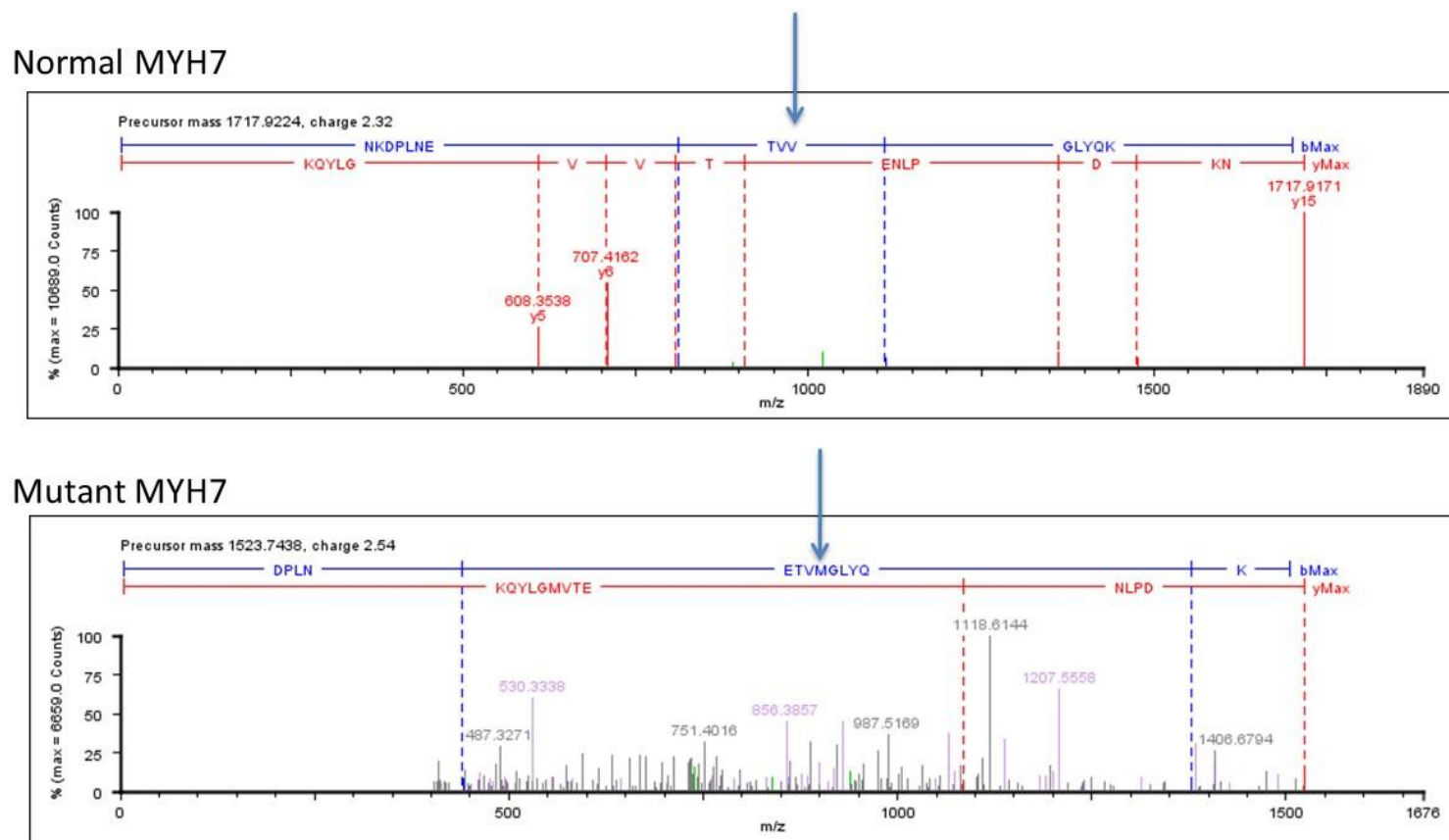
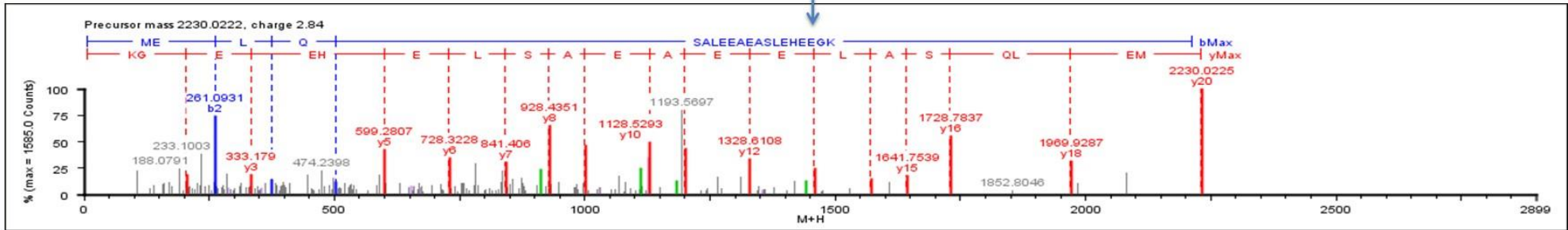


Figure 7.6 Spectra from mutant and normal peptide in V606M mutation

The top spectra show the normal peptide (LNETVVGLYQ) and the bottom spectra the abnormal peptide (LNETVMGLYQ). The arrow indicates the position of the substitution of valine (V) for methionine (M) caused by a mutation in the myosin 7, MYH7 gene.

Normal MYH7



Mutant MYH7

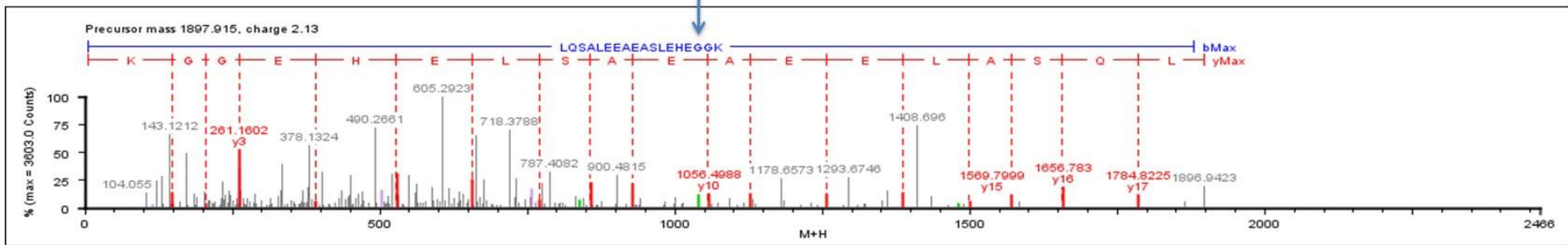


Figure 7.7 Spectra from mutant and normal peptide in E1555G mutation

The top spectra show the normal peptide (SALEEAEASLEHEEGK) and the bottom spectra the abnormal peptide (SALEEAEASLEHEGGK). The arrow indicates the position of the substitution of glutamate (E) for glycine (G) caused by a mutation in the myosin 7, MYH7 gene.

The present study was not designed to quantify the amount of mutant peptide within each sample. In order to do this, it would be necessary to use isotopically labelled AQUA (Absolute QUAntitation) peptides representing the mutant and wildtype sequences and develop a MRM-MS/MS assay for relative quantitation using peak areas. Ideally this should be performed on a myosin enriched sample. Such a technique has been used by one group to demonstrate that the fraction of mutated myosin protein varies according to specific mutation (Tripathi, Schultz et al. 2011).

7.5 Conclusion

No consistent relationship was evident between genetic substrate and proteomic profile, underscoring the broad heterogeneity of HCM from both molecular and clinical perspectives. Proteomic analysis did however support different disease mechanisms in the *MYH7* and *MYBPC3* sub-groups.

8 Clinical Application of Novel Biomarkers

Selected biomarkers markers that were previously validated were taken forwards for further development because they were considered to be potential clinical biomarkers or relevant to understanding the biology of the disease. A variety of approaches were used including: correlation with clinical parameters, measurement of peptides in urine and scrutiny of protein localisation using immunohistochemistry. Methods are currently in development in our laboratory to measure these biomarkers in plasma.

8.1 Visualisation of proteins in myocardium

8.1.1 Methods

Immunohistochemistry was performed to validate the presence or absence of specific proteins of interest. The antibodies used and their working dilutions are shown in Table 8.1. A tissue microarray was constructed using the methods described in Section 3.3. Thirty-eight patients (age 20 to 77 years, 63 % male) with HCM were included. Eighteen patients had been included in the proteomic study (internal validation) and twenty had not (external validation). Where possible, two samples were selected from each patient representing an area of normal myocardium and an area of myocyte disarray (Figure 8.1).

Antibody	Product details	Dilutions
Annexin A6	AbCam ab31026	1:50
Annexin A2	AbCam ab54771	1:250
SERCA2 ATPase	AbCam ab3625	1:250
Lumican	AbCam ab98067	1:400
Gelsolin	G1538 Sigma	1:2000
Carbonic Anhydrase III	AbCam ab118428	1:1000

Table 8.1 Working dilutions of antibodies used and product codes

Where heterogeneity of staining was observed an objective analysis of the myocardial samples was carried out using a “quickscore” method to quantify the results. Each case was scored according to the intensity of staining and the proportion of cells involved.

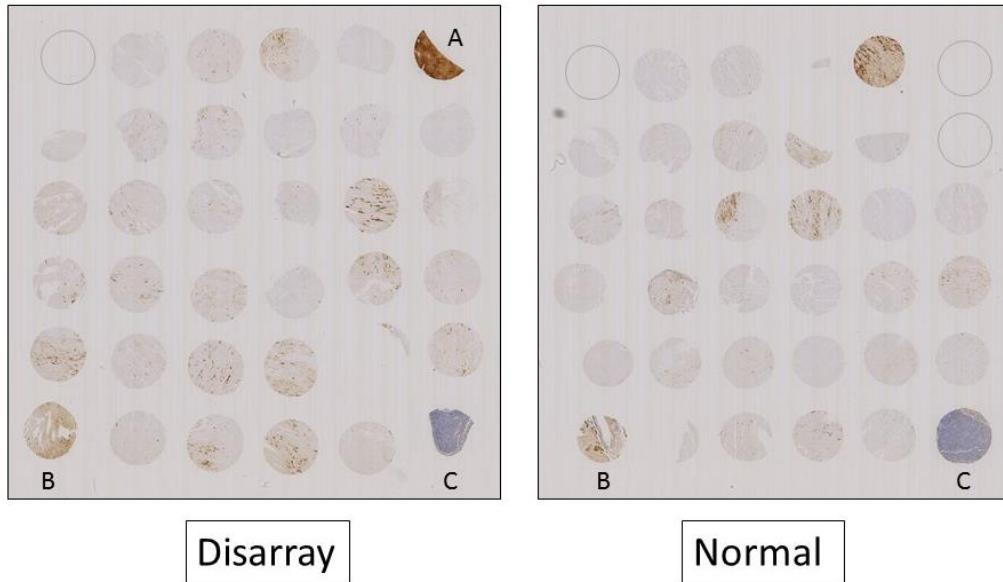


Figure 8.1 Design of tissue microarray

Myocardial samples arranged in two separate paraffin blocks according to the presence or absence of disarray. Marker tissue was included: A) Liver, B) Heart, C) Kidney.

8.1.2 Results

8.1.2.1 Annexins

Annexins are a family of 13 proteins known to bind phospholipids in a calcium dependent way. By global proteomic profiling several of these annexin proteins were identified in the heart. On differential expression analysis we found lower levels of A1, A2, A3, A5, A6, A7 and A11 in HCM hearts compared with controls (Annexin A2, 1.43 fold and Annexin A6, 1.65 fold, $p < 0.001$). Using immunohistochemistry we were able to illustrate the variable staining of A6 amongst patients with HCM and confirm the localisation of A2 to the intramyocardial capillaries and extracellular matrix (Figure 8.2). Annexin A6 has been reported to be both increased at the onset of heart failure in animal models and decreased in failing human hearts (Camors,

Monceau et al. 2005). There are no reports of Annexin A6 in HCM but a recent study has shown Annexin A6 has a regulatory role in maintaining the balance between hypertrophy and cell death in cardiomyocytes (Banerjee, Chander et al. 2015).

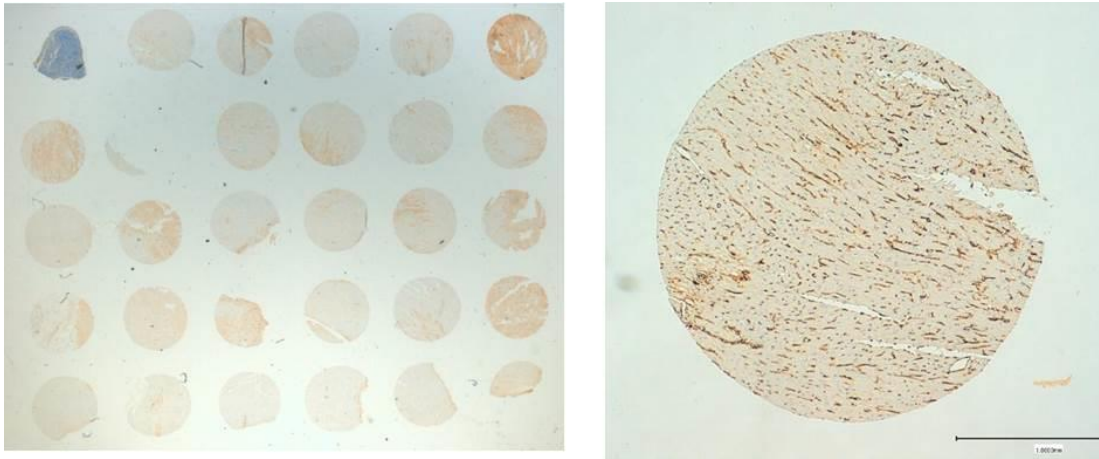


Figure 8.2 Annexin staining in HCM

LEFT: Annexin A6 stained tissue microarray array illustrating different patterns and staining intensity. Normal heart is represented top right. LEFT: Example of Annexin A2 stained myocardium extracellular matrix within the myocardium.

8.1.2.2 Carbonic anhydrase III

The presence of carbonic anhydrase (CA) in heart muscle is well established however CA II and IV are the predominant isoforms reported in the literature. The role of CA is discussed in detail in Section 8.3.2.3. Human skeletal muscle is known to contain large amounts of CA III but only trace levels have been detected in human heart (Harju, Botorabi et al. 2013). By proteomic profiling CA III was found to be upregulated in HCM hearts compared with controls. This finding was confirmed by immunohistochemistry (Figure 8.3). Carbonic anhydrase III was found to be localised to the cytoplasm of the cardiomyocytes (Figure 8.4). There was marked heterogeneity not only between samples but also within the same sample.

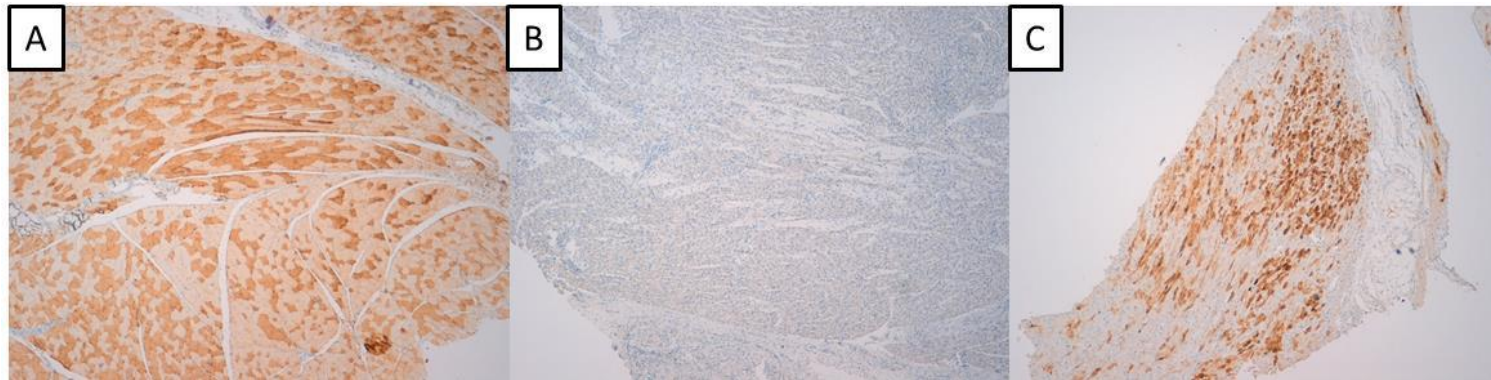


Figure 8.3 Histological validation of CA III in HCM.

Carbonic anhydrase III staining in A) skeletal muscle (positive control), B) normal heart, C) HCM Heart (x 10 magnification)

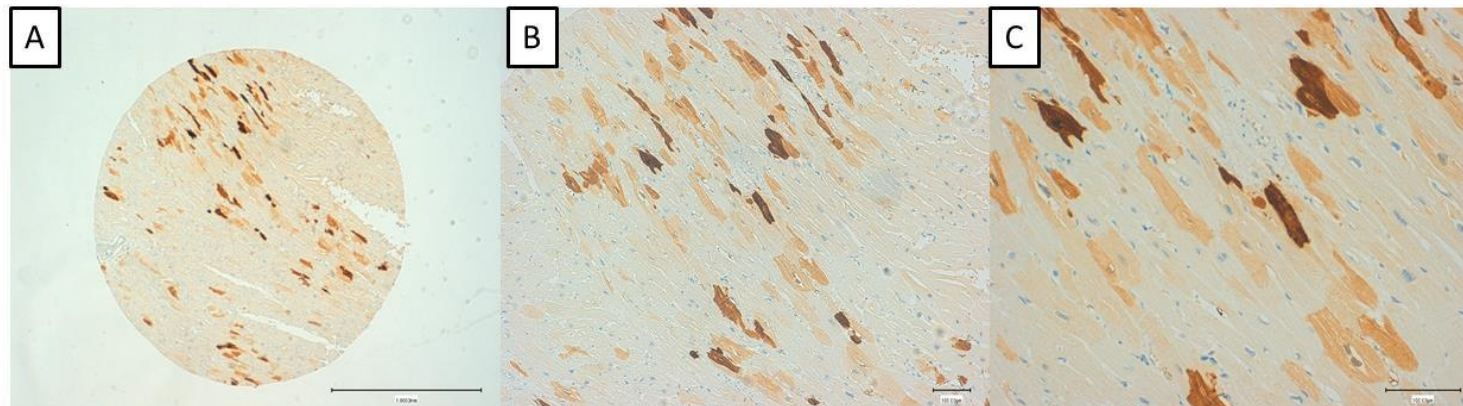


Figure 8.4 Localisation of CAIII within cytoplasm of cardiomyocytes

CAIII staining in non-disarrayed heart from a 65-year-old female. at A) low (x 4), B) medium (x 10) and C) high power(x 20 magnification..

Heterogeneity of CA III expression

Each case on the TMA was scored (0 to 3) according to the intensity of staining of the cytoplasm of the cardiomyocytes (Figure 8.5) and the proportion of cells involved (0=nil, 1= less than 25 %, 2 = 25-50 %, 3 = more than 75%). A combined “quick-score” was given for each sample as the sum of the intensity and proportion scores. Comparison was made between the disarrayed and non-disarrayed samples using a paired t-test.

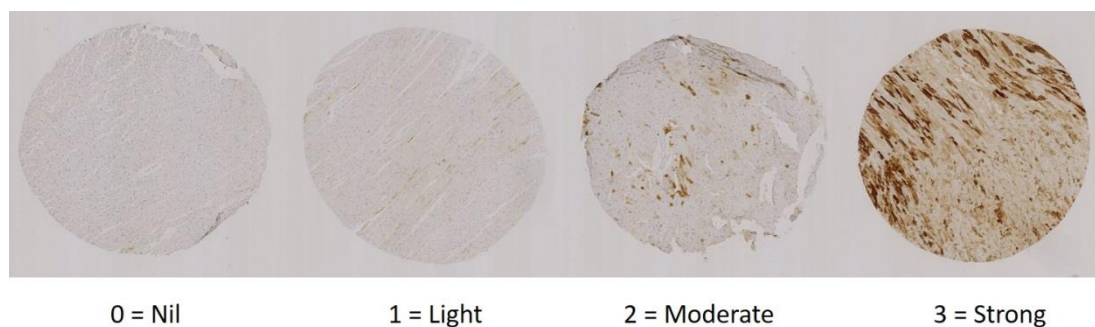


Figure 8.5 Quantification of CA III staining intensity

Examples of grading for CA III staining of cardiomyocytes, including a) no staining, b) light scanty, c) moderate patchy, d) strong diffuse.

There was a significantly higher score observed in the disarrayed group (3.2 ± 1.6) compared with the non-disarrayed group (2.1 ± 2.1 , $p=0.047$). There was no significant correlation with age, sex or genotype. Increased CA III expression was also observed in samples not included in the proteomic study providing further validation for our earlier finding.

8.1.2.3 Lumican

Although the antibody work up was successful, and we were able to confirm the presence of lumican in the heart (Figure 8.6), the staining was not considered to be reliable for quantitative analysis on the tissue array. The intense staining appeared to be extra-cellular but it was not clearly associated with regions of fibrosis.

8.1.2.4 Gelsolin

Immunohistochemistry confirmed gelsolin as a smooth muscle vascular marker in HCM (Figure 8.6), but the staining of myocytes was not reproducible between samples. It was not taken forward for further analysis.

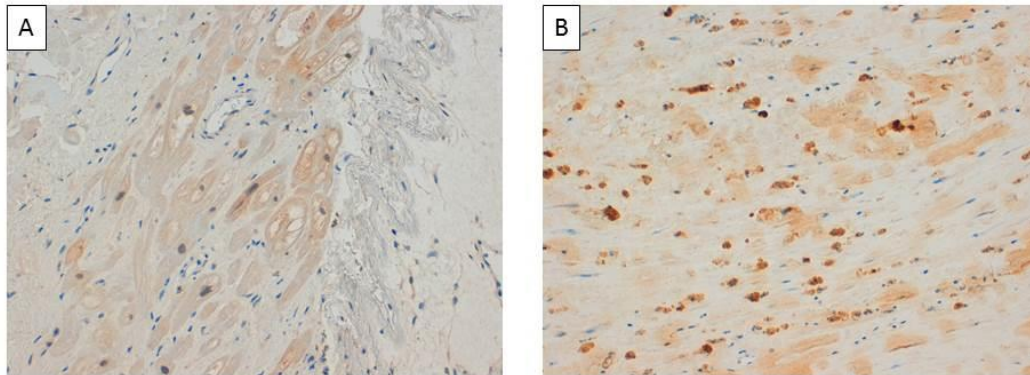


Figure 8.6 Immunohistochemical validation of myocardial proteins

Immunohistochemical analysis of paraffin embedded human myocardium from a patient with HCM labelling A) Gelsolin with G1538 (Sigma) at 1/2000, and B) Lumican with ab98067 (AbCam) at 1/400 and. Magnification x 20.

8.1.3 Conclusion

Tissue microarray technology is a fast and effective way to simultaneously analyse large numbers of specimens. The advantages of this approach are that it resolves the issue of blood contamination, allows localisation of proteins, uses archived FFPE tissue and enables direct translational into clinical use. The heterogeneity of protein expression suggests that not only does cardiomyocyte behaviour differ between patients but also within different regions of an individual's myocardium. By sampling microscopic cores of tissue there is a chance that the sample may not be representative of the whole organ. However, we have confirmed the localisation of CA III and can confidently report its increased expression in adult hypertrophic cardiomyopathy.

8.2 Screening biomarkers in urine

8.2.1 Methods

To determine whether proteins of interest were detectable in peripheral body fluids, a small study was conducted applying the multiplexed MRM-MS assay developed in Chapter 6 to urine samples. Additional biomarkers reported to be altered in heart disease and detectable in urine (albumin, cystatin C and osteopontin) were also included (Valente, Damman et al. 2012). Urine samples obtained from 19 patients with HCM (age 18-41 years, 74 % male) and 16 age-matched healthy volunteers were studied. Urine was prepared as described in Section 3.2.6. Yeast enolase was spiked into the samples prior to digestion and used as an internal standard. Urinary concentrations of compounds of interest were corrected for dilution differences by measuring the urinary creatinine concentration. Differences between the two groups were evaluated using a Mann Whitney U test and graphs were constructed using Graph Pad Prism (Version 5).

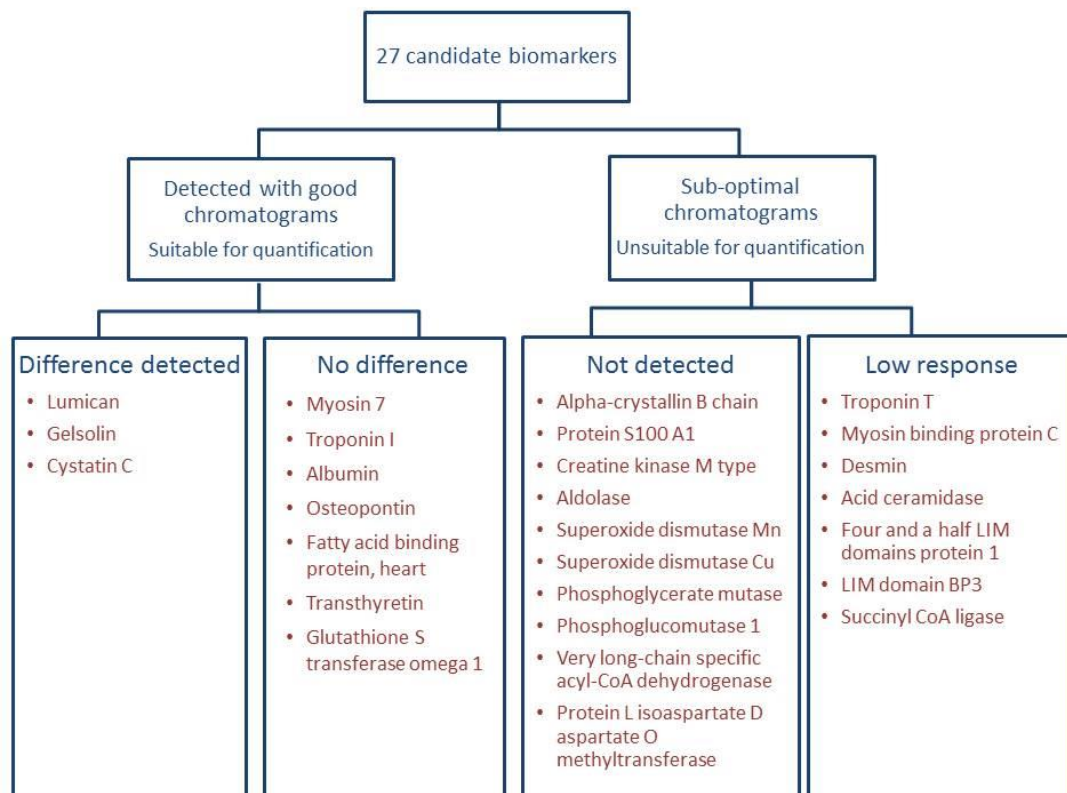


Figure 8.7 Biomarker screening in urine using UPLC-MS/MS assay

Twenty seven candidate biomarkers were screened in urine samples from patients with HCM (n=19) and healthy volunteers (n=16).

8.2.2 Results

A multiplex and targeted UPLC-MS/MS proteomic assay using 1 mL of urine was used to screen twenty-seven potential biomarker peptides. Ten of these were easily detectable in the urine with chromatography suitable for quantitative analysis. Seventeen peptides were not suitable for comparison because they were either undetectable or the chromatography was sub-optimal for accurate peak integration (Figure 8.7). A statistically significant difference between disease and control groups was found for three proteins: in the HCM group lumican was 5.6-fold higher, gelsolin was 3.7-fold higher and Cystatin C was 1.6-fold higher, compared with controls (Figure 8.8). The MRM parameters for the tryptic peptides representing these three biomarker proteins in urine are shown in Table 8.2.

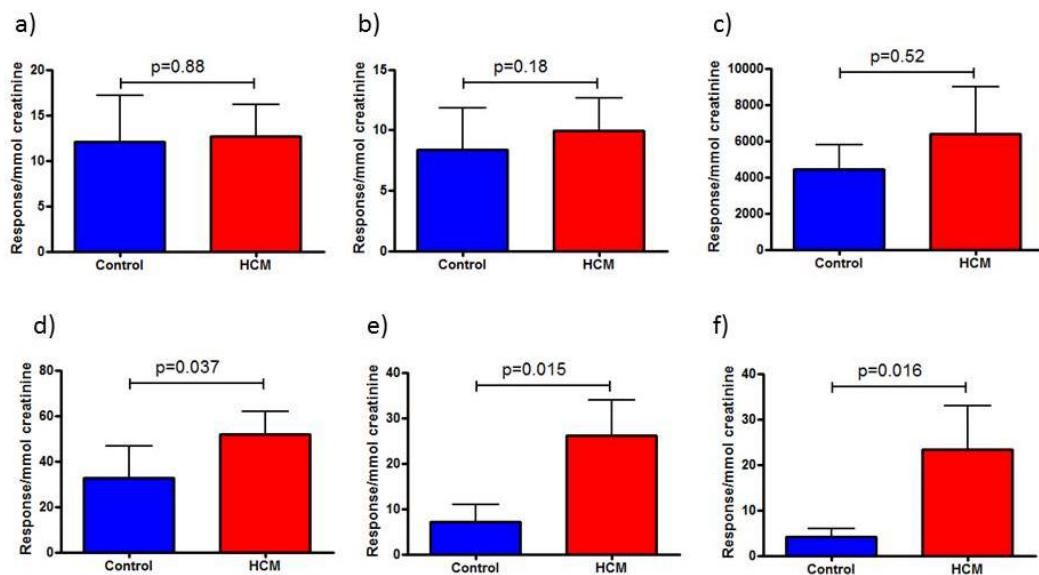


Figure 8.8 Targeted MRM LC-MS/MS analysis in HCM urine samples

Bar charts showing mean \pm SEM of proteins detected in urine from patients with HCM ($n=19$) and healthy volunteers ($n=16$). No difference was detected in abundance of a) fatty acid binding protein heart, b) myosin 7, c) osteopontin, but significantly higher levels of d) cystatin C, e) gelsolin and f) lumican were detected in the HCM group.

Protein Peptide sequence	Charge	Precursor ion m/z	Cone (V)	Product ion m/z	Collision (V)
Lumican NIPTVNENLENYYLEVNQLEK	2+	1268.69	58	1155.06	36
Gelsolin TPSAAYLWVGTGASEAEK	2+	919.90	8	849.67	26
Cystatin C ALDFAVGEYNK	2+	614.12	48	709.45	18

Table 8.2 Potential urinary biomarkers in HCM

MRM transitions used in the UPLC-MS/MS assay. The precursor (parent) and product (daughter) ion m/z and the cone and collision energy used for fragmentation are shown.

8.2.3 Conclusion

The multiplexed targeted proteomic assay designed to validate proteins in myocardial tissue was successfully applied to screen potential biomarkers identified in the urine. Although less than half the biomarkers examined were suitable for quantification, three proteins were found to be significantly different with a directional change similar to that observed in tissue. Importantly all have potential pathological mechanisms associated with HCM. Further optimisation of the urine assay could be performed to detect those peptides with a lower response. Overall this proof of principle study highlights the potential role of urine as a biofluid for biomarker detection in this disease and illustrates the value of an MRM-MS based biomarker pipeline for rapid translation into a clinical test.

8.3 Correlation with clinical parameters

Clinical characteristics of patients and operation details were obtained retrospectively from hospital notes. The targeted MRM-MS data was compared with clinical parameters to determine whether phenotypic differences could account for the heterogeneity in specific protein levels.

8.3.1 Clinical relevance of lumican in HCM

Histological reports and operation notes were initially reviewed to gain insight into the microscopic phenotype of patients with elevated lumican concentration. The highest level was measured in a 73-year-old female with familial HCM due a mutation (K847E) in *MYH7* gene. At operation, the surgeon described the myocardium as “gritty, speckled and fibrous”. On microscopic examination, the pathologist reported the myocardium to show “florid myocyte hypertrophy with extensive interstitial fibrous scarring”. The lowest lumican concentration was observed in a 56-year-old male with non-familial HCM and no identifiable genetic mutation. Histological examination showed “features of hypertrophy but no myocyte disarray and no significant fibrosis”. This led to the hypothesis that patients with elevated myocardial lumican concentration could be identified using cardiac magnetic resonance (CMR) imaging with late gadolinium enhancement.

8.3.1.1 Relation to late enhancement on CMR

Lumican was observed to be significantly elevated in heart samples from 51 patients with HCM (283 ± 176 $\mu\text{mol/l}$) compared with 7 controls (132 ± 26 $\mu\text{mol/l}$). Thirty-six patients underwent CMR imaging with administration of late gadolinium contrast. Mean concentration was significantly higher in the group of patients ($n=22$) with late enhancement on CMR (315 ± 144 $\mu\text{mol/l}$, $p=0.02$) compared to patients ($n=12$) with no detectable LGE (208 ± 121 $\mu\text{mol/l}$, $p=0.02$). There was one notable outlier in the group without LGE and some overlap between the groups (Figure 8.9). Lumican concentration was positively correlated with left atrial area ($r=0.314$, $p=0.025$) but there was no significant correlation between lumican and age, NT-proBNP concentration or maximal LV wall thickness.

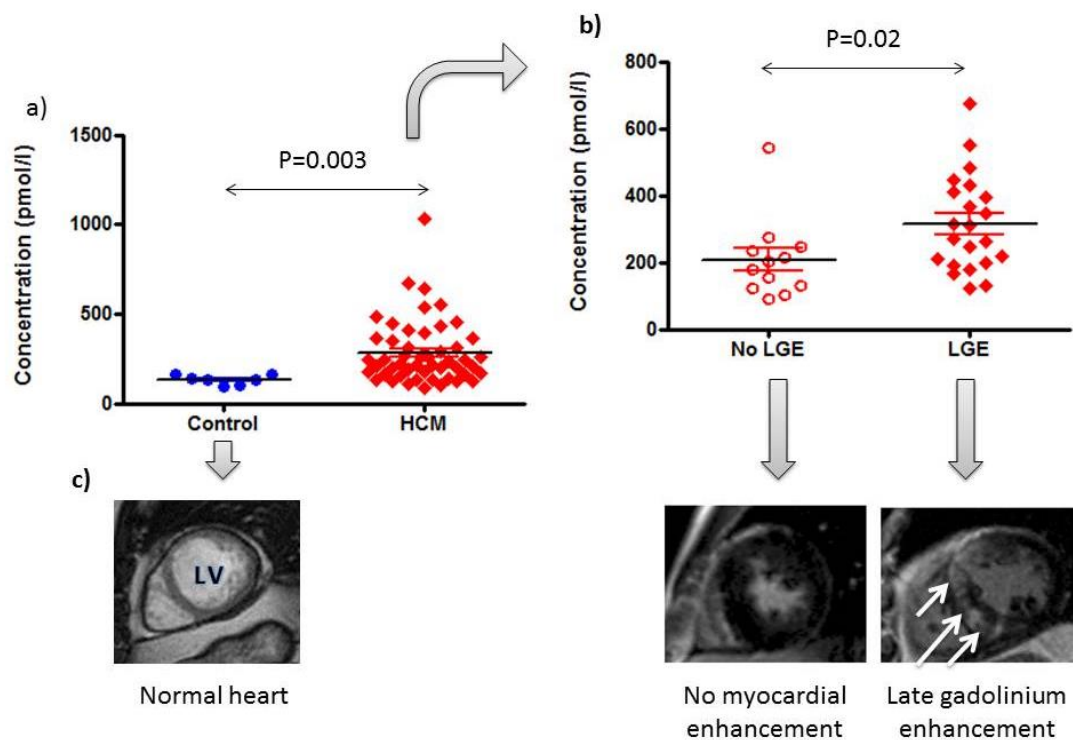


Figure 8.9 Lumican concentration and fibrosis on CMR

a) Lumican concentration in myocardial tissue from 51 patients with HCM and 7 controls, b) Myocardial lumican concentration in a sub-group of 36 patients with HCM stratified according to presence of late gadolinium contrast on CMR imaging, c) representative examples of CMR imaging showing (left to right) a normal heart (LV=left ventricle), hypertrophic cardiomyopathy with no LGE and HCM with late gadolinium enhancement in the septum (indicated by white arrows).

8.3.1.2 Biological relevance of lumican in HCM

The extracellular matrix (ECM) of the heart provides structural and functional integrity by mediating the mechanical connections between cardiomyocytes, cardiac fibroblasts and blood vessels. It also transmits extracellular mechanical signals to cardiomyocytes and contains cytokines and growth factors that modulate cardiac function and determine the fate of cardiac cells. Lumican mRNA has been identified in human heart and appears to significantly upregulated in HCM and DCM compared with controls (Hwang, Allen et al. 2002). In a mouse model lumican protein levels were shown to increase in the left ventricle after aortic banding, and reduced after de-banding (Engelbrechtsen, Waehre et al. 2013) implying that production can be altered in

response to external stimuli. In a separate study by the same group, stimulation of cardiac fibroblasts with recombinant glycosylated lumican, caused an increase in type I collagen, decrease in activity of the collagen-degrading enzyme matrix metalloproteinase-9 and an increase in the phosphorylation of fibrosis-inducing SMAD3 (Engebretsen, Lunde et al. 2013). This direct link between lumican and molecules known to be important for cardiac remodelling and fibrosis implies a role in disease pathogenesis. Our discovery that lumican is significantly raised in HCM hearts provides new insight into the abnormal collagen homeostasis that characterises this disease.

8.3.2 Clinical relevance of CA III in HCM

Heterogeneity in expression levels of CA III was demonstrated across HCM samples in both the histological and proteomic study. Lower levels of CA III were found in the control and the aortic stenosis (AS) samples (Figure 6.9).

8.3.2.1 Relation to clinical parameters

In patients with both HCM (n=51) and aortic stenosis (n=7), CA III levels were positively correlated with recognised heart failure markers including serum NT-pro BNP ($r=0.43$, $p<0.001$, $n=58$) and left atrial diameter ($r=0.37$, $p=0.004$, $n=58$). There was no correlation with age ($r=0.15$, $p=0.23$, $n=58$) or time on cardiopulmonary bypass ($r=-0.17$, $p=0.213$, $n=58$) both potential confounders of underlying myocardial metabolic and acid-base status. There was no significant difference in CA III level in the group of patients with late enhancement on CMR compared to those with no detectable LGE.

8.3.2.2 Relation to other biomarkers

There was a positive correlation between % skeletal actin and CA III ($r=0.271$, $p=0.04$) and levels of FHL1 and CA III ($r=0.352$, $p=0.07$). Both skeletal actin and FHL1 are part of the fetal gene program, recognised to be activated in pathological cardiac hypertrophy and adult heart failure (Razeghi, Young et al. 2001).

8.3.2.3 Biological relevance of CA III in HCM

In the heart, cellular or mitochondrial acidosis can impair contractility by reducing calcium sensitivity and compromising energetics. By transporting carbon dioxide out of cells into nearby capillaries, acidosis is prevented and intracellular pH maintained. The correlation between CA III levels and established biomarkers of heart failure such as NT-proBNP could be a sign that these hearts are increasingly metabolically active in an attempt to compensate for their decrease in function. The association of CA III levels with other proteins involved in fetal reprogramming suggests this is likely to be a protective mechanism that enables cells to survive in a compromised environment and that CA III is upregulated to clear the excess acid produced by anaerobic glycolysis.

Murine studies

Mice lacking CA III have been shown to be viable, fertile and have a normal life-expectancy even when exposed to exercise stress or hyperoxia (Kim, Lee et al. 2004). Transgenic mice overexpressing CA III in the heart show normal cardiac phenotypes under non-stress conditions but when exposed to low pH environment demonstrate better tolerance to acidosis with less impaired heart function than wild type controls (Feng and Jin 2015). These data suggest CA III may function as part of the defence system and is required where there is demand for increased demand of carbon dioxide removal.

Human studies

CA III has not been specifically studied in human hearts. However, there is emerging evidence that other isoforms are activated in heart disease. Increased expression of CA II and CA IV mRNA and protein has been demonstrated in both hypertrophic and dilated failing human hearts using surgical biopsy specimens (Alvarez, Quon et al. 2013). Increased expression of CA I and CA II has been found in diabetic patients undergoing coronary revascularisation, compared with non-diabetic controls, where CA II levels were associated with greater myocyte hypertrophy and apoptosis (Torella, Ellison et al. 2014). In skeletal muscle, Ca III exhibits a fibre type-specific expression pattern with preferential expression in slow, type 1 fibres (Harju, Botorabi et al. 2013).

A potential drug target

Many of the CA enzymes are intriguing as drug targets, and some CA inhibitors are already clinically used as therapeutic agents in the treatment of various diseases including glaucoma and epilepsy. The membrane permeant carbonic anhydrase inhibitor, acetazolamide, was historically used as a diuretic in patients with severe congestive heart failure (Relman, Leaf et al. 1954). Carbonic anhydrase inhibition has been shown to prevent and revert cardiomyocyte hypertrophy in cultured myocytes (Alvarez, Johnson et al. 2007). It remains to be determined whether carbonic anhydrase activity can or should be modulated in HCM.

8.4 Conclusion

We have provided further validation for two novel biomarkers identified in the proteomic study. Lumican is elevated in HCM, can be detected in urine and correlates with the presence of scar on cardiac magnetic resonance imaging. Carbonic anhydrase III is elevated in HCM, localises to the cytoplasm of myocytes in regions of abnormal myocardium and correlates with markers of heart failure. An unbiased proteomic approach has led to the discovery and validation of clinically relevant markers in HCM important to the underlying biology of the disease.

9 Discussion

Proteomics provides a novel way to understand the molecular basis of disease. This study represents a first attempt to characterise the proteome of human HCM. The work presented in this thesis provides supporting evidence for many of the known disease mechanisms in HCM but also identifies differences in the expression of several proteins not described previously. It has generated a large amount of data that will be used for future clinical and basic science studies. There are a number of challenges and limitations with proteomic research which are addressed in this chapter with particular reference to our data.

9.1 Proteome coverage

Over 1500 unique proteins were identified in this study. Many of these were mitochondrial, reflecting the high energetic demands of cardiac tissue and are comparable with other fresh-frozen heart tissue studies (Aye, Scholten et al. 2010, Hammer, Goritzka et al. 2011, Polden, McManus et al. 2011). At the start of our study, the deepest coverage achieved so far used a multifaceted approach combining different fractionation and digestion methods to identify 3584 distinct proteins, of which only 520 were consistently identified with all methods. (Aye, Scholten et al. 2010). The human proteome, defined for the first time in 2014, contains closer to 20,000 unique proteins (Kim, Pinto et al. 2014, Wilhelm, Schlegl et al. 2014). We used stringent criteria and multiple quantitative methods to identify proteins reliably which may have been at the expense of greater proteome coverage. Undoubtedly, further work needs to be done before the entire cardiac “proteome” can be examined in a single experiment.

9.1.1 Optimising proteome coverage

There are several strategies for maximising proteome coverage including depletion, fractionation and enrichment techniques. The three techniques explored in the preliminary work (Section 4.2) illustrate the advantages and disadvantages of some

of these methods. In general, the more abundant a protein is in a sample the more likely it is to be detected by MS/MS. We found albumin and myoglobin to be the most abundant proteins within the myocardium. By performing an entirely label-free proteomic study, all fractions contained these proteins, likely masking the lower abundant proteins within the sample matrix. An advantage of combining pre-fractionation using gel electrophoresis is that the molecular weight (and isoelectric point with 2D separation) of a protein provides further insight into its biological function. Differences between a protein's observed molecular weight and that predicted by its full length amino acid sequence can be the result of different types of post-translational events, such as alternative splicing and post-translational modifications. These were not specifically examined in the present study, but would be very relevant to appreciate disease mechanisms. For example, the contractile abnormalities observed in HCM and heart failure have been in part attributed to reduced phosphorylation of MyBP-C (Copeland, Sadayappan et al. 2010).

9.1.2 Sub-proteomes

There was a noticeable absence of extracellular matrix (ECM) proteins in our proteome, except for a few proteins that are secreted e.g. MMP12 and MMP-3 (involved in collagen degradation) and Galectin-1 (regulates apoptosis). These were found in low abundance and not identified in all samples. This likely reflects the methods used to extract and digest proteins from the tissue. Proteomic characterisation of the cardiac ECM has only recently been reported in the literature (Barallobre-Barreiro, Didangelos et al. 2012). Biochemical sub-fractionation can be performed using a de-cellularisation step to reduce contamination of other cellular material, prior to a denaturing extraction using a guanidine hydrochloride buffer. This method has enabled identification of over 100 extra-cellular space proteins (Didangelos, Yin et al. 2010). Because fibrosis is an early, potentially reversible process, better understanding of the ECM proteome in HCM has important implications.

9.1.3 Technological advances

Future advances in proteomic methodology and technology will allow deeper coverage, using smaller samples, in less time. New pre-fractionation techniques already being used in our laboratory include direct coupling of the mass spectrometer to an online high resolution LC system and the introduction of ion mobility mass spectrometry (IM-MS). Ion mobility separation includes a gas phase separation of ions (based on size and geometry) hence adding a third dimension, “drift time” to the separation of complex mixtures, over and above retention time and m/z ratio. These advances make in depth profiling simpler, quicker, more sensitive and accurate (Chouinard, Wei et al. 2016). A major challenge in proteome research is reproducibility due to the variation in sample preparation, MS instrumentation and data analysis tools. In 2005 criteria for reporting proteomic experiments were standardised (Taylor, Paton et al. 2007) so that protocol details are published alongside results. This will help to increase reliability and accuracy of data.

9.1.4 Gene annotation

Gene Ontology (GO) is the most widely used vocabulary to describe genes and gene products, with GO terms displayed by all major sequence databases. Although databases are constantly being updated, annotation of the human genome is incomplete and predicted to take several decades (Baumgartner, Cohen et al. 2007). There have been initiatives to improve annotations of cardiovascular associated genes with manual curating (Lovering, Dimmer et al. 2009) however approximately 5 % of proteins identified in this study still had no associated GO annotations. There are also recognised biases associated with manual curation, for example proteins involved in cardiac development have been particularly well annotated because of a dedicated study (Khodiyar, Hill et al. 2011).

9.1.4.1 Pathway analysis

Most proteins are assigned multiple GO annotations, so to assess their relevance within a sample they must be compared to an established reference dataset such as the full proteome. Rather than clustering proteins according to GO terms (Section 5.2.2) it is preferable to recognise proteins that are over-represented or statistically

enrichment according to the pathways they represent e.g. mitochondrial dysfunction (Section 5.2.5). This type of analysis benefits by inputting a large number of proteins, but in turn also relies on understanding protein-protein interactions and accurate annotation of pathway databases. One of the many problems with pathway knowledge databases is that they are curated from experiments performed in different cell types under different experimental conditions. The cardiovascular community has gone some way to addressing this by the development of a specific resource for cardiovascular proteomics (www.heartproteome.org) (Zong, Li et al. 2013).

9.2 Differences between HCM and Controls

Clear differences were found in the myocardial proteome of HCM compared with controls. The most significant differences were seen in the myofilament and metabolic proteins signifying both a structural and biological perturbation in HCM. This is in accord with our current understanding of the disease. Increased expression of fetal gene program markers support the hypothesis that hypertrophy in HCM is a secondary phenotype, and common pathways are involved in the induction of cardiac hypertrophy regardless of the genetic and non-genetic forms. However, sample size and sample heterogeneity is an important limitation that needs to be considered when interpreting the results from proteomic experiments. By profiling samples individually rather than pooling them into disease and control groups we were able to examine some of this heterogeneity.

9.2.1 Adequacy of control tissue

Differential protein expression between HCM and control samples may in part reflect the circumstances under which tissues were obtained. The HCM tissue was taken from a cardiopleged heart with patients maintained on cardiopulmonary by-pass; the normal heart tissue was obtained following brain death where activation of the Cushing reflex may involve acute ischemia-reperfusion stress and therefore contribute to oxidative protein modification. The time from sampling-to-freezing heart tissue may have varied between the groups. Furthermore the age of samples

may have impacted on results with older samples being subject to more degradation than more recently obtained samples.

9.2.2 Protein estimation and cell type

Myocardial homogenate was quantified using a standard protein assay (Section 3.2.2). However, it is incorrect to assume that cardiomyocytes made up the same proportion of total protein present in all the samples being studied. We know from non-invasive imaging and histology of HCM hearts that interstitial and/or replacement myocardial fibrosis is a common occurrence. More than 50 % of the cells of the heart are cardiac fibroblasts (Vliegen, van der Laarse et al. 1991). In addition, endothelial cells form the endocardium or inner layer of the heart, and smooth muscle cells contribute to the coronary arteries; both of which are present throughout the myocardial tissue. We did not routinely determine the presence of non-cardiomyocyte components in the myocardium. Fibrosis could be estimated by measuring hydroxyproline content and/or performing sirrus red histological staining on the tissue. A newer approach is to combine proteomics with laser capture microdissection to examine pure cell populations from complex tissue sections (Datta, Malhotra et al. 2015). This approach has been used to study specific populations of cells in cancer tissues and to classify amyloid sub-types in clinical specimens (Vrana, Gamez et al. 2009).

9.2.3 Blood contamination

Heart tissue is very vascular making blood contamination inevitable. This is illustrated by the presence of proteins such as albumin and haemoglobin in our samples. This unavoidable issue has been addressed previously by proposing a rank is assigned to all proteins based on their abundance in the plasma versus the heart (Aye, Scholten et al. 2010). Thus, lowly abundant or absent plasma proteins that rank higher in the heart e.g. heart specific fatty acid binding protein, may be selected as potential tissue leakage biomarkers.

In our study, despite washing of the samples, we found that haemoglobin concentration in the heart homogenate, correlated with haemoglobin concentration of the blood (Section 7.2.1.1). Although this indicated there was residual blood

contamination it suggested this was probably uniform across all samples, from the point at which they were prepared for MS analysis in our laboratory.

Using diseased tissue rather than plasma to identify biomarkers is usually preferred because it reduces signal to noise; plasma is dominated by a few highly abundant proteins as well as being a highly complex fluid containing proteins leaked from tissues around the body (thus diluting particular proteins of interest). Global plasma profiling in HCM has recently been undertaken in our laboratory and it will be valuable to compare the differences between tissue and plasma proteomes to guide future biomarker research.

9.2.4 Myocardial metabolism

It is widely appreciated that pathologic hypertrophy is associated with reprogramming of cardiac energy metabolism (Lopaschuk, Ussher et al. 2010). Fatty acids represent the predominant fuel for the adult heart under normal physiologic conditions. However, hypertrophied and failing hearts are characterised by a switch from fatty acid oxidation to glycolysis and lipid accumulation (Davila-Roman, Vedala et al. 2002, Krishnan, Suter et al. 2009). Although it is not possible to determine the extent to which individual pathways contribute to metabolic demand, there is convincing evidence that the energetic demands in HCM are greater than in normal hearts. This primary “energy compromise” hypothesis proposes all HCM mutations result in inefficient use of ATP and an result in an increased energetic cost of contraction, even without apparent hypertrophy (Crilley, Boehm et al. 2003). This is supported by the fact that several inherited metabolic syndromes (e.g. glycogen storage disorders) in which mitochondrial energy production is impaired can result in asymmetric cardiac hypertrophy akin to that seen in HCM (Frey, Luedde et al. 2012). Our proteomic work confirmed that not only does the HCM heart revert to a fetal gene program where glycolysis predominates (Taegtmeyer, Sen et al. 2010) but also that it’s mitochondrial energy producing capabilities are severely reduced (Section 5.3.2). It is tempting to link the profound metabolic derangement directly to the underlying genetic defect however we also found a reduction in overall muscle creatine kinase (Section 6.4.8) and it is feasible that this underlies the reduced

capability for ATP production. A potential criticism of our study is that the myocardial samples were obtained from patients with HCM undergoing surgery; this in itself represents advanced disease. Only by studying early disease, without hypertrophy is it possible to distinguish whether these profound changes in cardiac energy metabolism are a primary or secondary phenomenon.

9.3 Biomarker Discovery

Many biomarkers identified to date have been stuck in the research setting and not carried over to clinical use. It is increasingly recognised that not many proteins will be stand-alone biomarkers, like troponin or NT-proBNP. Rather a panel of biomarkers might be more effective to describe the multifaceted features of disease. Increasingly proteomic approaches have therefore been employed to identifying novel cardiovascular biomarkers (Domanski, Percy et al. 2012, Marshall, Edwards et al. 2014). The large volumes of data generated by discovery based proteomic studies can provide many candidate markers that require further validation.

9.3.1 Validation by mass spectrometry

A lot of proteomic laboratories identify proteins by mass spectrometry then validate findings by immunoblotting. This is probably influenced by the wide use of immunoblotting in molecular biology and cell signalling research. Although this technique is reliable to detect an individual protein in a complex mixture it assumes the antibody epitope is specific for that particular isoform, but we know that many antibodies show cross reactivity to non-specific proteins. We therefore employed a tandem mass spectrometry based MRM-MS approach to validate our findings.

9.3.2 Advantage of a multiplexed approach

By using an MRM-MS technique, multiple peptides (representing proteins) were measured simultaneously. The assay designed in this study was able to detect and quantify 35 proteins in a sample over a 10-minute run. Additional markers have since been added to the assay which is likely to be further improved as more clinically relevant biomarkers are established and less relevant ones discarded. This can be

useful to correlate the various biomarkers with each other. Another advantage of such an approach is that it can be applied to other biological samples such as blood and urine. These peripheral samples are easier to collect and can offer rapid validation in much larger population. Furthermore, it enables rapid translation into the clinical arena, since most hospital laboratories already have access to a tandem mass spectrometer.

9.3.2.1 Challenges of plasma proteomics

One of the challenges presented by measurement of peptides by MRM-MS in plasma is the huge dynamic range of proteins it contains, ranging over several orders of magnitude from albumin (mg/ml concentration) to cytokines (ng/ml concentration). To overcome this problem some researchers have used depletion methods to remove the top 10 or 12 most abundant proteins by strong cation exchange (SCX) fractionation (Huttenhain, Soste et al. 2012). This approach has been applied to detect six known cardiac biomarkers from patients with HCM undergoing alcohol septal ablation (Keshishian, Addona et al. 2009). Another potential limitation of using an MRM-MS approach is the requirement for protein digestion prior to MS analysis. Samples were incubated with trypsin overnight for our assay which would limit the ability of such a test being used for urgent clinical decision making e.g. measuring troponin concentration in the setting of acute myocardial infarction.

9.3.3 Clinical validation

In addition to measuring protein concentrations, detailed phenotype information is needed for novel biomarkers to be clinically valid. A context specific biomarker is more likely to be feasible for translation to the clinic than random sampling.

9.3.3.1 Linking biomarkers to pathophysiology

In this study lumican was an attractive marker to scrutinise because of its potential link with fibrosis, an early and adverse prognostic feature in HCM. The correlation between myocardial lumican concentration and the presence of replacement fibrosis determined non-invasively using cardiac magnetic resonance imaging affirmed this (Section 8.3.1.1). The stimulus to increased lumican secretion in HCM remains unclear, but there are some characteristics of the disease that may be pertinent. In

in vitro studies of cardiac fibroblasts suggest that mechanical stress may be an important stimulus to lumican secretion which is of relevance in HCM where wall stress is increased by left ventricular outflow obstruction, asymmetric hypertrophy and cardiomyocyte disarray (Engebretsen, Lunde et al. 2013). Inflammatory cytokines (specifically IL-1b) also stimulate production of lumican and its release from cells. Several investigators have examined the role of inflammation in HCM and have reported elevated levels of CRP, IL-6 and TNF α (Hogye, Mandi et al. 2004, Dimitrow, Undas et al. 2008). The relation between inflammatory cytokine signalling and lumican remains to be determined but clearly represents another potential stimulus to its production in HCM.

9.3.3.2 Detection of disease specific markers in peripheral fluids

Despite recognition of novel biomarkers in myocardial tissue, these can only become clinically useful if they can be readily detected in peripheral samples such as saliva, blood or urine. In addition, they must have a definite role in diagnosing and/or monitoring certain features of the disease, over and above current clinical markers.

Urinary biomarkers

Detection of markers in the urine was evaluated in this project because it did not have the aforementioned challenges associated with plasma measurement and protocols had already been developed in our laboratory. The MRM-MS validation assay could be applied to rapidly determine whether markers of interest could be detected in urine. Although only a small pilot study was conducted we were able to detect lumican reproducibly in urine, perhaps reflecting its small size and ability to be filtered by the glomerulus. Larger studies are needed to confirm this finding and determine whether it is a reliable and clinically useful marker. Proteinuria (the presence of serum proteins in urine) is a well-established risk factor for cardiovascular mortality (Jackson, Solomon et al. 2009). Urinary natriuretic peptides have also been shown to have a diagnostic role in heart failure (Ng, Geeranavar et al. 2004). The urinary proteome is relatively unexplored in cardiovascular disease and warrants further investigation.

9.4 Future work

This project has generated a large amount of data that can be taken forwards to either better understand disease mechanisms or to evaluate the clinical role of biomarkers. The key areas I plan to focus on are discussed below.

9.4.1 Extending proteome coverage by studying ECM.

The extracellular matrix (ECM) of the heart provides structural and functional integrity by transmitting mechanical forces generated by cardiomyocytes to the cardiac cavities. It also contains cytokines and growth factors that modulate cardiac function and determine the fate of cardiac cells. Disruption of the ECM in various disease states compromises the structural integrity and function of the heart. The accumulation of ECM structural proteins in the form of fibrosis increases myocardial stiffness and contributes to diastolic and systolic dysfunction (Schaper, Mollnau et al. 1995). Both dilated and hypertrophic cardiomyopathies appear to continually form fibrotic tissue as the disease progresses, although the mechanism for this are not well understood (Kapelko 2001).

Applying LC-MS with specific ECM enrichment methods is likely to improve our understanding of the types and quality of the collagens involved in this disease. This research could be conducted in an animal model with tissue obtained at different stages of disease, or using human samples grouped according to the presence of replacement fibrosis on cardiac magnetic resonance imaging. A challenge will be obtaining tissue from early (or non-surgical) disease. Given that fibrosis is implicated early in the pathogenesis of HCM, better understanding the biology of the ECM is important to target treatment.

9.4.2 Understanding the role of CA III in the heart

The finding of elevated CA III in HCM was not expected. The differential expression analysis (Section 5.2.3) was felt to be robust because several unique peptides were identified, each with a significant fold change associated with a high confidence score for the protein. The peptide selected for inclusion in the validation assay did not align with any other isoforms of carbonic anhydrase. Furthermore, secondary

validation was performed using immunohistochemistry in a broad range of samples with a good positive control.

Although, CA III is abundant in skeletal muscle, adipocytes and liver, its exact function is not known. It has been implicated in augmenting fatty acid metabolism and may also have a role in the cellular response to oxidative stress when it is rapidly glutathionylated (Lyons, Buckingham et al. 1991, Cabiscol and Levine 1995). Inactivation of the CA III gene in mice did not produce any major phenotype; contractile properties of muscle and the carbonyl content (reflecting oxidative damage) of heart and other organs were unchanged. The mice were exposed to exercise stress, but no other cardiac stressors were examined (Kim, Lee et al. 2004). It is possible that CA III is required for an effective response to specific stimuli e.g. hypertrophy which could be studied pharmacologically or induced by aortic banding. The porcine CA III gene is associated with muscle fibre size and pork meat quality traits and has been studied for gene-assisted commercial breeding (Wimmers, Murani et al. 2007). Preliminary work in our cohort has also suggested that CA III is regulated at gene level and this may be a potential modifier of the HCM phenotype. Further studies are required to evaluate the role of CA III in the normal, failing and hypertrophic heart.

9.4.3 Large clinical biomarker studies

To determine the proteomic signature of disease it is necessary to identify subsets of patients by focusing on a specific phenotype. In HCM one of the major clinical challenges is in distinguishing those patients prone to arrhythmias or heart failure from others who follow an entirely benign clinical course. In addition, a mounting population of non-affected gene carriers has been identified through cascade screening with little knowledge about when and if they will develop a clinical phenotype. Longitudinal studies linked to genotype are required to answer these fundamental questions. Probably the best approach will be to study the plasma from patients with HCM and their relatives and follow them for a period of time, comparing the proteome of those who develop overt hypertrophy, arrhythmia and heart failure, against those who remain stable during follow-up. Urine is an attractive

bio fluid to investigate as it is easily attainable in large quantities and is more stable at $-20\text{ }^{\circ}\text{C}$ compared with plasma. Urinary peptides and low molecular weight proteins are also generally soluble and can be analysed by mass spectrometry without prior tryptic digestion. Urinary biomarkers would be particularly useful in children (who should have normal renal function) and would avoid the need for blood sampling.

10 Conclusion

This is the first comprehensive global proteomic study of human hypertrophic cardiomyopathy. It has identified differences in the expression of several proteins in the myocardium not described previously, but highly relevant to pathophysiology of this disease. Like other rare diseases, genetic cardiomyopathies provide a valuable model to study more common diseases. Pathways that lead to cardiomyocyte hypertrophy, arrhythmogenesis and ventricular remodelling are all relevant to understanding mechanisms of heart failure, which affects 1-2% people in the United Kingdom (McMurray and Stewart 2000, Stewart, Jenkins et al. 2002, BHF 2014).

Our knowledge of proteins has improved greatly over recent years. It is clear that a single gene does not produce a single protein, rather several gene products which may have different biological functions. Proteome research has been driven by developments in mass spectrometry and bioinformatics but there are still many technical, intellectual, and economic limitations to overcome. Datasets are increasingly large and acquired rapidly, but must be translated to meaningful biological outputs in order to benefit clinical practice. Because many proteins are drug targets it is hoped that proteomics will facilitate future drug discovery and development. Like genomics, proteomics requires a coordinated and collaborative international effort to reach its potential in scientific research.

11 References

- Abdellatif, M. (2000). "Leading the way using microarray: a more comprehensive approach for discovery of gene expression patterns." Circ Res **86**(9): 919-920.
- Abozguia, K., et al. (2010). "Metabolic modulator perhexiline corrects energy deficiency and improves exercise capacity in symptomatic hypertrophic cardiomyopathy." Circulation **122**(16): 1562-1569.
- Adzhubei, I. A., et al. (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-249.
- Aebersold, R. and M. Mann (2003). "Mass spectrometry-based proteomics." Nature **422**(6928): 198-207.
- Agbulut, O., et al. (1996). "Analysis of skeletal and cardiac muscle from desmin knock-out and normal mice by high resolution separation of myosin heavy-chain isoforms." Biol Cell **88**(3): 131-135.
- Ahmadian, A., et al. (2013). "Cardiac manifestations of subarachnoid hemorrhage." Heart Lung Vessel **5**(3): 168-178.
- Akki, A., et al. (2012). "Creatine kinase overexpression improves ATP kinetics and contractile function in postischemic myocardium." Am J Physiol Heart Circ Physiol **303**(7): H844-852.
- Alvarez, B. V., et al. (2007). "Carbonic anhydrase inhibition prevents and reverts cardiomyocyte hypertrophy." J Physiol **579**(Pt 1): 127-145.
- Alvarez, B. V., et al. (2013). "Quantification of carbonic anhydrase gene expression in ventricle of hypertrophic and failing human heart." BMC Cardiovasc Disord **13**: 2.
- Andersen, J. S., et al. (2003). "Proteomic characterization of the human centrosome by protein correlation profiling." Nature **426**(6966): 570-574.
- Arai, M., et al. (1994). "Sarcoplasmic reticulum gene expression in cardiac hypertrophy and heart failure." Circ Res **74**(4): 555-564.
- Aronow, B. J., et al. (2001). "Divergent transcriptional responses to independent genetic causes of cardiac hypertrophy." Physiol Genomics **6**(1): 19-28.
- Ashrafian, H., et al. (2003). "Hypertrophic cardiomyopathy: a paradigm for myocardial energy depletion." Trends Genet **19**(5): 263-268.

Askevold, E. T., et al. (2014). "The cardiokine secreted Frizzled-related protein 3, a modulator of Wnt signalling, in clinical and experimental heart failure." J Intern Med **275**(6): 621-630.

Authors/Task Force, m., et al. (2014). "2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC)." Eur Heart J **35**(39): 2733-2779.

Aye, T. T., et al. (2010). "Proteome-wide protein concentrations in the human heart." Mol Biosyst **6**(10): 1917-1927.

Baker, C. S., et al. (1992). "A human myocardial two-dimensional electrophoresis database: protein characterisation by microsequencing and immunoblotting." Electrophoresis **13**(9-10): 723-726.

Banerjee, P., et al. (2015). "Balancing functions of annexin A6 maintain equilibrium between hypertrophy and apoptosis in cardiomyocytes." Cell Death Dis **6**: e1873.

Banerjee, S. and S. Mazumdar (2012). "Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte." Int J Anal Chem **2012**: 282574.

Barallobre-Barreiro, J., et al. (2012). "Proteomics analysis of cardiac extracellular matrix remodeling in a porcine model of ischemia/reperfusion injury." Circulation **125**(6): 789-802.

Baumgartner, W. A., Jr., et al. (2007). "Manual curation is not sufficient for annotation of genomic databases." Bioinformatics **23**(13): i41-48.

Bayes-Genis, A., et al. (2014). "Head-to-head comparison of 2 myocardial fibrosis biomarkers for long-term heart failure risk stratification: ST2 versus galectin-3." J Am Coll Cardiol **63**(2): 158-166.

Becker, A. E. and G. Caruso (1982). "Myocardial disarray. A critical review." Br Heart J **47**(6): 527-538.

Becker, E., et al. (2007). "Quantification of mutant versus wild-type myosin in human muscle biopsies using nano-LC/ESI-MS." Anal Chem **79**(24): 9531-9538.

Beuerle, J. R., et al. (2000). "Characteristics of myoglobin, carbonic anhydrase III and the myoglobin/carbonic anhydrase III ratio in trauma, exercise, and myocardial infarction patients." Clin Chim Acta **294**(1-2): 115-128.

BHF (2014). Cardiovascular Disease Statistics 2014, British Heart Foundation.

Binder, J., et al. (2007). "Usefulness of brain natriuretic peptide levels in the clinical evaluation of patients with hypertrophic cardiomyopathy." Am J Cardiol **100**(4): 712-714.

Binder, J., et al. (2006). "Echocardiography-guided genetic testing in hypertrophic cardiomyopathy: septal morphological features predict the presence of myofilament mutations." Mayo Clin Proc **81**(4): 459-467.

Birney, E., et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Blair, E., et al. (2001). "Mutations in the gamma(2) subunit of AMP-activated protein kinase cause familial hypertrophic cardiomyopathy: evidence for the central role of energy compromise in disease pathogenesis." Hum Mol Genet **10**(11): 1215-1220.

Bostanci, N., et al. (2010). "Application of label-free absolute quantitative proteomics in human gingival crevicular fluid by LC/MS E (gingival exudatome)." J Proteome Res **9**(5): 2191-2199.

Bousette, N., et al. (2009). "Large-scale characterization and analysis of the murine cardiac proteome." J Proteome Res **8**(4): 1887-1901.

Burnett, J. C., Jr., et al. (1986). "Atrial natriuretic peptide elevation in congestive heart failure in the human." Science **231**(4742): 1145-1147.

Cabiscol, E. and R. L. Levine (1995). "Carbonic anhydrase III. Oxidative modification in vivo and loss of phosphatase activity during aging." J Biol Chem **270**(24): 14742-14747.

Cambronero, F., et al. (2009). "Biomarkers of pathophysiology in hypertrophic cardiomyopathy: implications for clinical management and prognosis." Eur Heart J **30**(2): 139-151.

Camelliti, P., et al. (2005). "Structural and functional characterisation of cardiac fibroblasts." Cardiovasc Res **65**(1): 40-51.

Camors, E., et al. (2005). "Annexins and Ca²⁺ handling in the heart." Cardiovasc Res **65**(4): 793-802.

Chapman, J. R. (1996). Protein and peptide analysis by mass spectrometry. Totowa, N.J., Humana Press.

Charron, P., et al. (2010). "Genetic counselling and testing in cardiomyopathies: a position statement of the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases." Eur Heart J **31**(22): 2715-2726.

Charron, P., et al. (1998). "Genotype-phenotype correlations in familial hypertrophic cardiomyopathy. A comparison between mutations in the cardiac protein-C and the beta-myosin heavy chain genes." Eur Heart J **19**(1): 139-145.

Chen, Y. R. and J. L. Zweier (2014). "Cardiac mitochondria and reactive oxygen species generation." Circ Res **114**(3): 524-537.

Chouinard, C. D., et al. (2016). "Ion Mobility in Clinical Analysis: Current Progress and Future Perspectives." Clin Chem **62**(1): 124-133.

Christodoulou, D. C., et al. (2014). "5'RNA-Seq identifies Fhl1 as a genetic modifier in cardiomyopathy." J Clin Invest **124**(3): 1364-1370.

Chu, P. H., et al. (2000). "Expression patterns of FHL/SLIM family members suggest important functional roles in skeletal muscle and cardiovascular system." Mech Dev **95**(1-2): 259-265.

Cieniewski-Bernard, C., et al. (2008). "Proteomic analysis in cardiovascular diseases." Clin Exp Pharmacol Physiol **35**(4): 362-366.

Coats, C. J. and P. M. Elliott (2013). "Genetic biomarkers in hypertrophic cardiomyopathy." Biomark Med **7**(4): 505-516.

Coats, C. J., et al. (2013). "Relation between serum N-terminal pro-brain natriuretic peptide and prognosis in patients with hypertrophic cardiomyopathy." Eur Heart J **34**(32): 2529-2537.

Coats, C. J., et al. (2015). "Current applications of biomarkers in cardiomyopathies." Expert Rev Cardiovasc Ther **13**(7): 825-837.

Coats, C. J. and A. Hollman (2008). "Hypertrophic cardiomyopathy: lessons from history." Heart **94**(10): 1258-1263.

Cohen Freue, G. V. and C. H. Borchers (2012). "Multiple reaction monitoring (MRM): principles and application to coronary artery disease." Circ Cardiovasc Genet **5**(3): 378.

Colan, S. D., et al. (2007). "Epidemiology and cause-specific outcome of hypertrophic cardiomyopathy in children: findings from the Pediatric Cardiomyopathy Registry." Circulation **115**(6): 773-781.

Connors, L. H., et al. (2016). "Heart Failure Resulting From Age-Related Cardiac Amyloid Disease Associated With Wild-Type Transthyretin: A Prospective, Observational Cohort Study." Circulation **133**(3): 282-290.

Consortium, I. H. G. S. (2004). "Finishing the euchromatic sequence of the human genome." Nature **431**(7011): 931-945.

- Copeland, O., et al. (2010). "Analysis of cardiac myosin binding protein-C phosphorylation in human heart muscle." J Mol Cell Cardiol **49**(6): 1003-1011.
- Cornwell, G. G., 3rd, et al. (1983). "Frequency and distribution of senile cardiovascular amyloid. A clinicopathologic correlation." Am J Med **75**(4): 618-623.
- Corthals, G. L., et al. (2000). "The dynamic range of protein expression: a challenge for proteomic research." Electrophoresis **21**(6): 1104-1115.
- Cowling, B. S., et al. (2008). "Identification of FHL1 as a regulator of skeletal muscle mass: implications for human myopathy." J Cell Biol **183**(6): 1033-1048.
- Craig, R., et al. (2004). "Open source system for analyzing, validating, and storing protein identification data." J Proteome Res **3**(6): 1234-1242.
- Crilley, J. G., et al. (2003). "Hypertrophic cardiomyopathy due to sarcomeric gene mutations is characterized by impaired energy metabolism irrespective of the degree of hypertrophy." J Am Coll Cardiol **41**(10): 1776-1782.
- D'Amati, G., et al. (1992). "Altered distribution of desmin filaments in hypertrophic cardiomyopathy: an immunohistochemical study." Mod Pathol **5**(2): 165-168.
- D'Eustachio, P. (2011). "Reactome knowledgebase of human biological pathways and processes." Methods Mol Biol **694**: 49-61.
- Datta, S., et al. (2015). "Laser capture microdissection: Big data from small samples." Histol Histopathol **30**(11): 1255-1269.
- Davies, M. J. and W. J. McKenna (1994). "Hypertrophic cardiomyopathy: an introduction to pathology and pathogenesis." Br Heart J **72**(6 Suppl): S2-3.
- Davila-Roman, V. G., et al. (2002). "Altered myocardial fatty acid and glucose metabolism in idiopathic dilated cardiomyopathy." J Am Coll Cardiol **40**(2): 271-277.
- de Lemos, J. A., et al. (2003). "B-type natriuretic peptide in cardiovascular disease." Lancet **362**(9380): 316-322.
- Debold, E. P., et al. (2007). "Hypertrophic and dilated cardiomyopathy mutations differentially affect the molecular force generation of mouse alpha-cardiac myosin in the laser trap assay." Am J Physiol Heart Circ Physiol **293**(1): H284-291.
- Didangelos, A., et al. (2010). "Proteomics characterization of extracellular space components in the human aorta." Mol Cell Proteomics **9**(9): 2048-2062.

- Dimitrow, P. P., et al. (2008). "Obstructive hypertrophic cardiomyopathy is associated with enhanced thrombin generation and platelet activation." Heart **94**(6): e21.
- Domanski, D., et al. (2012). "MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma." Proteomics **12**(8): 1222-1243.
- Dorn, G. W., et al. (2003). "Phenotyping hypertrophy - Eschew obfuscation." Circ Res **92**(11): 1171-1175.
- Driesen, R. B., et al. (2009). "Re-expression of alpha skeletal actin as a marker for dedifferentiation in cardiac pathologies." J Cell Mol Med **13**(5): 896-908.
- Durany, N., et al. (1996). "Activity of phosphoglycerate mutase and its isoenzymes in serum after acute myocardial infarction." Clin Mol Pathol **49**(5): M298-300.
- Elliott, P., et al. (2008). "Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases." Eur Heart J **29**(2): 270-276.
- Elliott, P. M., et al. (2014). "2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: The Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC)." Eur Heart J.
- Elliott, P. M., et al. (2006). "Historical trends in reported survival rates in patients with hypertrophic cardiomyopathy." Heart **92**(6): 785-791.
- Elliott, P. M., et al. (2006). "Left ventricular outflow tract obstruction and sudden death risk in patients with hypertrophic cardiomyopathy." Eur Heart J **27**(16): 1933-1941.
- Emdin, M., et al. (2009). "Old and new biomarkers of heart failure." Eur J Heart Fail **11**(4): 331-335.
- Engelbrechtsen, K. V., et al. (2013). "Lumican is increased in experimental and clinical heart failure, and its production by cardiac fibroblasts is induced by mechanical and proinflammatory stimuli." FEBS J **280**(10): 2382-2398.
- Engelbrechtsen, K. V., et al. (2013). "Decorin, lumican, and their GAG chain-synthesizing enzymes are regulated in myocardial remodeling and reverse remodeling in the mouse." J Appl Physiol (1985) **114**(8): 988-997.
- Fananapazir, L. and N. D. Epstein (1994). "Genotype-phenotype correlations in hypertrophic cardiomyopathy. Insights provided by comparisons of kindreds with distinct and identical beta-myosin heavy chain gene mutations." Circulation **89**(1): 22-32.

- Feng, H. and J. Jin (2015). "Transgenic over-expression of carbonic anhydrase III in cardiac muscle demonstrates a mechanism to resist acidosis." Biophysical Journal **108**(2, Supp 1): 595a.
- Fenn, J. B., et al. (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**(4926): 64-71.
- Francalanci, P., et al. (1995). "The pattern of desmin filaments in myocardial disarray." Hum Pathol **26**(3): 262-266.
- Frey, N., et al. (2004). "Hypertrophy of the heart: a new therapeutic target?" Circulation **109**(13): 1580-1589.
- Frey, N., et al. (2012). "Mechanisms of disease: hypertrophic cardiomyopathy." Nat Rev Cardiol **9**(2): 91-100.
- Friedrich, F. W., et al. (2012). "Evidence for FHL1 as a novel disease gene for isolated hypertrophic cardiomyopathy." Hum Mol Genet **21**(14): 3237-3254.
- Geisterfer-Lowrance, A. A., et al. (1990). "A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation." Cell **62**(5): 999-1006.
- Geoghegan, K. F., et al. (2012). "Deconstruction of activity-dependent covalent modification of heme in human neutrophil myeloperoxidase by multistage mass spectrometry (MS(4))." Biochemistry **51**(10): 2065-2077.
- Gerdes, A. M. (2002). "Cardiac myocyte remodeling in hypertrophy and progression to failure." J Card Fail **8**(6): S264-S268.
- Gersh, B. J., et al. (2011). "2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines." Circulation **124**(24): 2761-2796.
- Geske, J. B., et al. (2013). "B-type natriuretic peptide and survival in hypertrophic cardiomyopathy." J Am Coll Cardiol **61**(24): 2456-2460.
- Gladka, M. M., et al. (2012). "Small changes can make a big difference - microRNA regulation of cardiac hypertrophy." J Mol Cell Cardiol **52**(1): 74-82.
- Goldfarb, L. G., et al. (1998). "Missense mutations in desmin associated with familial cardiac and skeletal myopathy." Nat Genet **19**(4): 402-403.
- Goudarzi, M., et al. (2011). "Development of a novel proteomic approach for mitochondrial proteomics from cardiac tissue from patients with atrial fibrillation." J Proteome Res **10**(8): 3484-3492.

- Gramolini, A. O., et al. (2008). "Comparative proteomics profiling of a phospholamban mutant mouse model of dilated cardiomyopathy reveals progressive intracellular stress responses." Mol Cell Proteomics **7**(3): 519-533.
- Green, J. J., et al. (2012). "Prognostic value of late gadolinium enhancement in clinical outcomes for hypertrophic cardiomyopathy." JACC Cardiovasc Imaging **5**(4): 370-377.
- Group, B. D. W. (2001). "Biomarkers and surrogate endpoints: preferred definitions and conceptual framework." Clin Pharmacol Ther **69**(3): 89-95.
- Gupta, M. P. (2007). "Factors controlling cardiac myosin-isoform shift during hypertrophy and heart failure." J Mol Cell Cardiol **43**(4): 388-403.
- Gygi, S. P., et al. (1999). "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags." Nat Biotechnol **17**(10): 994-999.
- Gygi, S. P., et al. (1999). "Correlation between protein and mRNA abundance in yeast." Mol Cell Biol **19**(3): 1720-1730.
- Haim, T. E., et al. (2007). "Independent FHC-related cardiac troponin T mutations exhibit specific alterations in myocellular contractility and calcium kinetics." J Mol Cell Cardiol **42**(6): 1098-1110.
- Hammer, E., et al. (2011). "Characterization of the human myocardial proteome in inflammatory dilated cardiomyopathy by label-free quantitative shotgun proteomics of heart biopsies." J Proteome Res **10**(5): 2161-2171.
- Hargreaves, I. P., et al. (1999). "Mitochondrial respiratory chain defects are not accompanied by an increase in the activities of lactate dehydrogenase or manganese superoxide dismutase in paediatric skeletal muscle biopsies." J Inherit Metab Dis **22**(8): 925-931.
- Harju, A. K., et al. (2013). "Carbonic anhydrase III: a neglected isozyme is stepping into the limelight." J Enzyme Inhib Med Chem **28**(2): 231-239.
- Hayman, R., et al. (2000). "Desmin as a possible immunohistochemical marker for feline hypertrophic cardiomyopathy." J Vet Med Sci **62**(3): 343-346.
- Heinke, M. Y., et al. (1999). "Changes in myocardial protein expression in pacing-induced canine heart failure." Electrophoresis **20**(10): 2086-2093.
- Helms, A. S., et al. (2014). "Sarcomere mutation-specific expression patterns in human hypertrophic cardiomyopathy." Circ Cardiovasc Genet **7**(4): 434-443.
- Heywood, W. E., et al. (2011). "2D DIGE analysis of maternal plasma for potential biomarkers of Down Syndrome." Proteome Sci **9**: 56.

- Hinojar, R., et al. (2015). "T1 Mapping in Discrimination of Hypertrophic Phenotypes: Hypertensive Heart Disease and Hypertrophic Cardiomyopathy: Findings From the International T1 Multicenter Cardiovascular Magnetic Resonance Study." Circ Cardiovasc Imaging **8**(12).
- Hinz, B., et al. (2007). "The myofibroblast: one function, multiple origins." Am J Pathol **170**(6): 1807-1816.
- Ho, C. Y., et al. (2015). "Genetic advances in sarcomeric cardiomyopathies: state of the art." Cardiovasc Res **105**(4): 397-408.
- Ho, C. Y., et al. (2015). "Diltiazem treatment for pre-clinical hypertrophic cardiomyopathy sarcomere mutation carriers: a pilot randomized trial to modify disease expression." JACC Heart Fail **3**(2): 180-188.
- Ho, C. Y., et al. (2010). "Myocardial fibrosis as an early manifestation of hypertrophic cardiomyopathy." N Engl J Med **363**(6): 552-563.
- Ho, C. Y. and C. E. Seidman (2006). "A contemporary approach to hypertrophic cardiomyopathy." Circulation **113**(24): e858-862.
- Hogye, M., et al. (2004). "Comparison of circulating levels of interleukin-6 and tumor necrosis factor-alpha in hypertrophic cardiomyopathy and in idiopathic dilated cardiomyopathy." Am J Cardiol **94**(2): 249-251.
- Huang da, W., et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.
- Huttenhain, R., et al. (2012). "Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics." Sci Transl Med **4**(142): 142ra194.
- Huxley, A. F. and R. Niedergerke (1954). "Structural changes in muscle during contraction; interference microscopy of living muscle fibres." Nature **173**(4412): 971-973.
- Hwang, D. M., et al. (1997). "A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes." Circulation **96**(12): 4146-4203.
- Hwang, J. J., et al. (2002). "Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure." Physiol Genomics **10**(1): 31-44.
- Ioannidis, J. P. and O. A. Panagiotou (2011). "Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses." JAMA **305**(21): 2200-2210.

- Jackson, C. E., et al. (2009). "Albuminuria in chronic heart failure: prevalence and prognostic importance." Lancet **374**(9689): 543-550.
- Jacques, A. M., et al. (2008). "The molecular phenotype of human cardiac myosin associated with hypertrophic obstructive cardiomyopathy." Cardiovasc Res **79**(3): 481-491.
- Jaffe, A. S., et al. (2000). "It's time for a change to a troponin standard." Circulation **102**(11): 1216-1220.
- James, A. and C. Jorgensen (2010). "Basic design of MRM assays for peptide quantification." Methods Mol Biol **658**: 167-185.
- Jarcho, J. A., et al. (1989). "Mapping a gene for familial hypertrophic cardiomyopathy to chromosome 14q1." N Engl J Med **321**(20): 1372-1378.
- Jeffery, S. and N. Carter (1980). "Isoenzymes of human muscle carbonic anhydrase III generated by thiol interaction." Biochem Soc Trans **8**(5): 643.
- Jeffery, S., et al. (1980). "Distribution of CAIII in fetal and adult human tissue." Biochem Genet **18**(9-10): 843-849.
- Jiang, J., et al. (2013). "Allele-specific silencing of mutant Myh6 transcripts in mice suppresses hypertrophic cardiomyopathy." Science **342**(6154): 111-114.
- Jung, W. I., et al. (2000). "Differences in cardiac energetics between patients with familial and nonfamilial hypertrophic cardiomyopathy." Circulation **101**(12): E121.
- Kalsi, K. K., et al. (1999). "Energetics and function of the failing human heart with dilated or hypertrophic cardiomyopathy." Eur J Clin Invest **29**(6): 469-477.
- Kapelko, V. I. (2001). "Extracellular matrix alterations in cardiomyopathy: The possible crucial role in the dilative form." Exp Clin Cardiol **6**(1): 41-49.
- Karamanlidis, G., et al. (2011). "Impaired mitochondrial biogenesis precedes heart failure in right ventricular hypertrophy in congenital heart disease." Circ Heart Fail **4**(6): 707-713.
- Kato, T. S., et al. (2012). "Cardiac transplantation in patients with hypertrophic cardiomyopathy." Am J Cardiol **110**(4): 568-574.
- Keshishian, H., et al. (2009). "Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution." Mol Cell Proteomics **8**(10): 2339-2349.
- Khodiyar, V. K., et al. (2011). "The representation of heart development in the gene ontology." Dev Biol **354**(1): 9-17.

- Khoury, G. A., et al. (2011). "Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database." Sci Rep **1**.
- Kim, G., et al. (2004). "Carbonic anhydrase III is not required in the mouse for normal growth, development, and life span." Mol Cell Biol **24**(22): 9942-9947.
- Kim, M. S., et al. (2014). "A draft map of the human proteome." Nature **509**(7502): 575-581.
- Kislinger, T., et al. (2006). "Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling." Cell **125**(1): 173-186.
- Kline, K. G., et al. (2008). "High quality catalog of proteotypic peptides from human heart." J Proteome Res **7**(11): 5055-5061.
- Knowlton, A. A., et al. (1998). "Differential expression of heat shock proteins in normal and failing human hearts." J Mol Cell Cardiol **30**(4): 811-818.
- Kolwicz, S. C., Jr. and R. Tian (2011). "Glucose metabolism and cardiac hypertrophy." Cardiovasc Res **90**(2): 194-201.
- Kong, S. W., et al. (2005). "Genetic expression profiles during physiological and pathological cardiac hypertrophy and heart failure in rats." Physiol Genomics **21**(1): 34-42.
- Krishnan, J., et al. (2009). "Activation of a HIF1alpha-PPARgamma axis underlies the integration of glycolytic and lipid anabolic pathways in pathologic cardiac hypertrophy." Cell Metab **9**(6): 512-524.
- Kubo, T., et al. (2013). "Significance of high-sensitivity cardiac troponin T in hypertrophic cardiomyopathy." J Am Coll Cardiol **62**(14): 1252-1259.
- Kuster, D. W., et al. (2013). "MicroRNA transcriptome profiling in cardiac tissue of hypertrophic cardiomyopathy patients with MYBPC3 mutations." J Mol Cell Cardiol **65**: 59-66.
- Kuusisto, J., et al. (2012). "Low-grade inflammation and the phenotypic expression of myocardial fibrosis in hypertrophic cardiomyopathy." Heart **98**(13): 1007-1013.
- Kwapiszewska, G., et al. (2008). "Fhl-1, a new key protein in pulmonary hypertension." Circulation **118**(11): 1183-1194.
- Ladue, J. S., et al. (1954). "Serum glutamic oxaloacetic transaminase activity in human acute transmural myocardial infarction." Science **120**(3117): 497-499.

- Lam, L., et al. (2010). "Differential protein expression profiling of myocardial tissue in a mouse model of hypertrophic cardiomyopathy." J Mol Cell Cardiol **48**(5): 1014-1022.
- Lan, F., et al. (2013). "Abnormal calcium handling properties underlie familial hypertrophic cardiomyopathy pathology in patient-specific induced pluripotent stem cells." Cell Stem Cell **12**(1): 101-113.
- Lazarides, E. and B. D. Hubbard (1976). "Immunological characterization of the subunit of the 100 A filaments from muscle cells." Proc Natl Acad Sci U S A **73**(12): 4344-4348.
- Lele, S. S., et al. (1995). "Exercise capacity in hypertrophic cardiomyopathy. Role of stroke volume limitation, heart rate, and diastolic filling characteristics." Circulation **92**(10): 2886-2894.
- Levin, D. S., et al. (2007). "Using a nanoelectrospray-differential mobility spectrometer-mass spectrometer system for the analysis of oligosaccharides with solvent selected control over ESI aggregate ion formation." J Am Soc Mass Spectrom **18**(3): 502-511.
- Li, H., et al. (2012). "Proteomic analysis of effluents from perfused human heart for transplantation: identification of potential biomarkers for ischemic heart damage." Proteome Sci **10**(1): 21.
- Li, W., et al. (2012). "Proteomic analysis of metabolic, cytoskeletal and stress response proteins in human heart failure." J Cell Mol Med **16**(1): 59-71.
- Lie-Venema, H., et al. (2007). "Origin, fate, and function of epicardium-derived cells (EPDCs) in normal and abnormal cardiac development." ScientificWorldJournal **7**: 1777-1798.
- Lim, D. S., et al. (2001). "Expression profiling of cardiac genes in human hypertrophic cardiomyopathy: insight into the pathogenesis of phenotypes." J Am Coll Cardiol **38**(4): 1175-1180.
- Lindsey, M. L., et al. (2015). "Transformative Impact of Proteomics on Cardiovascular Health and Disease: A Scientific Statement From the American Heart Association." Circulation.
- Linhart, A. and P. M. Elliott (2007). "The heart in Anderson-Fabry disease and other lysosomal storage disorders." Heart **93**(4): 528-535.
- Liu, N. and E. N. Olson (2010). "MicroRNA regulatory networks in cardiovascular development." Dev Cell **18**(4): 510-525.

- Lombardi, R., et al. (2003). "Myocardial collagen turnover in hypertrophic cardiomyopathy." Circulation **108**(12): 1455-1460.
- Lopaschuk, G. D., et al. (1992). "Developmental changes in energy substrate use by the heart." Cardiovasc Res **26**(12): 1172-1180.
- Lopaschuk, G. D., et al. (2010). "Myocardial fatty acid metabolism in health and disease." Physiol Rev **90**(1): 207-258.
- Lopes, L. R., et al. (2015). "Novel genotype-phenotype associations demonstrated by high-throughput sequencing in patients with hypertrophic cardiomyopathy." Heart **101**(4): 294-301.
- Lopes, L. R., et al. (2013). "Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing." J Med Genet **50**(4): 228-239.
- Lovering, R. C., et al. (2009). "Improvements to cardiovascular gene ontology." Atherosclerosis **205**(1): 9-14.
- Lyons, G. E., et al. (1991). "Carbonic anhydrase III, an early mesodermal marker, is expressed in embryonic mouse skeletal muscle and notochord." Development **111**(1): 233-244.
- Maher, B. (2012). "ENCODE: The human encyclopaedia." Nature **489**(7414): 46-48.
- Maisel, A. S., et al. (2002). "Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure." N Engl J Med **347**(3): 161-167.
- Makarov, A. (2000). "Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis." Anal Chem **72**(6): 1156-1162.
- Manwaring, V., et al. (2013). "The identification of new biomarkers for identifying and monitoring kidney disease and their translation into a rapid mass spectrometry-based test: evidence of presymptomatic kidney disease in pediatric Fabry and type-I diabetic patients." J Proteome Res **12**(5): 2013-2021.
- Maron, B. J., et al. (1995). "Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary Artery Risk Development in (Young) Adults." Circulation **92**(4): 785-789.
- Maron, B. J. and M. S. Maron (2013). "Hypertrophic cardiomyopathy." Lancet **381**(9862): 242-255.
- Maron, B. J. and W. C. Roberts (1979). "Quantitative analysis of cardiac muscle cell disorganization in the ventricular septum of patients with hypertrophic cardiomyopathy." Circulation **59**(4): 689-706.

- Maron, B. J., et al. (2007). "Implantable cardioverter-defibrillators and prevention of sudden cardiac death in hypertrophic cardiomyopathy." JAMA **298**(4): 405-412.
- Maron, B. J., et al. (2004). "Usefulness of B-type natriuretic peptide assay in the assessment of symptomatic state in hypertrophic cardiomyopathy." Circulation **109**(8): 984-989.
- Maron, M. S., et al. (2009). "Hypertrophic cardiomyopathy phenotype revisited after 50 years with cardiovascular magnetic resonance." J Am Coll Cardiol **54**(3): 220-228.
- Maron, M. S., et al. (2003). "Effect of left ventricular outflow tract obstruction on clinical outcome in hypertrophic cardiomyopathy." N Engl J Med **348**(4): 295-303.
- Marshall, K. D., et al. (2014). "Proteomic mapping of proteins released during necrosis and apoptosis from cultured neonatal cardiac myocytes." Am J Physiol Cell Physiol **306**(7): C639-647.
- Marston, S., et al. (2012). "How do MYBPC3 mutations cause hypertrophic cardiomyopathy?" J Muscle Res Cell Motil **33**(1): 75-80.
- Marston, S., et al. (2009). "Evidence from human myectomy samples that MYBPC3 mutations cause hypertrophic cardiomyopathy through haploinsufficiency." Circ Res **105**(3): 219-222.
- Masson, S., et al. (2012). "Serial measurement of cardiac troponin T using a highly sensitive assay in patients with chronic heart failure: data from 2 large randomized clinical trials." Circulation **125**(2): 280-288.
- Mayr, M., et al. (2009). "Proteomics, metabolomics, and immunomics on microparticles derived from human atherosclerotic plaques." Circ Cardiovasc Genet **2**(4): 379-388.
- McBride, H. M., et al. (2006). "Mitochondria: more than just a powerhouse." Curr Biol **16**(14): R551-560.
- McGrath, M. J., et al. (2006). "Four and a half LIM protein 1 binds myosin-binding protein C and regulates myosin filament formation and sarcomere assembly." J Biol Chem **281**(11): 7666-7683.
- McGregor, E. and M. J. Dunn (2003). "Proteomics of heart disease." Hum Mol Genet **12 Spec No 2**: R135-144.
- McGregor, E. and M. J. Dunn (2006). "Proteomics of the heart: unraveling disease." Circ Res **98**(3): 309-321.

- McKenna, W. J., et al. (1997). "Experience from clinical genetics in hypertrophic cardiomyopathy: proposal for new diagnostic criteria in adult members of affected families." Heart **77**(2): 130-132.
- McMurray, J. J. and S. Stewart (2000). "Epidemiology, aetiology, and prognosis of heart failure." Heart **83**(5): 596-602.
- Melacini, P., et al. (2010). "Clinicopathological profiles of progressive heart failure in hypertrophic cardiomyopathy." Eur Heart J **31**(17): 2111-2123.
- Merril, C. R., et al. (1984). "Gel protein stains: silver stain." Methods Enzymol **104**: 441-447.
- Mills, K., et al. (2005). "Measurement of urinary CDH and CTH by tandem mass spectrometry in patients hemizygous and heterozygous for Fabry disease." J Inherit Metab Dis **28**(1): 35-48.
- Mogensen, J., et al. (2004). "Frequency and clinical expression of cardiac troponin I mutations in 748 consecutive families with hypertrophic cardiomyopathy." J Am Coll Cardiol **44**(12): 2315-2325.
- Montoro-Garcia, S., et al. (2012). "Growth differentiation factor-15, a novel biomarker related with disease severity in patients with hypertrophic cardiomyopathy." Eur J Intern Med **23**(2): 169-174.
- Moon, J. C., et al. (2003). "Toward clinical risk assessment in hypertrophic cardiomyopathy with gadolinium cardiovascular magnetic resonance." J Am Coll Cardiol **41**(9): 1561-1567.
- Moreno, V., et al. (2010). "Serum levels of high-sensitivity troponin T: a novel marker for cardiac remodeling in hypertrophic cardiomyopathy." J Card Fail **16**(12): 950-956.
- Morita, H., et al. (2008). "Shared genetic causes of cardiac hypertrophy in children and adults." N Engl J Med **358**(18): 1899-1908.
- Nag, S., et al. (2013). "Gelsolin: the tail of a molecular gymnast." Cytoskeleton (Hoboken) **70**(7): 360-384.
- Nascimben, L., et al. (2004). "Mechanisms for increased glycolysis in the hypertrophied rat heart." Hypertension **44**(5): 662-667.
- Neubauer, S. (2007). "The failing heart--an engine out of fuel." N Engl J Med **356**(11): 1140-1151.
- Neubauer, S., et al. (1997). "Myocardial phosphocreatine-to-ATP ratio is a predictor of mortality in patients with dilated cardiomyopathy." Circulation **96**(7): 2190-2196.

- Neubauer, S., et al. (1997). "Cardiac high-energy phosphate metabolism in patients with aortic valve disease assessed by ³¹P-magnetic resonance spectroscopy." J Investig Med **45**(8): 453-462.
- Ng, L. L., et al. (2004). "Diagnosis of heart failure using urinary natriuretic peptides." Clin Sci (Lond) **106**(2): 129-133.
- Ng, P. C. and S. Henikoff (2003). "SIFT: Predicting amino acid changes that affect protein function." Nucleic Acids Res **31**(13): 3812-3814.
- Niimura, H., et al. (2002). "Sarcomere protein gene mutations in hypertrophic cardiomyopathy of the elderly." Circulation **105**(4): 446-451.
- Nunez, L., et al. (2013). "Somatic MYH7, MYBPC3, TPM1, TNNT2 and TNNI3 mutations in sporadic hypertrophic cardiomyopathy." Circ J **77**(9): 2358-2365.
- O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins." J Biol Chem **250**(10): 4007-4021.
- O'Hanlon, R., et al. (2010). "Prognostic significance of myocardial fibrosis in hypertrophic cardiomyopathy." J Am Coll Cardiol **56**(11): 867-874.
- O'Mahony, C., et al. (2014). "A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD)." Eur Heart J **35**(30): 2010-2020.
- Obayashi, T., et al. (1992). "Point mutations in mitochondrial DNA in patients with hypertrophic cardiomyopathy." Am Heart J **124**(5): 1263-1269.
- Oliver, I. T. (1955). "A spectrophotometric method for the determination of creatine phosphokinase and myokinase." Biochemical Journal **61**(1): 116-122.
- Olivotto, I., et al. (2009). "Developmental origins of hypertrophic cardiomyopathy phenotypes: a unifying hypothesis." Nat Rev Cardiol **6**(4): 317-321.
- Ommen, S. R., et al. (2005). "Long-term effects of surgical septal myectomy on survival in patients with obstructive hypertrophic cardiomyopathy." J Am Coll Cardiol **46**(3): 470-476.
- Page, S. P., et al. (2012). "Cardiac myosin binding protein-C mutations in families with hypertrophic cardiomyopathy: disease expression in relation to age, gender, and long term outcome." Circ Cardiovasc Genet **5**(2): 156-166.
- Palacin, M., et al. (2011). "Profile of microRNAs differentially produced in hearts from patients with hypertrophic cardiomyopathy and sarcomeric mutations." Clin Chem **57**(11): 1614-1616.

- Paya, E., et al. (2008). "Variables associated with contrast-enhanced cardiovascular magnetic resonance in hypertrophic cardiomyopathy: clinical implications." J Card Fail **14**(5): 414-419.
- Peng, Y., et al. (2014). "Top-down mass spectrometry of cardiac myofilament proteins in health and disease." Proteomics Clin Appl **8**(7-8): 554-568.
- Penhoet, E., et al. (1967). "The subunit structure of mammalian fructose diphosphate aldolase." Biochemistry **6**(9): 2940-2949.
- Pieroni, M., et al. (2007). "Increased brain natriuretic peptide secretion is a marker of disease progression in nonobstructive hypertrophic cardiomyopathy." J Card Fail **13**(5): 380-388.
- Polden, J., et al. (2011). "A 2-D gel reference map of the basic human heart proteome." Proteomics **11**(17): 3582-3586.
- Rajan, S., et al. (2006). "Microarray analysis of gene expression during early stages of mild and severe cardiac hypertrophy." Physiol Genomics **27**(3): 309-317.
- Ramirez-Correa, G. A., et al. (2014). "Targeted proteomics of myofilament phosphorylation and other protein posttranslational modifications." Proteomics Clin Appl **8**(7-8): 543-553.
- Rapezzi, C., et al. (2013). "Diagnostic work-up in cardiomyopathies: bridging the gap between clinical phenotypes and final diagnosis. A position statement from the ESC Working Group on Myocardial and Pericardial Diseases." Eur Heart J **34**(19): 1448-1458.
- Ravenscroft, G., et al. (2008). "Expression of cardiac alpha-actin spares extraocular muscles in skeletal muscle alpha-actin diseases--quantification of striated alpha-actins by MRM-mass spectrometry." Neuromuscul Disord **18**(12): 953-958.
- Razeghi, P., et al. (2001). "Metabolic gene expression in fetal and failing human heart." Circulation **104**(24): 2923-2931.
- Relman, A. S., et al. (1954). "Oral administration of a potent carbonic anhydrase inhibitor (Diamox). II. Its use as a diuretic in patients with severe congestive heart failure." N Engl J Med **250**(19): 800-804.
- Rockey, D. C., et al. (2015). "Fibrosis--a common pathway to organ injury and failure." N Engl J Med **372**(12): 1138-1149.
- Ross, J. S., et al. (2004). "Targeted therapy in breast cancer: the HER-2/neu gene and protein." Mol Cell Proteomics **3**(4): 379-398.

- Ross, P. L., et al. (2004). "Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents." Mol Cell Proteomics **3**(12): 1154-1169.
- Ruberg, F. L. and J. L. Berk (2012). "Transthyretin (TTR) cardiac amyloidosis." Circulation **126**(10): 1286-1300.
- Salameh, J., et al. (2013). "Phosphoglycerate mutase deficiency with tubular aggregates in a patient from Panama." Muscle Nerve **47**(1): 138-140.
- Sata, M. and M. Ikebe (1996). "Functional analysis of the mutations in the human cardiac beta-myosin that are responsible for familial hypertrophic cardiomyopathy. Implication for the clinical outcome." J Clin Invest **98**(12): 2866-2873.
- Schaper, J., et al. (1995). "[Interactions between cardiomyocytes and extracellular matrix in the failing human heart]." Z Kardiol **84 Suppl 4**: 33-38.
- Scheibe, R. J., et al. (2006). "Expression of membrane-bound carbonic anhydrases IV, IX, and XIV in the mouse heart." J Histochem Cytochem **54**(12): 1379-1391.
- Schneider, M. D. and R. J. Schwartz (2000). "Chips ahoy: gene expression in failing hearts surveyed by high-density microarrays." Circulation **102**(25): 3026-3027.
- Seger, C., et al. (2013). "Mass spectrometry and NMR spectroscopy: modern high-end detectors for high resolution separation techniques--state of the art in natural product HPLC-MS, HPLC-NMR, and CE-MS hyphenations." Nat Prod Rep **30**(7): 970-987.
- Seidman, C. E. and J. G. Seidman (2011). "Identifying sarcomere gene mutations in hypertrophic cardiomyopathy: a personal history." Circ Res **108**(6): 743-750.
- Seidman, J. G. and C. Seidman (2001). "The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms." Cell **104**(4): 557-567.
- Semsarian, C., et al. (2002). "The L-type calcium channel inhibitor diltiazem prevents cardiomyopathy in a mouse model." J Clin Invest **109**(8): 1013-1020.
- Shao, B. and J. W. Heinecke (2011). "Impact of HDL oxidation by the myeloperoxidase system on sterol efflux by the ABCA1 pathway." J Proteomics **74**(11): 2289-2299.
- Sharma, P., et al. (2013). "Recent advances in cardiovascular proteomics." J Proteomics **81**: 3-14.
- Shevchenko, A., et al. (1996). "Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels." Anal Chem **68**(5): 850-858.

Shirani, J., et al. (2000). "Morphology and significance of the left ventricular collagen network in young patients with hypertrophic cardiomyopathy and sudden cardiac death." J Am Coll Cardiol **35**(1): 36-44.

Siddiqi, N., et al. (2013). "Cardiac metabolism in hypertrophy and heart failure: implications for therapy." Heart Fail Rev **18**(5): 595-606.

Sly, W. S. and P. Y. Hu (1995). "Human carbonic anhydrases and carbonic anhydrase deficiencies." Annu Rev Biochem **64**: 375-401.

Somura, F., et al. (2001). "Reduced myocardial sarcoplasmic reticulum Ca(2+)-ATPase mRNA expression and biphasic force-frequency relations in patients with hypertrophic cardiomyopathy." Circulation **104**(6): 658-663.

Song, L., et al. (2014). "MiR-451 is decreased in hypertrophic cardiomyopathy and regulates autophagy by targeting TSC1." J Cell Mol Med.

Spinale, F. G. (2007). "Myocardial matrix remodeling and the matrix metalloproteinases: influence on cardiac form and function." Physiol Rev **87**(4): 1285-1342.

Spindler, M., et al. (1998). "Diastolic dysfunction and altered energetics in the alphaMHC403/+ mouse model of familial hypertrophic cardiomyopathy." J Clin Invest **101**(8): 1775-1783.

Spudich, J. A. (2014). "Hypertrophic and dilated cardiomyopathy: four decades of basic research on muscle lead to potential therapeutic approaches to these devastating genetic diseases." Biophys J **106**(6): 1236-1249.

Spudich, J. A. (2015). "The myosin mesa and a possible unifying hypothesis for the molecular basis of human hypertrophic cardiomyopathy." Biochem Soc Trans **43**(1): 64-72.

Stanley, W. C., et al. (2005). "Myocardial substrate metabolism in the normal and failing heart." Physiol Rev **85**(3): 1093-1129.

Steen, H. and M. Mann (2004). "The ABC's (and XYZ's) of peptide sequencing." Nat Rev Mol Cell Biol **5**(9): 699-711.

Stewart, S., et al. (2002). "The current cost of heart failure to the National Health Service in the UK." Eur J Heart Fail **4**(3): 361-371.

Sundstrom, J., et al. (2009). "Cardiac troponin-I and risk of heart failure: a community-based cohort study." Eur Heart J **30**(7): 773-781.

Supuran, C. T. and A. Scozzafava (2007). "Carbonic anhydrases as targets for medicinal chemistry." Bioorg Med Chem **15**(13): 4336-4350.

- Suurmeijer, A. J., et al. (2003). "Alpha-actin isoform distribution in normal and failing human heart: a morphological, morphometric, and biochemical study." J Pathol **199**(3): 387-397.
- Swenson, E. R. (1997). "Carbonic anhydrase and the heart." Cardiologia **42**(5): 453-462.
- Syrjala, H., et al. (1990). "Carbonic anhydrase III as a serum marker for diagnosis of rhabdomyolysis." Clin Chem **36**(4): 696.
- Taegtmeyer, H., et al. (2010). "Return to the fetal gene program: a suggested metabolic link to gene expression in the heart." Ann N Y Acad Sci **1188**: 191-198.
- Takimoto, E. and D. A. Kass (2007). "Role of oxidative stress in cardiac hypertrophy and remodeling." Hypertension **49**(2): 241-248.
- Tan, F. L., et al. (2002). "The gene expression fingerprint of human heart failure." Proc Natl Acad Sci U S A **99**(17): 11387-11392.
- Tanaka, M., et al. (1986). "Quantitative analysis of myocardial fibrosis in normals, hypertensive hearts, and hypertrophic cardiomyopathy." Br Heart J **55**(6): 575-581.
- Tang, W. H., et al. (2011). "Plasma myeloperoxidase predicts incident cardiovascular risks in stable patients undergoing medical management for coronary artery disease." Clin Chem **57**(1): 33-39.
- Taylor, C. F., et al. (2007). "The minimum information about a proteomics experiment (MIAPE)." Nat Biotechnol **25**(8): 887-893.
- Taylor, S. W., et al. (2003). "Characterization of the human heart mitochondrial proteome." Nat Biotechnol **21**(3): 281-286.
- Teare, D. (1958). "Asymmetrical hypertrophy of the heart in young adults." Br Heart J **20**(1): 1-8.
- Theis, J. L., et al. (2009). "Expression patterns of cardiac myofilament proteins: genomic and protein analysis of surgical myectomy tissue from patients with obstructive hypertrophic cardiomyopathy." Circ Heart Fail **2**(4): 325-333.
- Thomas, P. D., et al. (2003). "PANTHER: a library of protein families and subfamilies indexed by function." Genome Res **13**(9): 2129-2141.
- Thomas, P. D., et al. (2006). "Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools." Nucleic Acids Res **34**(Web Server issue): W645-650.

- Thompson, A., et al. (2003). "Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS." Anal Chem **75**(8): 1895-1904.
- Thomson, J. J. (1921). Rays of Positive Electricity and their application to Chemical Analyses. [S.l.], Longmans, Green and Co.
- Thygesen, K., et al. (2012). "Third universal definition of myocardial infarction." Circulation **126**(16): 2020-2035.
- Tonge, R., et al. (2001). "Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology." Proteomics **1**(3): 377-396.
- Torella, D., et al. (2014). "Carbonic anhydrase activation is associated with worsened pathological remodeling in human ischemic diabetic cardiomyopathy." J Am Heart Assoc **3**(2): e000434.
- Tripathi, S., et al. (2011). "Unequal allelic expression of wild-type and mutated beta-myosin in familial hypertrophic cardiomyopathy." Basic Res Cardiol **106**(6): 1041-1055.
- Unlu, M., et al. (1997). "Difference gel electrophoresis: a single gel method for detecting changes in protein extracts." Electrophoresis **18**(11): 2071-2077.
- Unno, K., et al. (2009). "Relation of functional and morphological changes in mitochondria to myocardial contractile and relaxation reserves in asymptomatic to mildly symptomatic patients with hypertrophic cardiomyopathy." Eur Heart J **30**(15): 1853-1862.
- Vaananen, H. K., et al. (1990). "Serum carbonic anhydrase III and myoglobin concentrations in acute myocardial infarction." Clin Chem **36**(4): 635-638.
- Vaananen, H. K., et al. (1988). "Muscle-specific carbonic anhydrase III is a more sensitive marker of muscle damage than creatine kinase in neuromuscular disorders." Arch Neurol **45**(11): 1254-1256.
- Valente, M. A., et al. (2012). "Urinary proteins in heart failure." Prog Cardiovasc Dis **55**(1): 44-55.
- van Dijk, S. J., et al. (2009). "Cardiac myosin-binding protein C mutations and hypertrophic cardiomyopathy: haploinsufficiency, deranged phosphorylation, and cardiomyocyte dysfunction." Circulation **119**(11): 1473-1483.
- van Rooij, E., et al. (2006). "A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure." Proc Natl Acad Sci U S A **103**(48): 18255-18260.

Vandekerckhove, J. and K. Weber (1978). "Mammalian cytoplasmic actins are the products of at least two genes and differ in primary structure in at least 25 identified positions from skeletal muscle actins." Proc Natl Acad Sci U S A **75**(3): 1106-1110.

Varnava, A. M., et al. (2001). "Hypertrophic cardiomyopathy: histopathological features of sudden death in cardiac troponin T disease." Circulation **104**(12): 1380-1384.

Varnava, A. M., et al. (2000). "Hypertrophic cardiomyopathy: the interrelation of disarray, fibrosis, and small vessel disease." Heart **84**(5): 476-482.

Vliegen, H. W., et al. (1991). "Myocardial changes in pressure overload-induced left ventricular hypertrophy. A study on tissue composition, polyploidization and multinucleation." Eur Heart J **12**(4): 488-494.

Vogel, C. and E. M. Marcotte (2012). "Insights into the regulation of protein abundance from proteomic and transcriptomic analyses." Nat Rev Genet **13**(4): 227-232.

Vrana, J. A., et al. (2009). "Classification of amyloidosis by laser microdissection and mass spectrometry-based proteomic analysis in clinical biopsy specimens." Blood **114**(24): 4957-4959.

Vuotikka, P., et al. (2003). "Serum myoglobin/carbonic anhydrase III ratio as a marker of reperfusion after myocardial infarction." Int J Cardiol **91**(2-3): 137-144.

Wasinger, V. C., et al. (1995). "Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*." Electrophoresis **16**(7): 1090-1094.

Watkins, H., et al. (1995). "Mutations in the cardiac myosin binding protein-C gene on chromosome 11 cause familial hypertrophic cardiomyopathy." Nat Genet **11**(4): 434-437.

Watkins, H., et al. (1992). "Characteristics and prognostic implications of myosin missense mutations in familial hypertrophic cardiomyopathy." N Engl J Med **326**(17): 1108-1114.

Westbrook, J. A., et al. (2006). "The human heart proteome: Two-dimensional maps using narrow-range immobilised pH gradients." Electrophoresis **27**(8): 1547-1555.

Westbrook, J. A., et al. (2001). "Zooming-in on the proteome: very narrow-range immobilised pH gradients reveal more protein species and isoforms." Electrophoresis **22**(14): 2865-2871.

Westermann, D., et al. (2006). "Diltiazem treatment prevents diastolic heart failure in mice with familial hypertrophic cardiomyopathy." Eur J Heart Fail **8**(2): 115-121.

Wigle, E. D., et al. (1985). "Hypertrophic cardiomyopathy. The importance of the site and the extent of hypertrophy. A review." Prog Cardiovasc Dis **28**(1): 1-83.

Wilhelm, M., et al. (2014). "Mass-spectrometry-based draft of the human proteome." Nature **509**(7502): 582-587.

Wimmers, K., et al. (2007). "Associations of functional candidate genes derived from gene-expression profiles of prenatal porcine muscle tissue with meat quality and muscle deposition." Anim Genet **38**(5): 474-484.

Witjas-Paalberends, E. R., et al. (2014). "Gene-specific increase in the energetic cost of contraction in hypertrophic cardiomyopathy caused by thick filament mutations." Cardiovasc Res **103**(2): 248-257.

Woo, A., et al. (2003). "Mutations of the beta myosin heavy chain gene in hypertrophic cardiomyopathy: critical functional sites determine prognosis." Heart **89**(10): 1179-1185.

Wroblewski, F., et al. (1956). "Serum lactic dehydrogenase activity in acute transmural myocardial infarction." Science **123**(3208): 1122-1123.

Yang, J., et al. (2000). "Decreased SLIM1 expression and increased gelsolin expression in failing human hearts measured by high-density oligonucleotide arrays." Circulation **102**(25): 3046-3052.

Zen, K., et al. (2005). "Analysis of circulating apoptosis mediators and proinflammatory cytokines in patients with idiopathic hypertrophic cardiomyopathy: comparison between nonobstructive and dilated-phase hypertrophic cardiomyopathy." Int Heart J **46**(2): 231-244.

Zhao, Y. Y., et al. (2014). "UPLC-based metabolomic applications for discovering biomarkers of diseases in clinical chemistry." Clin Biochem **47**(15): 16-26.

Zimmermann, O., et al. (2009). "Myocardial inflammation and non-ischaemic heart failure: is there a role for C-reactive protein?" Basic Res Cardiol **104**(5): 591-599.

Zong, N. C., et al. (2013). "Integration of cardiac proteome biology and medicine by a specialized knowledgebase." Circ Res **113**(9): 1043-1053.

12 Appendix

This section details the significantly differentially expressed proteins of whole cardiac tissue lysate quantified by label-free mass spectrometry in HCM compared with controls. The number of peptides per protein is listed with unique peptides used for quantitation shown in parentheses where applicable. The score provides a measure of confidence of accurate protein identification. Fold change and significance levels are shown.

a) Proteins upregulated in HCM

Accession	Protein Description	Gene	Peptides	Score	Anova (p)*	Fold
RSSA_HUMAN	40S ribosomal protein SA	RPSA	2 (1)	7.37	0.05	1.22
ASAH1_HUMAN	Acid ceramidase	ASAH1	6	40.5	0.02	1.28
ACTS_HUMAN	Actin alpha skeletal muscle	ACTA1	146 (7)	531.74	0.05	1.41
ADT3_HUMAN	ADP ATP translocase 3	SLC25A6	52 (7)	218.69	0.05	1.45
FETUA_HUMAN	Alpha 2 HS glycoprotein	AHSG	1	6.81	0.01	2.16
ACTN2_HUMAN	Alpha actinin 2	ACTN2	147 (89)	817.23	0.05	1.59
AATM_HUMAN	Aspartate aminotransferase mitochondrial	GOT2	139 (111)	549.85	0.05	1.16
ATIF1_HUMAN	ATPase inhibitor mitochondrial	ATPIF1	3 (2)	13.09	0.02	1.71
ENOB_HUMAN	Beta enolase	ENO3	59 (32)	328.48	0.05	1.24
BICR2_HUMAN	Bicaudal D related protein 2	CCDC64B	1	6.3	0.04	3.05
CAH3_HUMAN	Carbonic anhydrase 3	CA3	7	45.61	1.52E-03	4.69
COF2_HUMAN	Cofilin 2	CFL2	26 (17)	170.61	7.32E-04	1.2
CSRP3_HUMAN	Cysteine and glycine rich protein 3	CSRP3	40 (24)	155.24	0.04	1.23
ECHM_HUMAN	Enoyl CoA hydratase mitochondrial	ECHS1	26 (19)	159.29	0.04	1.18
FHL1_HUMAN	Four and a half LIM domains protein 1	FHL1	37 (30)	222.42	1.22E-03	1.49
FUMH_HUMAN	Fumarate hydratase mitochondrial	FH	38 (28)	157.06	4.11E-03	1.23
ENOG_HUMAN	Gamma enolase	ENO2	40 (15)	227.9	3.89E-03	1.35
G6PI_HUMAN	Glucose 6 phosphate isomerase	GPI	47 (37)	242.23	0.01	1.32
HSP76_HUMAN	Heat shock 70 kDa protein 6	HSPA6	43 (8)	232.06	0.02	1.56
HSPB2_HUMAN	Heat shock protein beta 2	HSPB2	6 (4)	33.96	3.37E-04	1.73
HSPB6_HUMAN	Heat shock protein beta 6	HSPB6	21 (19)	97.39	2.46E-03	1.43
HBD_HUMAN	Hemoglobin subunit delta	HBD	72 (16)	243.38	2.73E-03	1.31
IGHG3_HUMAN	Ig gamma 3 chain C region	IGHG3	22 (1)	102.81	3.70E-03	2.4
IGKC_HUMAN	Ig kappa chain C region	IGKC	16 (11)	88.88	0.04	1.44
KV401_HUMAN	Ig kappa chain V IV region Fragment	IGKV4 1	1	6.75	0.05	1.42
LAC3_HUMAN	Ig lambda 3 chain C regions	IGLC3	17 (4)	77.03	0.04	1.3
JKIP3_HUMAN	Janus kinase and microtubule interacting protein 3	JAKMIP3	1	13.34	3.56E-03	1.71
KINH_HUMAN	Kinesin 1 heavy chain	KIF5B	1	6.43	2.92E-03	3.16
LRC39_HUMAN	Leucine rich repeat containing protein 39	LRRC39	1	12.98	0.03	1.64
LDB3_HUMAN	LIM domain binding protein 3	LDB3	42 (32)	244.13	2.63E-04	1.33
LUM_HUMAN	Lumican	LUM	8 (6)	63.93	0.02	1.76
MAP4_HUMAN	Microtubule associated protein 4	MAP4	2	13.24	0.04	1.4
MCMBP_HUMAN	Mini chromosome maintenance complex binding protein	MCMBP	1	6.22	0.01	1.73
MYPC3_HUMAN	Myosin binding protein C cardiac type	MYBPC3	83 (61)	441.04	1.33E-03	1.36
MYL3_HUMAN	Myosin light chain 3	MYL3	77 (64)	293.66	0.02	1.47
MYL4_HUMAN	Myosin light chain 4	MYL4	24 (9)	94.98	0.02	1.89
MLRV_HUMAN	Myosin regulatory light chain 2 ventricular cardiac muscle isoform	MYL2	67 (47)	254.79	0.03	1.46
MTMR1_HUMAN	Myotubularin related protein 1	MTMR1	1	6.18	0.01	2.33
NEST_HUMAN	Nestin	NES	2	6.34	0.02	1.41
PDLI1_HUMAN	PDZ and LIM domain protein 1	PDLIM1	9	48.75	2.43E-03	1.7

PPIF_HUMAN	Peptidyl prolyl cis trans isomerase F mitochondrial	PPIF	1	6.66	4.56E-03	1.41
PGM5_HUMAN	Phosphoglucomutase like protein 5	PGM5	9 (5)	42.47	2.73E-03	1.18
POTEF_HUMAN	POTE ankyrin domain family member F	POTEF	103 (10)	471.62	6.36E-03	1.45
POTEJ_HUMAN	POTE ankyrin domain family member J	POTEJ	41 (1)	127.85	0.04	1.42
PROF1_HUMAN	Profilin 1	PFN1	12 (11)	48.6	0.04	1.22
PSB2_HUMAN	Proteasome subunit beta type 2	PSMB2	1	7.22	0.03	1.24
PDIA1_HUMAN	Protein disulfide isomerase	P4HB	4 (3)	25.59	0.03	1.23
PDIA6_HUMAN	Protein disulfide isomerase A6	PDIA6	2 (1)	12.38	0.04	2.25
WIT1_HUMAN	Putative Wilms tumor upstream neighbor 1 gene protein	WT1 AS	1	11.7	0.05	1.84
PWP2B_HUMAN	PWWP domain containing protein 2B	PWWP2B	1	6.42	3.27E-03	1.78
QOR_HUMAN	Quinone oxidoreductase	CRYZ	2	13.96	2.59E-03	1.41
ALBU_HUMAN	Serum albumin	ALB	349 (296)	1107.96	0.03	1.44
SRBS2_HUMAN	Sorbin and SH3 domain containing protein 2	SORBS2	1	6.77	0.05	1.36
SODC_HUMAN	Superoxide dismutase Cu Zn	SOD1	9	45.42	6.93E-03	1.25
THIO_HUMAN	Thioredoxin	TXN	3 (2)	20.07	3.73E-03	1.37
TERA_HUMAN	Transitional endoplasmic reticulum ATPase	VCP	32 (20)	195.58	0.01	1.17
TTHY_HUMAN	Transthyretin	TTR	13 (12)	90.43	0.04	1.47
TPIS_HUMAN	Triosephosphate isomerase	TPI1	74 (59)	378.87	5.64E-03	1.31
TNNC1_HUMAN	Troponin C slow skeletal and cardiac muscles	TNNC1	12 (8)	51.85	0.04	1.56
TNNI3_HUMAN	Troponin I cardiac muscle	TNNI3	35 (24)	152.45	0.03	1.66
TBB5_HUMAN	Tubulin beta chain	TUBB	22 (3)	146.22	7.12E-04	1.72
VINC_HUMAN	Vinculin	VCL	32 (25)	175.13	0.02	1.27

b) Proteins downregulated in HCM

Accession	Protein Description	Gene	Peptides	Score	Anova (p)*	Fold
DECR_HUMAN	2 4 dienoyl CoA reductase mitochondrial	DECR1	35 (27)	240.98	3.11E-05	1.42
RS13_HUMAN	40S ribosomal protein S13	RPS13	1	7.01	1.95E-03	1.31
ACS2L_HUMAN	Acetyl coenzyme A synthetase 2 like mitochondrial	ACSS1	2 (1)	13.02	0.04	1.26
KAD1_HUMAN	Adenylate kinase isoenzyme 1	AK1	28 (20)	135.15	0.04	1.3
ADT1_HUMAN	ADP ATP translocase 1	SLC25A4	68 (20)	297.04	3.58E-05	1.86
A1AG1_HUMAN	Alpha 1 acid glycoprotein 1	ORM1	23 (15)	114.07	2.92E-03	1.34
AACT_HUMAN	Alpha 1 antichymotrypsin	SERPINA3	15 (12)	90.18	1.18E-03	1.7
AOFB_HUMAN	Amine oxidase flavin containing B	MAOB	3 (2)	13.11	2.61E-04	2.11
ANXA1_HUMAN	Annexin A1	ANXA1	3	18.48	9.71E-05	2.65
ANXA2_HUMAN	Annexin A2	ANXA2	11 (10)	66.92	1.96E-03	1.43
ANXA3_HUMAN	Annexin A3	ANXA3	1	7.19	8.51E-05	1.45
ANXA5_HUMAN	Annexin A5	ANXA5	5	44.37	3.49E-04	1.45
ANXA6_HUMAN	Annexin A6	ANXA6	11 (9)	64.07	3.56E-04	1.65
ATP6_HUMAN	ATP synthase subunit a	MT ATP6	2	14.42	1.73E-04	2.11
ATPA_HUMAN	ATP synthase subunit alpha mitochondrial	ATP5A1	82 (63)	329.38	0.02	1.22
AT5F1_HUMAN	ATP synthase subunit b mitochondrial	ATP5F1	13 (9)	55.97	1.09E-03	1.43
ATP5I_HUMAN	ATP synthase subunit e mitochondrial	ATP5I	5 (2)	36.86	9.84E-05	1.87
ATPK_HUMAN	ATP synthase subunit f mitochondrial	ATP5J2	4	38.39	0.01	1.73
ATP5L_HUMAN	ATP synthase subunit g mitochondrial	ATP5L	6 (4)	15.44	3.64E-05	1.95
ATPG_HUMAN	ATP synthase subunit gamma mitochondrial	ATP5C1	12 (10)	78.44	2.42E-03	1.38
ATPO_HUMAN	ATP synthase subunit O mitochondrial	ATP5O	26 (24)	135.93	3.53E-05	1.38
CADH2_HUMAN	Cadherin 2	CDH2	2	12.86	4.95E-03	1.8
CMC1_HUMAN	Calcium binding mitochondrial carrier protein Aralar1	SLC25A12	5 (4)	34.6	3.57E-04	3.25
KCC2D_HUMAN	Calcium calmodulin dependent protein kinase type II subunit delta	CAMK2D	2	13.49	0.04	1.43
CAV1_HUMAN	Caveolin 1	CAV1	3	14.7	9.49E-03	1.57
QCR1_HUMAN	Cytochrome b c1 complex subunit 1 mitochondrial	UQCRC1	29 (23)	176.1	2.29E-04	1.59
QCR2_HUMAN	Cytochrome b c1 complex subunit 2 mitochondrial	UQCRC2	31 (23)	190.44	2.40E-04	1.74
QCR6_HUMAN	Cytochrome b c1 complex subunit 6 mitochondrial	UQCRH	6 (2)	36.74	1.11E-03	1.8
QCR7_HUMAN	Cytochrome b c1 complex subunit 7	UQCRB	8 (6)	59.85	1.76E-05	1.6
QCR8_HUMAN	Cytochrome b c1 complex subunit 8	UQCRQ	2 (1)	14.82	6.18E-05	4.91
QCR9_HUMAN	Cytochrome b c1 complex subunit 9	UQCR10	6 (5)	23.61	9.44E-05	1.54
UCRI_HUMAN	Cytochrome b c1 complex subunit Rieske mitochondrial	UQCRFS1	7 (5)	49.57	0.02	1.49
COX2_HUMAN	Cytochrome c oxidase subunit 2	MT CO2	3	7.85	2.97E-03	1.96
COX4I_HUMAN	Cytochrome c oxidase subunit 4 isoform 1 mitochondrial	COX4I1	12 (6)	45.67	2.06E-04	1.81

COX5A_HUMAN	Cytochrome c oxidase subunit 5A mitochondrial	COX5A	8 (6)	42.57	2.29E-03	1.68
COX5B_HUMAN	Cytochrome c oxidase subunit 5B mitochondrial	COX5B	7 (5)	38.59	8.92E-03	1.33
COX6C_HUMAN	Cytochrome c oxidase subunit 6C	COX6C	5 (4)	14.94	0.02	1.23
CX7A1_HUMAN	Cytochrome c oxidase subunit 7A1 mitochondrial	COX7A1	3	15.06	5.04E-05	1.92
CX7A2_HUMAN	Cytochrome c oxidase subunit 7A2 mitochondrial	COX7A2	3	16.21	1.49E-03	1.44
COX7C_HUMAN	Cytochrome c oxidase subunit 7C mitochondrial	COX7C	3 (2)	20.92	4.57E-04	1.89
CY1_HUMAN	Cytochrome c1 heme protein mitochondrial	CYC1	3 (2)	24.6	3.36E-04	2.22
DJC25_HUMAN	DnaJ homolog subfamily C member 25	DNAJC25	2	13.8	0.03	2
FABP5_HUMAN	Fatty acid binding protein epidermal	FABP5	9 (6)	49.02	0.01	1.37
FIBG_HUMAN	Fibrinogen gamma chain	FGG	3 (2)	6.96	0.03	1.91
ALDOA_HUMAN	Fructose bisphosphate aldolase A	ALDOA	97 (60)	440.99	0.02	1.17
ALDOC_HUMAN	Fructose bisphosphate aldolase C	ALDOC	42 (18)	261.88	8.11E-04	1.18
GPD1L_HUMAN	Glycerol 3 phosphate dehydrogenase 1 like protein	GPD1L	2 (1)	12.57	0.03	1.88
PYGM_HUMAN	Glycogen phosphorylase muscle form	PYGM	8 (3)	45.98	6.52E-03	1.5
HSDL2_HUMAN	Hydroxysteroid dehydrogenase like protein 2	HSDL2	11	49.8	2.09E-03	1.34
IDHP_HUMAN	Isocitrate dehydrogenase NADP mitochondrial	IDH2	88 (71)	348.24	0.01	1.46
LDHA_HUMAN	L lactate dehydrogenase A chain	LDHA	21 (14)	135.67	3.08E-04	1.33
ACSL1_HUMAN	Long chain fatty acid CoA ligase 1	ACSL1	1	6.99	2.60E-05	2.31
LC7L3_HUMAN	Luc7 like protein 3	LUC7L3	1	0	0.02	1.66
MGST3_HUMAN	Microsomal glutathione S transferase 3	MGST3	3	41.42	0.02	1.25
M2OM_HUMAN	Mitochondrial 2 oxoglutarate malate carrier protein	SLC25A11	18 (14)	101.87	1.40E-03	1.3
IMMT_HUMAN	Mitochondrial inner membrane protein	IMMT	3 (2)	14.22	3.37E-04	1.84
MYH3_HUMAN	Myosin 3	MYH3	107 (6)	444.25	0.02	1.13
NNTM_HUMAN	NAD P transhydrogenase mitochondrial	NNT	21 (14)	103.66	9.20E-04	1.64
NB5R1_HUMAN	NADH cytochrome b5 reductase 1	CYB5R1	1	6.67	1.87E-03	1.5
NDUA2_HUMAN	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 2	NDUFA2	3	23.78	0.01	2.01
NDUA4_HUMAN	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 4	NDUFA4	9 (6)	53.51	0.01	1.86
NDUA5_HUMAN	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 5	NDUFA5	4 (3)	20.33	0.02	1.56
NDUA8_HUMAN	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 8	NDUFA8	1	7.13	6.17E-03	4.82
NDUA9_HUMAN	NADH dehydrogenase ubiquinone 1 alpha subcomplex subunit 9 mitochondrial	NDUFA9	4	33.36	6.39E-03	2.11
NDUBA_HUMAN	NADH dehydrogenase ubiquinone 1 beta subcomplex subunit 10	NDUFB10	3	7.71	1.71E-03	1.97
NDUBB_HUMAN	NADH dehydrogenase ubiquinone 1 beta subcomplex subunit 11 mitochondrial	NDUFB11	2 (1)	6.3	9.83E-05	2.1
NDUB4_HUMAN	NADH dehydrogenase ubiquinone 1 beta subcomplex subunit 4	NDUFB4	2	21.54	1.15E-05	2.27
NDUB9_HUMAN	NADH dehydrogenase ubiquinone 1 beta subcomplex subunit 9	NDUFB9	2	6.14	0.01	1.94
NDUV1_HUMAN	NADH dehydrogenase ubiquinone flavoprotein 1 mitochondrial	NDUFV1	8 (5)	40.74	1.80E-03	1.58
NDUS2_HUMAN	NADH dehydrogenase ubiquinone iron sulfur protein 2 mitochondrial	NDUFS2	6 (5)	28.07	6.84E-03	1.35
NDUS3_HUMAN	NADH dehydrogenase ubiquinone iron sulfur protein 3 mitochondrial	NDUFS3	11 (8)	64.17	8.66E-05	1.78
NDUS4_HUMAN	NADH dehydrogenase ubiquinone iron sulfur protein 4 mitochondrial	NDUFS4	3	32.28	3.19E-03	1.42

NDUS5_HUMAN	NADH dehydrogenase ubiquinone iron sulfur protein 5	NDUFS5	3 (2)	15.79	7.39E-04	2.24
NDUS1_HUMAN	NADH ubiquinone oxidoreductase 75 kDa subunit mitochondrial	NDUFS1	12 (10)	93.97	0.01	1.24
NU4M_HUMAN	NADH ubiquinone oxidoreductase chain 4	MT ND4	2	14.78	3.77E-05	1.95
NU5M_HUMAN	NADH ubiquinone oxidoreductase chain 5	MT ND5	3	14.69	1.78E-05	1.99
NCOR2_HUMAN	Nuclear receptor corepressor 2	NCOR2	6 (2)	57.19	0.01	1.43
PTMS_HUMAN	Parathyrosin	PTMS	4 (3)	5.98	0.04	3.34
PRDX6_HUMAN	Peroxiredoxin 6	PRDX6	28 (23)	158.39	0.01	1.24
MPCP_HUMAN	Phosphate carrier protein mitochondrial	SLC25A3	11	93.19	1.27E-04	1.9
PRAF3_HUMAN	PRA1 family protein 3	ARL6IP5	1	6.93	0.05	1.54
GP158_HUMAN	Probable G protein coupled receptor 158	GPR158	1	5.62	0.05	1.27
PCCB_HUMAN	Propionyl CoA carboxylase beta chain mitochondrial	PCCB	2 (1)	13.14	0.02	1.52
ARHL1_HUMAN	Protein ADP ribosylarginine hydrolase like protein 1	ADPRHL1	5 (4)	34.33	0.05	1.22
RAB1B_HUMAN	Ras related protein Rab 1B	RAB1B	4 (3)	21.51	8.36E-04	1.5
GDIR1_HUMAN	Rho GDP dissociation inhibitor 1	ARHGDI1	1	7.12	0.01	1.38
AT2A2_HUMAN	Sarcoplasmic endoplasmic reticulum calcium ATPase 2	ATP2A2	20 (13)	126.67	1.20E-04	1.69
SBP1_HUMAN	Selenium binding protein 1	SELENBP1	5 (4)	13.43	1.08E-05	1.66
AT1A3_HUMAN	Sodium potassium transporting ATPase subunit alpha 3	ATP1A3	1	6.58	2.85E-05	1.96
SNX3_HUMAN	Sorting nexin 3	SNX3	1	6.31	0.03	1.65
ECHB_HUMAN	Trifunctional enzyme subunit beta mitochondrial	HADHB	22 (15)	127.86	3.57E-04	1.34
USMG5_HUMAN	Up regulated during skeletal muscle growth protein 5	USMG5	2 (1)	8.14	2.88E-03	1.45
ACADV_HUMAN	Very long chain specific acyl CoA dehydrogenase mitochondrial	ACADVL	53 (38)	288.34	9.60E-03	1.22
VDAC1_HUMAN	Voltage dependent anion selective channel protein 1	VDAC1	22 (16)	130.6	1.08E-04	1.59
VDAC2_HUMAN	Voltage dependent anion selective channel protein 2	VDAC2	14 (10)	103.05	4.09E-04	1.67
VDAC3_HUMAN	Voltage dependent anion selective channel protein 3	VDAC3	12 (8)	50.36	1.31E-04	1.76