4-9-2008

Serum Proteomic Profiles in Inflammatory and Non-Inflammatory Cardiomyopathies: A Novel Approach for Diagnostic Biomarker Discovery

Oyere Onuma

Serum Proteomic Profiles in Inflammatory and Non-Inflammatory Cardiomyopathies

A novel approach for diagnostic biomarker discovery

A thesis Submitted to the
Yale University School of Medicine
In Partial Fulfillment of the Requirements for the
Degree of Doctor of Medicine

By
Oyere Kalu Onuma S.B.
Class of 2007

1. ABSTRACT

The aim of this project is to develop a noninvasive serum test that predicts histologic forms of myocarditis (inflammatory) and dilated (non-inflammatory) cardiomyopathy using proteomic techniques to analyze serum proteins. Idiopathic dilated cardiomyopathy (DCM) and myocarditis (myocardial inflammation) represent a spectrum of heart muscle disease of various etiologies that usually present with progressive heart failure. Together, they constitute the leading cause of heart transplantation in the United States. Currently, the gold standard of diagnosis of myocarditis is by endomyocardial biopsy (EMB) and histopathological classification according to the Dallas Criteria[1]; however this diagnostic technique is severely limited by its invasiveness, a lack of sensitivity and an attendant sampling error, yielding diagnostic information in only 10-20% of the cases[2]. As such, the development of a non-invasive highly specific test for myocarditis is of great value and importance particularly in the diagnosis of giant cell myocarditis, a rare but very fulminant form of autoimmune myocarditis where timely institution of appropriate immunosuppressive therapy significantly increases transplant-free survival.[3]

We proposed, using an observational case-control study, to undertake a proteomic analysis to compare serum proteomic profiles - determined by mass spectroscopy and isotope tagging- with histologic findings on endomyocardial biopsy. Our hypothesis is that different forms of myocarditis and dilated cardiomyopathy have distinct serum protein profiles and that these unique profiles which correlate with specific histologic types, will allow for noninvasive diagnosis of major forms of myocarditis and DCM.

Acknowledgements

This work was undertaken with funding from the Sarnoff Foundation's Cardiovascular Research Fellowships for Medical Students.

I would like to extend my sincere thanks and gratitude to my mentor, Dr. Leslie Cooper, Jr. , Professor of Medicine, Mayo Clinic Rochester and also Xuan-Mai Persson and Larry Ward at the Proteomics Core Laboratory, Mayo Clinic, Rochester.

I would also like to acknowledge my adviser, Dr. Forrester Lee and the Yale School of Medicine Student Research Office for continued support throughout the years.

Finally, for my family and friends, without your support, I cannot be. Thank you very much.

SERUM PROTEOMIC PROFILES IN INFLAMMATORY AND NON-INFLAMMATORY CARDIOMYOPATHIES

2. SPECIFIC AIMS/HYPOTHESES

Idiopathic dilated cardiomyopathy and myocarditis together constitute the leading cause of heart transplantation in the United States. Worldwide, approximately 45% of all heart transplants are performed for IDC and up to 8% for myocarditis[4]. The current gold standard of diagnosis with endomyocardial biology and classification by the Dallas criteria is invasive - with a significant risk for major complications - and has low sensitivity and specificity for diagnosis. In the patient with acute dilated cardiomyopathy, prognosis of DCM varies with the histopathology found on biopsy. Thus, the central question addressed by this project is: "are there noninvasive tests that can accurately predict histolopathology in patients with acute cardiomyopathy?" Such tests would be of great benefit in clinical management and the assessment of prognosis in patients presenting with acute cardiomyopathy. We proposed to address this central question by testing the following hypothesis:

We hypothesized - based on the distinct clinical and histopathology characteristics of giant cell myocarditis (GCM), lymphocytic myocarditis (LM) and idiopathic dilated cardiomyopathy (DCM) - that there are specific proteome modifications induced by or associated with each disease state which results in differential expression of proteins in the serum.

To test our hypothesis, we proposed the following specific aims:

Specific Aim 1: To reproducibly characterize the quantitative and qualitative changes in serum protein expression in:

i) Acute giant cell myocarditis,

ii) Acute lymphocytic myocarditis,

iii) Idiopathic dilated cardiomyopathy

As compared to normal age and weight-matched controls.

Specifically, we used mass spectrometry based techniques for identification, isotope labeling for quantification and biostatistical methods for multivariate analysis of differential protein expression.

Specific Aim 2: To determine the sensitivity and specificity of the protein profiles for each disease state and identify potential serum cardiac biomarkers, which could be applied in novel diagnostic modalities, using statistical and biological correlation.

Our immediate emphasis in this part of the project, is on the characterization of the quantitative and qualitative differences between serum proteins in the different disease states. We proposed to determine the differences in the protein expression profiles in acute LM, acute GCM and idiopathic DCM as compared to normal subjects using broad-based proteomic screening techniques. Such differences would likely occur in up-regulation of immune proteins including the inflammatory cytokines, up-regulation of membrane proteins related to the formation of multinucleated giant cells in GCM, expression of viral proteins in LM and in DCM, the dysregulation of cytoskeletal proteins such as actin, lamin, dystrophin. Myosin light chain and other proteins linked to proven genetic mutations in DCM..

3. BACKGROUND/SIGNIFICANCE

Myocarditis is defined as an inflammatory process of the muscular walls of the myocardium which result in injury to the cardiac myocytes. Manifestations range from sub-clinical disease to sudden death. Myocarditis in association with cardiac dysfunction is classified as inflammatory cardiomyopathy and is usually caused by infections from viruses like enterovirus and adenovirus etc., autoimmune diseases, or responses to toxic substances. It is also thought to be a common cause of dilated cardiomyopathy and other cardiomyopathies[5] from evidence of viral persistence in the myocardium in patients with idiopathic dilated cardiomyopathy[6]. Idiopathic dilated cardiomyopathy is a heterogenous classification characterized by ventricular dilatation and diminished contractile function of unknown etiology. It is usually a chronic and histologically is not associated with active inflammatory infiltrates. In the USA, the estimated prevalence of DCM is 36.5 cases per 100,000[7] and in a review of 1230 cases of initially unexplained cardiomyopathy in the USA, 9% were thought to be due to myocarditis[8]. There are several different histopathologic forms of myocarditis. These include viral or lymphocytic myocarditis, most commonly due to adenoviruses like coxsackie virus, enteroviruses and several other viral and bacterial agents. It is characterized by interstitial lymphocytic infiltrates within the myocardium with little by way of myocardial necrosis as shown in Figure 1 below.

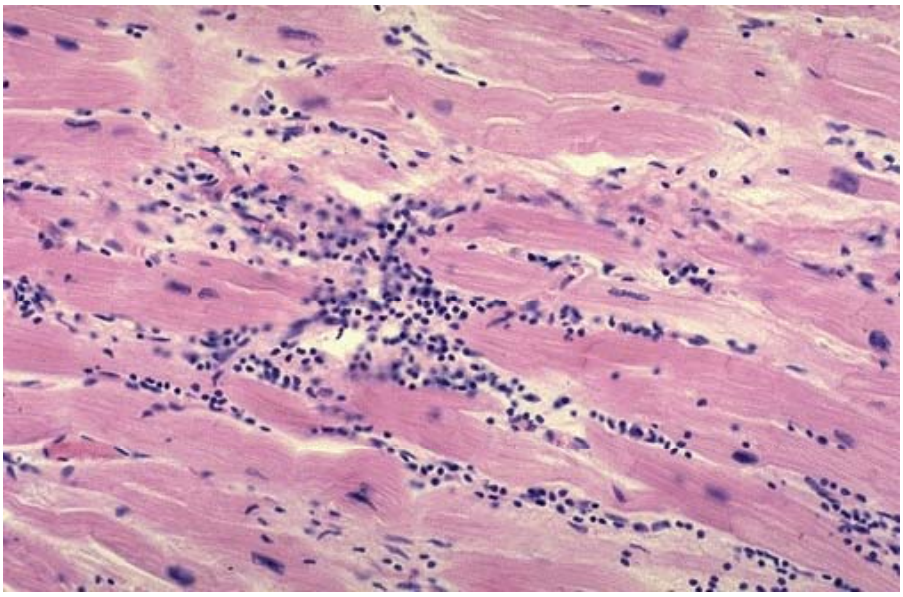


Figure 1: Lymphocytic Infiltrates in viral myocarditis. WebPath

Another distinct form of myocarditis is Giant Cell myocarditis, a form of autoimmune myocarditis characterized by the presence of giant cells in the myocardium with areas of myocyte necrosis and no signs of viral infiltration as shown in Figure 2 below.

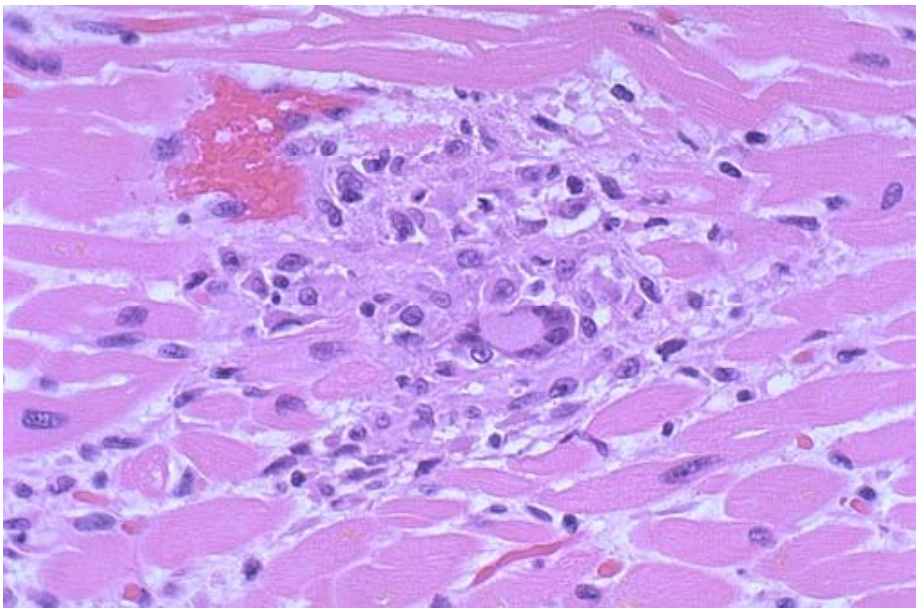


Figure 2: Giant cells in the myocardium with areas of myocyte necrosis and granuloma. WebPath.

In idiopathic dilated cardiomyopathy, the molecular features can include myocyte hypertrophy with myocyte degeneration and increased interstitial fibrosis.

Prognosis in myocarditis varies by histological type

Prognosis in myocarditis is dependent on the histological type diagnosed by biopsy and classification according to the Dallas Criteria. For instance, the 5-year survival rate in patients diagnosed with GCM is approximately 10% as compared with patients diagnosed with lymphocytic myocarditis shown in the Kaplan-Meier survival curve below.

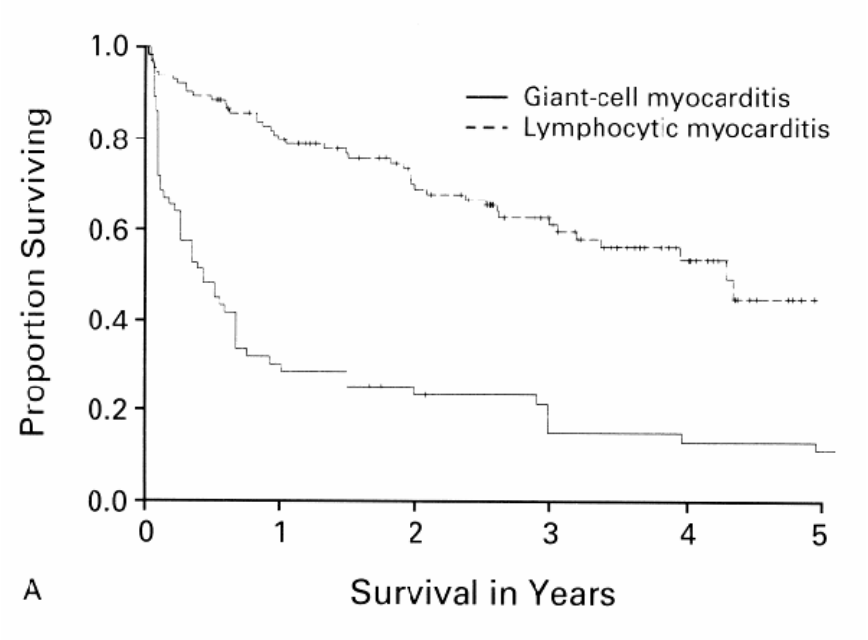


Figure 3,Transplant-free survival in GCM and LM. n=63 in GCM and n=111 in LM groups. P<.0001 by log-rank test. From Cooper, et. al., N Engl J Med 1997 *336:1860-66*.

Early diagnosis and appropriate treatment improves transplant-free survival especially in Giant Cell myocarditis (a rare but fatal autoimmune myocarditis) that requires immune suppression in addition to standard heart failure treatment as shown in the Kaplan-Meier curve below. This unpublished data from the GCM treatment trial showed a significant survival benefit for patients treated with a regimen consisting of cyclosporine, azathioprine and OK-T3.

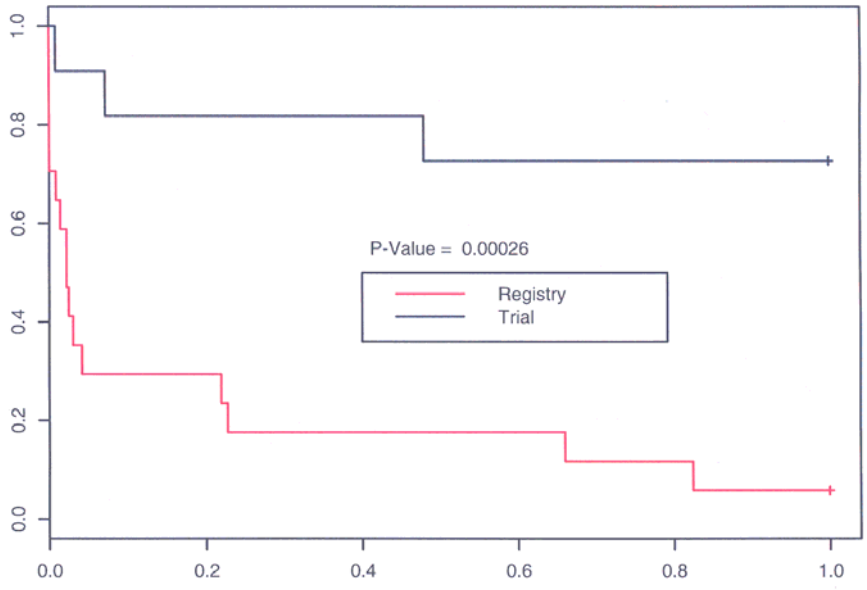


Figure 4. Transplant-free survival in 11 GCM patients treated prospectively with cyclosporine-based immunosuppression compared to 16 GCM historical registry patients treated with no immunosuppression or steroids alone. p=.00026. All patients were diagnosed by biopsy.

Diagnosis of Myocarditis by Endomyocardial Biopsy and the Dallas Criteria

The gold standard in the diagnosis of myocarditis is by endomyocardial biopsy and pathologic identification using the Dallas Criteria.[9] However this diagnostic standard is plagued by a lack of an attendant sampling error, a lack of sensitivity, and considerable intraobserver variability in the identification of inflammatory infiltrates, yielding diagnostic information in only 25% of cases.[10]

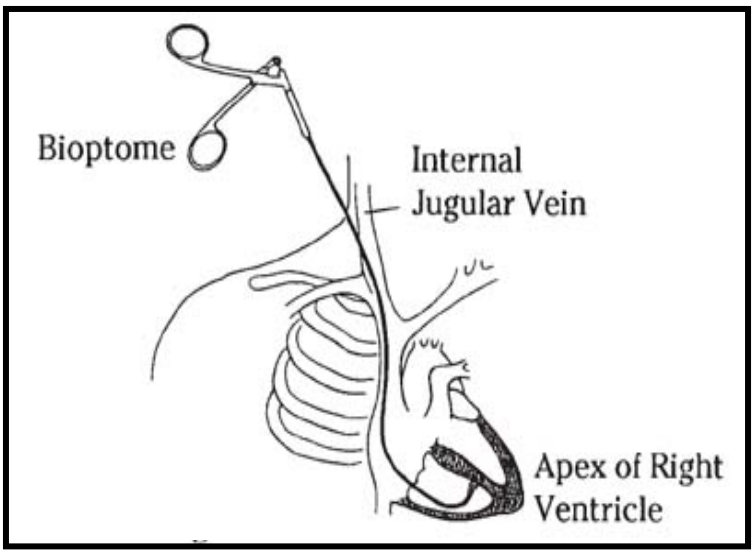


Figure 5: Right heart endomyocardial biopsy is gold standard for diagnosis of myocarditis. www.stanford.edu/biopsy

The Myocarditis Treatment Trial reported poor concordance between the clinical diagnosis of myocarditis and histopathological diagnosis from EMB with histopathologic evidence of myocyte inflammation in only 214 of 2233 patients enrolled in the trial. Furthermore the risks of biopsy include a 1 in 1000 risk of death and 1 in 250 risk of perforation in experienced hands[11] and also a risk of arrhythmias which preclude its routine clinical use. In effect, EMB is increasingly viewed "tarnished" standard for diagnosis of myocarditis, necessitating the development of better diagnostic standards[12]

Cardiac Biomarkers and Markers of immune up-regulation in myocarditis

There are several other biomarkers that have potential applicability in the clinical diagnosis of myocarditis. Standard markers of myocardial damage including Troponin – I, CK-MB have good specificity but limited sensitivity in the diagnosis of myocarditis. Lauer et al report for Troponin I, a sensitivity for detecting myocarditis of 53%, specificity of 94%, a positive predictive value of 93% and a negative predictive value of 56%[13]. Antibodies to a variety of cardiac autoantigens such as myosin heavy chain, the β-adrenergic receptor, mitochondrial antigens, and the adenosine diphosphate and triphosphate carrier proteins[14] have also been demonstrated in the sera of patients with acute myocarditis. Recent advances in the use of PCR and the description of HLA typing for the diagnosis of myocarditis have furthered diagnostic capabilities[15]. However, these techniques have limited clinical applicability due to a lack of specificity and still require invasive testing, in most cases, to obtain the myocardial tissue required for the tests.

Change in Protein Expression in Myocarditis.

The pathogenesis of myocarditis is thought to occur via several different mechanisms. These include a direct myocardial invasion by cardiotropic viruses, activation of CD4 cells leading to clonal expansion of B cells and production of cardiac auto-antibodies or the production of pro-inflammatory cytokines including IL-1, IL-2, TNF-∝and IFN-γ . In essence, the acute injury to the myocardium is accompanied by distinct changes in the protein expression profile and is more likely induce rapid post-translational modifications in the proteins as compared to the co-translational changes that occur with altered gene levels in chronic disease stated such as dilated cardiomyopathy[16].

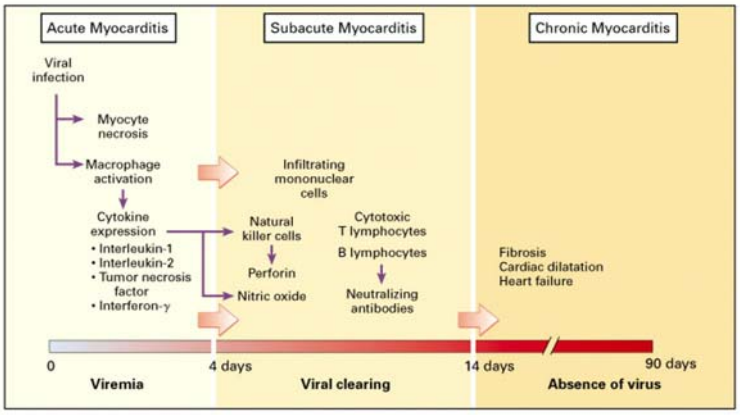


Figure 6: Pathogenesis of experimental autoimmune myocarditis varies in the acute, subacute and chronic stages of myocarditis. Feldman and McNamara NEJM, 343 (19): 1388

Recent work on the gene expression in giant cell myocarditis myocardial samples, as compared to normal hearts show an up-regulation in genes involved in immune response, transcriptional regulation and metabolism[17] suggesting that similar changes are likely in protein expression.

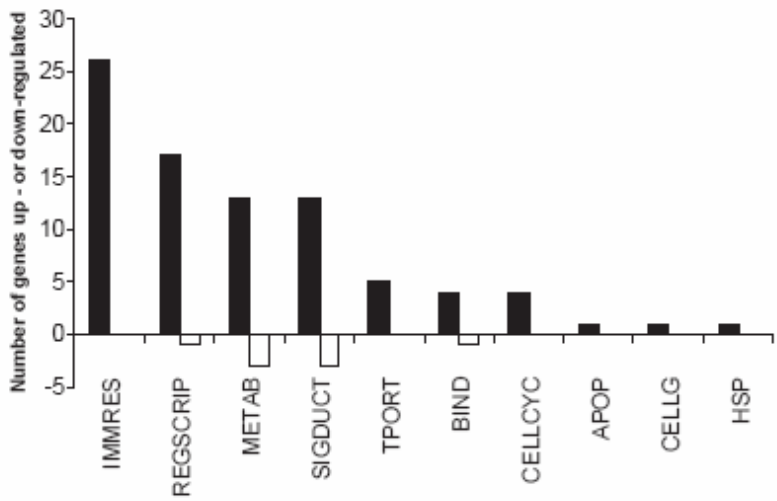


Figure 7; The gene expression profile in GCM is significant for a marked up-regulation in the immune response genes in GCM as compared to normal. Kittleson, M.M et al, Int. J Cardio 102 (2005)

Clinical Proteomics in Myocarditis

Clinical proteomics involves the identification of protein patterns of disease in order to improve patient care and public health through better assessment of disease susceptibility or selection of therapy for the individual. The National Heart Lung and Blood Institute (NHLBI) Clinical Proteomics Working Group[18] identified an opportunity in the use of modest numbers of well characterized clinical specimens for the discovery of such protein markers. Proteomic analysis provides a mechanism for the broad-based screening of the proteome to characterize the quantitative and qualitative modifications that are secondary to each disease process. Previous work in the analysis of myocardial tissue from animal and human models of cardiomyopathy have shown that the differential change in myocardial proteins such as the myosin light chain. This and other changes in the failing heart discovered through myocardial proteomics are cataloged online on the World Wide Web[19].

Although there have been some advances made in tissue proteomics in cardiovascular diseases, the proteomic analysis of the serum remains a novel approach in the investigation of diagnostic, prognostic and therapeutic markers of cardiovascular disease.

Given the ease of obtaining serum, and the widespread clinical use of other serum biomarkers in cardiovascular disease, it holds promise for the discovery of novel biomarkers. Recent advances in protein separation techniques, isotope tagging and increased sensitivity of mass spectrometry now allow for broad-based screening of the serum proteome[20]. More recently, classifications from the Human Proteome Organization (HUPO) have identified cardiovascular-related proteins found experimentally in the human serum[21] based on several proteomics platforms. Of relevance is the grouping of these proteins putatively by functional class into eight broad groups: markers of inflammation and/or cardiovascular disease, vascular & coagulation, signaling, growth & differentiation, cytoskeletal, transcription factors, channels/receptors and markers of heart failure. This provides the beginnings of a serum proteomic blueprint for future development of new diagnostic and prognostic markers for cardiovascular disease.

4. PRELIMINARY DATA

The research was conducted under the primary supervision of Dr. Leslie Cooper[22] who has an established interest and expertise in myocarditis research and has several current research protocols on the treatment of myocarditis (Giant Cell Treatment Trial, IMAC trial.) The team also included individuals with expertise in current proteomic technologies and techniques in clinical proteomics (Dr. Sreekumar Raghavakimal, Ph.D. and Xuan-Mai Persson of the Mayo Proteomics Core Facility).

State of the art proteomic research facilities and mass spectrometer were available through the Mayo Proteomics core laboratory and the General Clinical Research Center (GCRC). In addition, the team includes a biostatistician (Ann Oberg PhD) who had expertise in the design and analysis of proteomic data and a Research Nurse, Annette McNallan is to assist with the collection and coding of samples.

a. **Proteomic comparison of plasma samples of GCM and normal control:**

As proof of concept, we performed multidimensional liquid chromatography – quadrupole time of flight mass spectrometry (LC/MS/MS) analysis on plasma samples of one patient with GCM, one patient with DCM, one patient with LM and one normal control under a feasibility study protocol. Briefly, the samples were separately filtered through a Microbead separation column; the samples were labeled with an isotopic label, combined in equal protein amounts after labeling and analyzed by mass spectrometry with LC/MS/MS.

Protein and peptides were identified using two separate and independent search engines ProQuant (ABI) and MASCOT (Matrix Science) against the human non-redundant database. Protein and peptide quantifications from the labeled isotopes were obtained through the ProQuant software (ABI). Our results show the upregulation of inflammatory markers like CRP and cytokines in the GCM group, as expected. In each sample, approximately 2800 unique proteins were identified with approximately 200 proteins differentially expressed at a widely accepted arbitrary ratio of 1.2 or greater or a ratio of 0.8 or less.

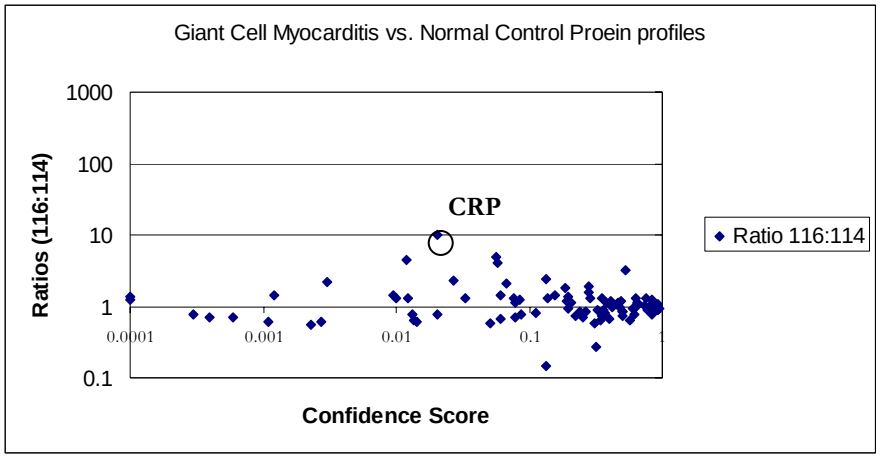


Figure 8: Difference in protein expression (ratios greater or less than one) between GCM and Normal controls at different confidence scores (p-values) Ratio >1 is up-regulated, Ratio < 1 are down-regulated, p-value <0.05 is considered significant

P Val 116:114	Ratio 116:114	Protein Name
0.0001	1.2456	hemopexin precursor
0.0001	1.3703	inter-alpha-trypsin inhibitor family heavy chain-related protein
0.0003	0.7645	Afamin precursor (Alpha-albumin) (Alpha-Alb)
0.0004	0.6965	gelsolin isoform b
0.0006	0.7131	proapolipoprotein
0.0011	0.6081	Inter-alpha-trypsin inhibitor heavy chain H2 precursor (ITI heavy chain H2) (Inter-alpha-inhibit or h)
0.0012	1.414	similar to hypothetical protein
0.0023	0.5505	alpha2-HS glycoprotein
0.0027	0.5951	Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin) (PRO 1400)
0.003	2.2001	Plasma protease C1 inhibitor precursor (C1 Inh) (C1Inh)
0.0096	**1.4609**	**Complement component C9 precursor**
0.0099	1.3213	Unnamed protein product
0.0118	4.4887	alpha1-antichymotrypsin
0.0123	**1.2789**	**complement component 6**
0.0133	0.7846	A Chain A, Antithrombin Iii
0.0136	0.6324	Insulin-like growth factor binding protein, acid labile subunit
0.0143	0.6147	A Chain A, Serum Amyloid P Component (Sap)
0.0203	0.7793	Thyroxine-binding globulin precursor (T4-binding globulin)
0.0205	**9.8728**	**CRP protein**
0.0267	2.3294	B Chain B, Cleaved Alpha-1-Antitrypsin Polymer
0.0331	**1.2829**	**complement component 5**
0.0504	0.5871	Tetranectin precursor (TN) (Plasminogen-kringle 4 binding protein)

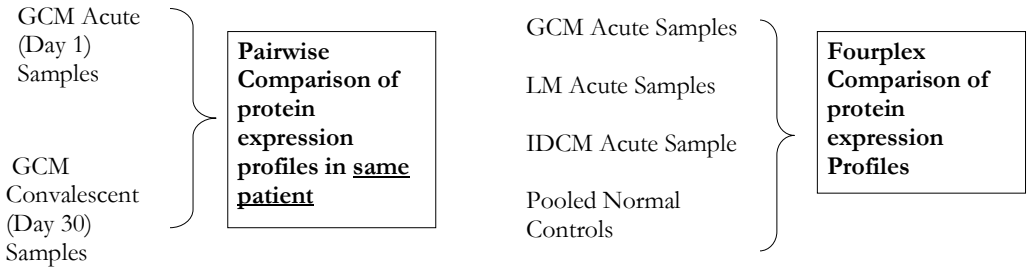
Table 1: Select protein expression ratios and p-values in a patient with GCM (116:114, ratio of GCM to Control) and a normal control, run twice (114 and 117). Inflammatory markers like CRP and complement are significantly elevated in disease subjects versus control.

5. STUDY DESIGN AND METHODS

Study Design:

This is an observational case-control study with matched cases and controls conducted in two phases. We studied:

- 10 cases of biopsy-proven giant cell myocarditis (GCM),
- 10 cases of biopsy-proven lymphocytic myocarditis (LM) and
- 10 cases of biopsy-proven idiopathic dilated cardiomyopathy (IDCM),
- 10 normal controls



PHASE I **PHASE II**

Each case was compared to a normal age and sex matched control. Given the inherent biological variations in serum protein profiles described with age, sex and body habitus, we controlled for these factors by one-to-one matching to minimize bias in our analysis. Cases were examined as a matched set for proteomic analysis using 4-isotope labeling. To minimize the effects of isotope-isotope interaction, isotope labels were randomly assigned to each disease or normal grouping.

Power and Sample Size:

Given the rarity of the conditions under study and the difficulty in ascertaining an effect size *a priori*, the calculation of sample size in this design is dependent on a back-calculation from the total number of cases available for study. Thus, with a study design of 10 patients in each group, and an estimated effect size of 1.2 or 0.8, with a co-efficient of variation of 80, we can achieve 80% power at 0.05 p-value. For the multivariate analysis of protein expression, we will consider a lower p-value of 0. 01 to account for multiple hypothesis testing.

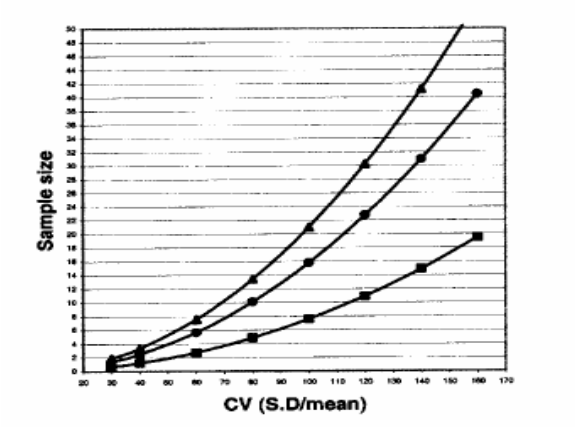


Figure 9: Relationship between sample size, variation and sample power.

Study Population:

Samples for the Proteomic analysis had already been obtained as part of previous research protocols and the project was approved by the IRB at the Mayo Clinic. All subjects had given written informed consent for their samples to be stored and used in the study of other disease conditions. Serum samples were available from patients with giant cell myocarditis, lymphocytic myocarditis and idiopathic dilated cardiomyopathy both in the acute and post-treatment phases of disease with corollary clinical data also available on the course of disease. Serum samples with similar storage time frames were also available from age and weight-matched normal controls for comparisons. The use of age and weight matched controls in this case was done to reduce the biologic variations in the proteome associated with those factors. This study used subjects with stored serum samples from previous Mayo Clinic protocols. The relevant protocols include:

2186-04 A Phase 1, Open-Label. Pilot Study to Assess the Safety of Immunoadsorption Using the Fresenius Immunoabsorba for Chronic Dilated Cardiomyopathy, L Cooper PI

2288-03 Genetic Modulation of LV Recovery Project, L Cooper PI

468-02 The Role of Autoantibodies and Cytokines in Myocarditis, L Cooper PI

754-00 Thromboangiitis Obliterans (Buerger's Diseases) Registry, L Cooper PI

147-99 A Multicenter Randomized Study to Evaluate the Efficacy of Monomurab-CD3, in Subjects with Giant Cell Myocarditis, L Cooper, PI

1792-03 UNIPATH study normal controls.

Inclusion Criteria:

We included subjects with the following criteria in the study:

- **Biopsy proven** cases of DCM, LM (by Dallas Criteria), IDCM
- Subjects with stored serum samples obtained during the acute phase of disease (i.e. within 7 days of diagnosis and before the institution of immuno-modulatory treatment)
- Age greater than 18 years
- Written informed consent available for use of samples.

Research Methods:

There are 3 major sequential steps in the proteomic analysis of serum which we employed in this study.

Step 1: **Sample Preparation**:

Protein Depletion: Given the complexity of the proteins in the plasma and the wide dynamic range of concentrations -- greater than 10 orders of magnitude -- it is important to simplify the composition of serum through protein depletion. This allows for the detection of low-abundance proteins after the removal of the 12 most abundant proteins that account for

95% of the protein content of the serum.

The protein depletion was done using Genway Biotech's Seppro Microbead-conjugated avian IgY antibodies to specifically remove human serum albumin, IgG, IgA, IgM, transferrin, fibrinogen, apolipoprotein A-I, apolipoprotein A-II, haptoglobin, α-1 antitrypsin, α-1 acid glycoprotein and α-2 macroglobulin[23]. The use of this system allowed for the recovery of all three separate fractions of proteins in the serum as shown below. Each fraction was analyzed separately, starting with the least abundant protein fraction, to characterize the full spectrum of serum protein.

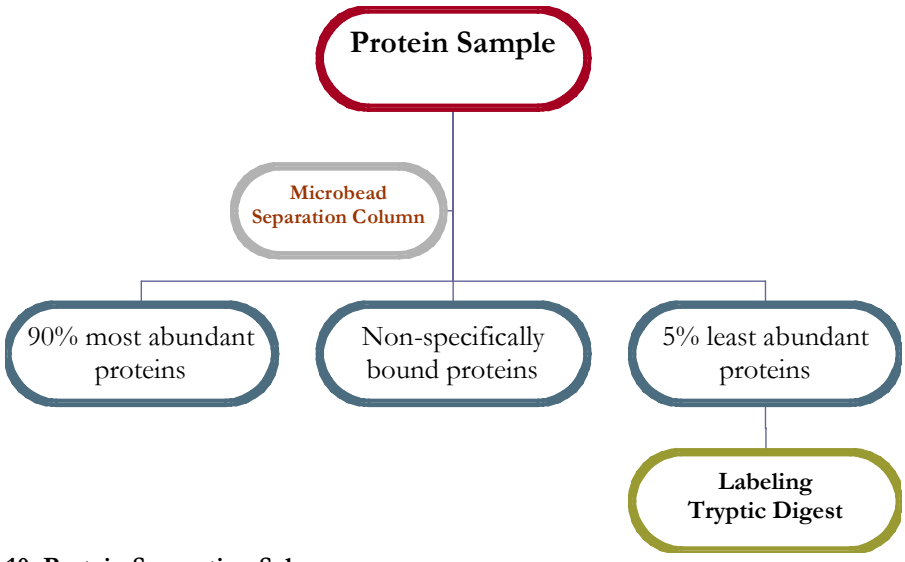


Figure 10: Protein Separation Schema

Isotope Labeling: Briefly, 100µg of protein post-depletion from each serum sample was separately reduced, denatured, cysteine blocked and tryptic digested according to the Applied Biosystems (ABI) iTRAQ protocol. The 4 samples were labeled with the 114, 115, 116 and 117 isotope iTRAQ reagents according to the randomization plan as described in the study design section. The samples are combined after labeling for further analysis on the mass spectrometry.

Step 2: **Liquid Chromatography/Mass Spectrometry analysis (LC/MS/MS)**

Following combination of the fractions, the peptides were fractionated on a strong cation exchange column prior to introducing to a capillary LC/MS/MS system (Ultimate, LC packings and QSTAR ABI.) Independent data acquisition from each fraction is obtained using the Analyst QS system.

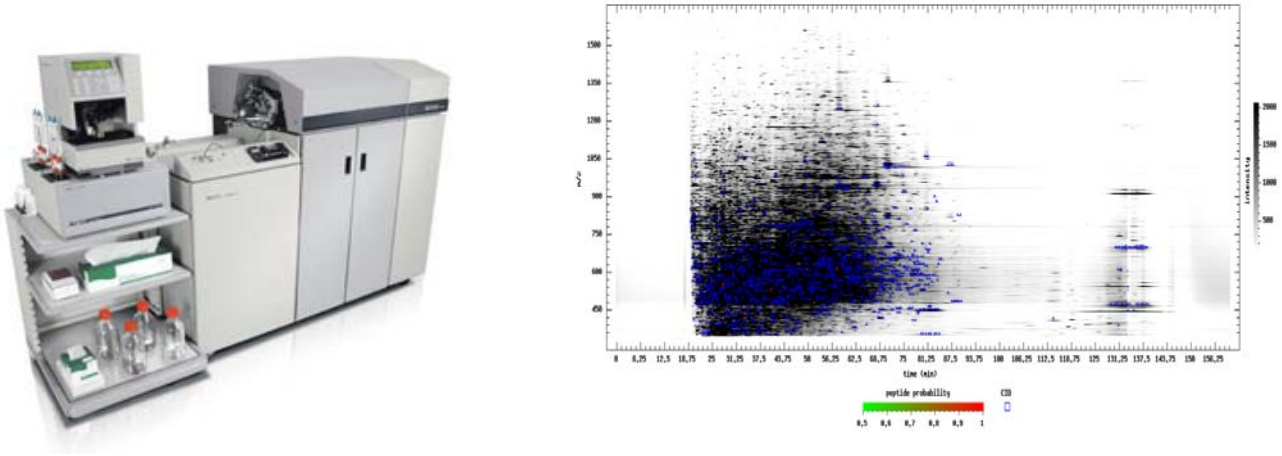


Figure 11: QStar ABI Mass Spectrometer, and graphical representation of a sample fraction after analysis on mass spect.

Step 3 : Protein Identification and Quantification:

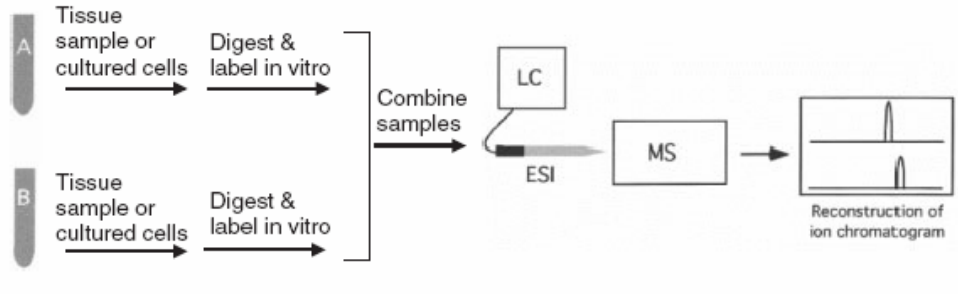


Figure 12: Mass Spectrometry Analysis of differential protein expression using in-vitro isotope tagging.[24]

Protein Identification:

Following the analysis of the samples on the mass spectrometry, the data obtained (peptide mass fingerprint after tandem MS/MS) was reconstructed into its parent peptide sequence and queried against known protein databases – SEQUEST and MASCOT using the Applied Biosystems (ABI) proprietary ProQuant software[25]. We chose to use this software because it qualifies the identified protein in a non-redundant score, assigning confidence scores (p-values) to the proteins such that species with higher scores have a higher likelihood of being present in the analyte and are easily identified. To validate our system, we will analyze the known fraction of specifically bound proteins (from the protein separation step) as an internal control.

Protein Quantification:

The quantification of various protein expression levels in our experimental set-up was carried out by analyzing the quantities of the isotope-tagged fragment of each protein. Isotope tagging was done using the proprietary iTRAQ system. This involves the use of an isobaric mass tag attached to a peptide reactive group (PRG). The isobaric tag is actually composed of a reporter mass and a balance mass which makes it isobaric. The reporter masses vary from 114 to 117, producing 4 separate masses that allow for the comparison of up to four separate samples simultaneously.

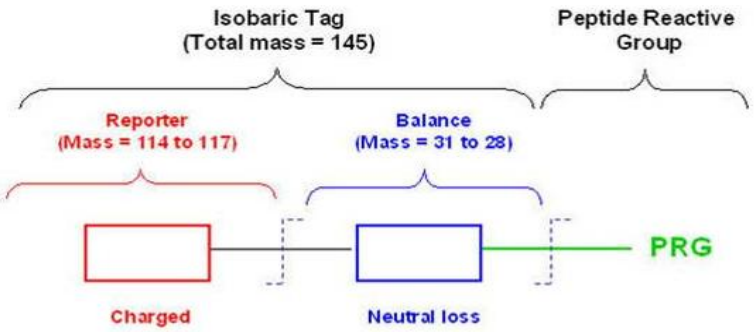


Figure 13: iTRAQ chemical tagging schema. http://www.proteome.soton.ac.uk/iTRAQ

The peptide reactive group attaches to the C terminal of the amino acid, thus attaching the isotope tag to all amino acids in the sample of interest. Following combination of the samples and mass spectrometry analysis, the unique fragmentation pattern identifies the peptide sequence in question, whilst the reporter or mass ions give a measure of the abundance of the peptide/protein species in the samples as shown in Figure 14 below.

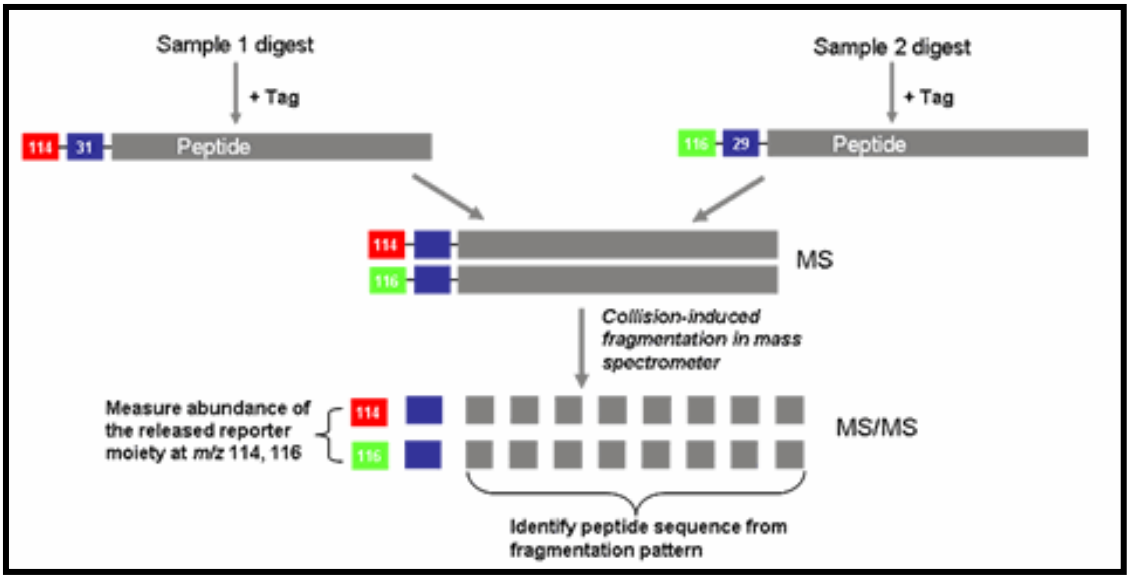


Figure 14: Principles of isotope tagging using iTRAQ. *http://www.proteome.soton.ac.uk/iTRAQ.htm*

This information is contained in the ProQuant analysis and is expressed as a ratio of a given tag to a pre-selected baseline tag, which is usually chosen to correspond to the normal control. Proteins are grouped into four categories from this analysis:

i. up-regulated (ratio >1.2)

ii. down-regulated (ratio <0.8)

iii. absent in patient but present in control

iv. present in patient but absent in control.

To validate the identification and quantification of the protein data obtained from the ProQuant Analyst software, we will use an independent protein analysis software, the i-Tracker[26] (open source : www.i-Tracker.com) to query the non-redundant databases and determine the accuracy of our analysis by correlation studies.

Statistical Analysis:

Specific Aim 1: To characterize the qualitative and quantitative changes in the serum expression profiles, we used two analytical methods:

Univariate Analysis - Average of Ratios: One dimensional analysis of the average change in the expression of proteins will be done by determining the average ratios of expression of a given protein across all ten sets of experiments (provided that the bias or error scores are similar between sets.) To normalize the distribution and provide a robust analysis of the ratios, conversion to the log space is necessary, as shown below.

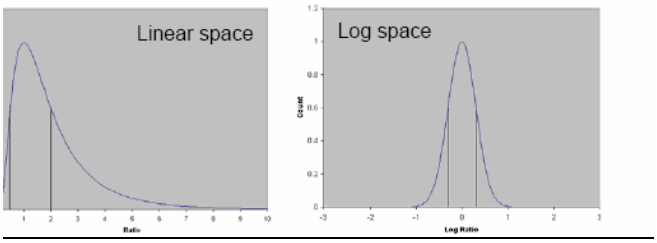


Figure 15: Transformation of ratios into log space for univariate analysis

Multi-dimensional analysis: The simple analysis of ratios fails to take into account the correlation between the expression levels of various proteins within the same disease state or amongst normal subjects i.e. it is likely that the changes in protein expression levels are correlated to a panel of proteins rather than a change in a single protein. Furthermore, the inherent biological variation can confound the ability to conclude that any differences are due to the true changes in the protein expression[27]. Thus to analyze this, we will use dimensional scaling or a dimensionality reduction procedure, primarily principal components

analysis (based on the pooled data on all subjects.)

This technique, which does not require prior knowledge of the disease state (i.e. blinded analysis) involves calculating the distance between individual samples in *n* –dimensional space using a variety of matrices such that similar samples cluster close to each other and dissimilar samples cluster far away. Using this scaling, one can test whether the group means for the multi-dimensional scaling coordinates are different among groups, using standard Multivariate ANOVA measure like a Hotelling T test.

Specific Aim 2:

To evaluate the sensitivity and specificity of the serum protein markers identified by profiling, boosted decision tree analysis and Euclidean distance vectors can be used to determine specific proteomic patterns. To ascertain the biological relevance of the differences in expression profile, we will do protein expression pathway analyses, localization and functional annotation to delineate the importance of the proteins within the metabolic pathway. These analyses can be performed on human proteins using the appropriate software (Pathway analysis from MetaCore, Pathway assistance from Stratagene and functional annotation from KEGG.)

Oyere K. Onuma performed all the experiments presented below, including the sample preparation and data analysis. The mass spectrometer was operated in the Proteomics Core Laboratory, Mayo Clinic Rochester. Biostatistical support was also provided by the Department of Biostatistics, Mayo Clinic Rochester.

RESULTS:

Protein separation using the Seppro column was reproducible.

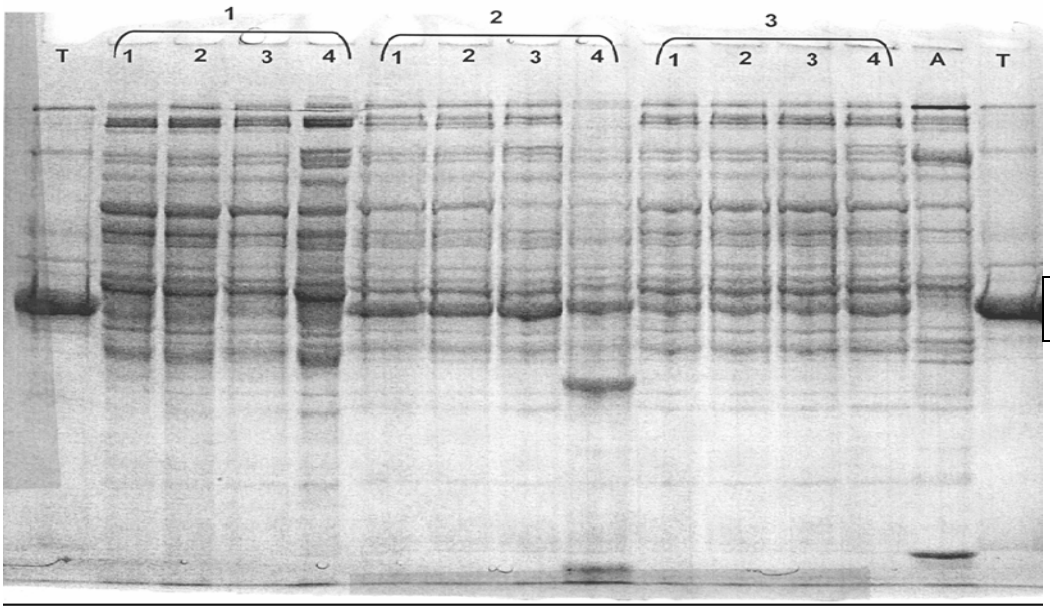


Figure 16: 1-D Gel showing serum fractions pre- and post- depletion

In the figure above, the depletion of the 12 most highly abundant proteins using the Seppro affinity column is highly consistent and reproducible. In the two end lanes in the gel shown above , the whole un-fractionated serum shows a heavy albumin signature. However post-depletion (lanes 2-14) albumin is markedly depleted, allowing for the appearance of less abundant proteins in the serum profile.

GCM (ACUTE VS. CONVALESCENT)

In Phase I studies, we compared serum samples in acute Giant cell myocarditis as compared to convalescent myocarditis. Serum samples from eight patients with histopathologic and clinical diagnosis of giant cell myocarditis were included in an observational case-control study in self-matched pairs. Briefly, the serum samples were separately filtered through an antibody-affinity microbead separation column; trypsinized, isotope labeled, combined and

analyzed by tandem mass spectrometry. Proteins with a fold change of 20% were defined as differentially expressed and pathway analysis of differential proteins was performed by comparison with published protein groupings. Of more than 600 differentially regulated proteins across the eight sets, there was marked up regulation in acute GCM of immune and defense proteins (n=217) such as C-Reactive Protein and the Complement proteins. Proteins involved in lipid, fatty acid and steroid metabolism, such as the apolipoproteins show significant down regulation in acute GCM as shown in Figure 17. Pathway analysis of 43 common differentially expressed proteins also highlighted activation of the immune response and lipid metabolism molecular pathways.

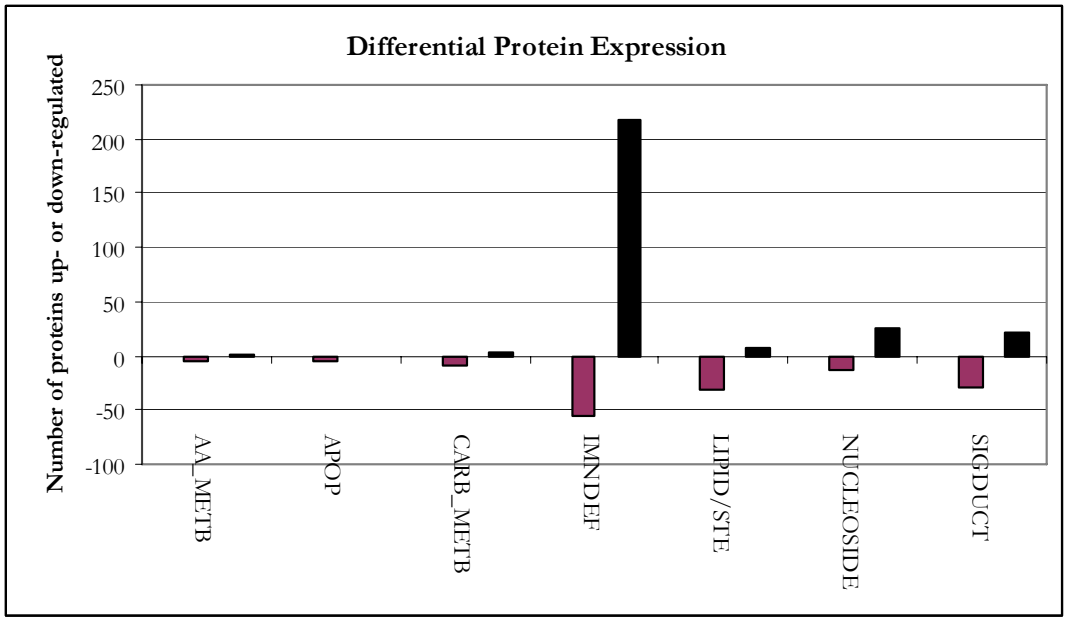
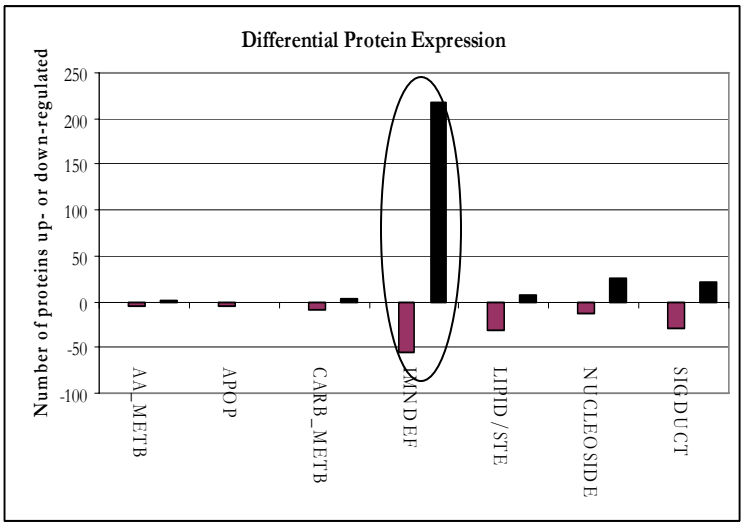
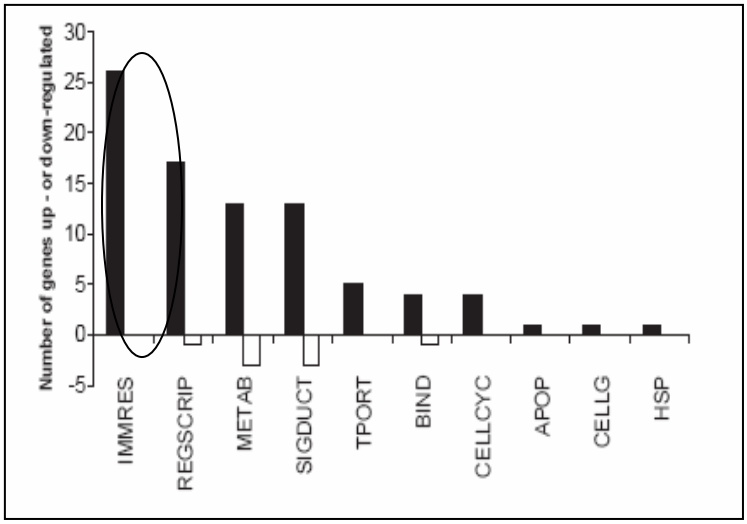


Figure 17. : Number of known proteins upregulated or downregulated in serum of acute versus convalescent patients with giant cell myocarditis classified by molecular function. AA_METB, amino acid metabolism; APOP, apoptosis; CARB_METB, carbohydrate metabolism; IMNDEF, immunity and defense; LIPID/STE, lipid, fatty acid and steroid metabolism, NUCLEOSIDE, nucleoside, nucleotide and nucleic acid metabolism, SIGDUCT, signal transduction.

When compared to previous results on GCM gene expression from work done by Kittleson et al, our results show an interesting concordance with the marked up-regulation in immune defense proteins across platforms (gene expression versus proteomics) and also across

biological samples (tissue versus serum) as shown below, thus validating the serum proteomics approach.





GCM SERUM PROTEINS **GCM TISSUE GENES**

Select Differentially Regulated Proteins in Acute Vs. Convalescent GCM.

Immunity and Defense

- *Complement component 4a*
- *Complement C4A precursor*
- *Complement 4B proprotein*
- *Complement C3 precursor*
- *Complement C5 precursor*
- *Beta-2-glycoprotein I precursor (Apolipoprotein H)*
- *Complement factor H precursor*
- *CRP protein*

B-cell and antibody-mediated immunity

- *anti-HBsAg immunoglobulin Fab kappa chain*
- *immunoglobulin lambda light chain VLJ region*
- *Ig L-chain V-region*
- *Ig heavy chain, variable region*
- *Ig kappa chain variable region*
- *immunoglobulin kappa chain variable region*

Complement mediated immunity

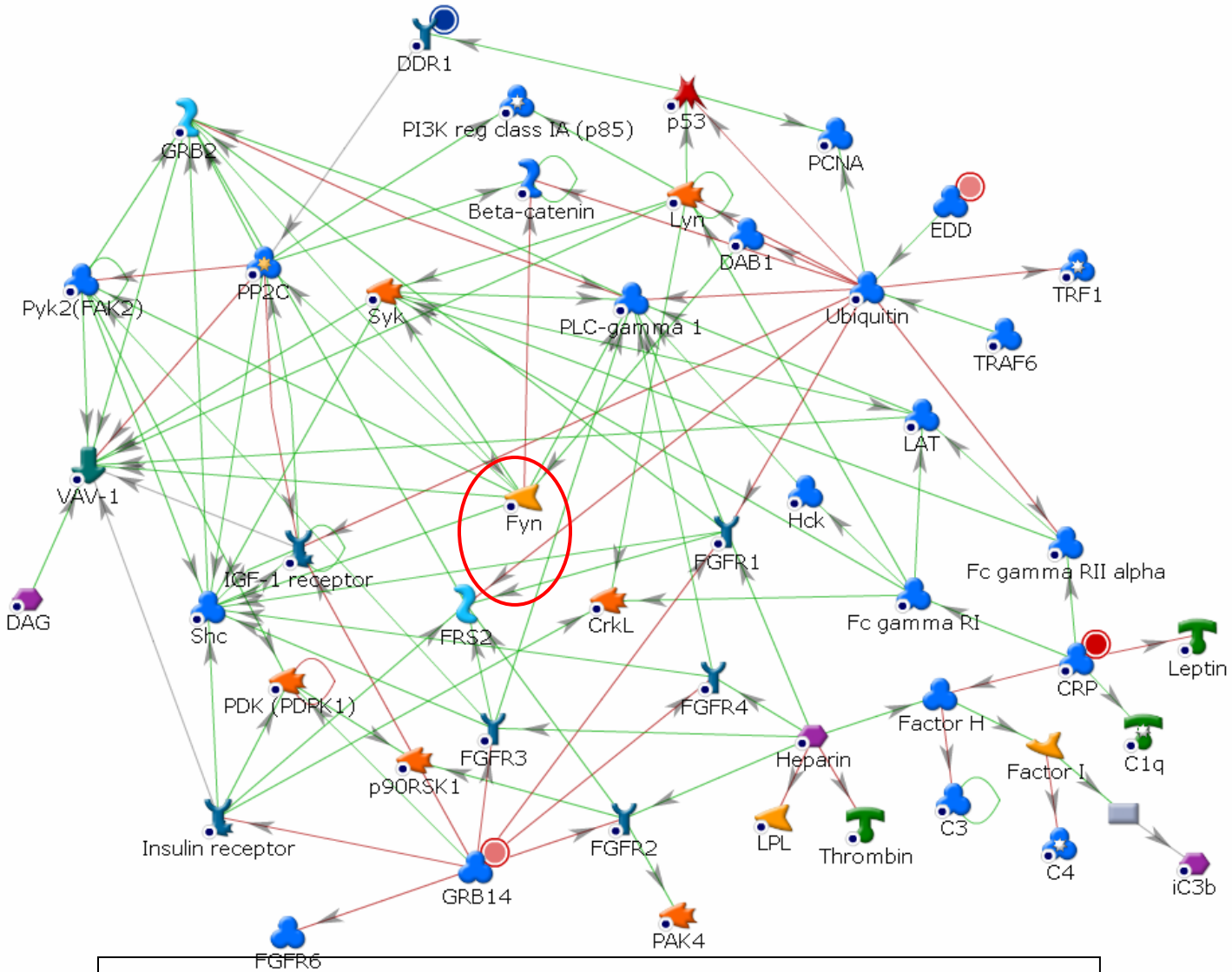
- *C9 complement protein*
- *complement C3d/Epstein-Barr virus receptor precursor - human*
- *immunoglobulin J chain*
- *complement component 1, q subcomponent, gamma polypeptide*
- *Complement component C9 precursor*
- *Complement component C7 precursor*

Lipid, Fatty Acid and Steroid Metabolism

- *proapo-A-I protein*
- *proapolipoprotein*
- *apolipoprotein D, apoD*
- *lipoprotein Gln I*
- *Apolipoprotein E precursor (Apo-E)*
- *Hypothetical protein (Fragment)*
- *Apolipoprotein A-IV precursor (Apo-AIV)*
- *Retinoic acid receptor RXR-gamma*

Pathway Analysis

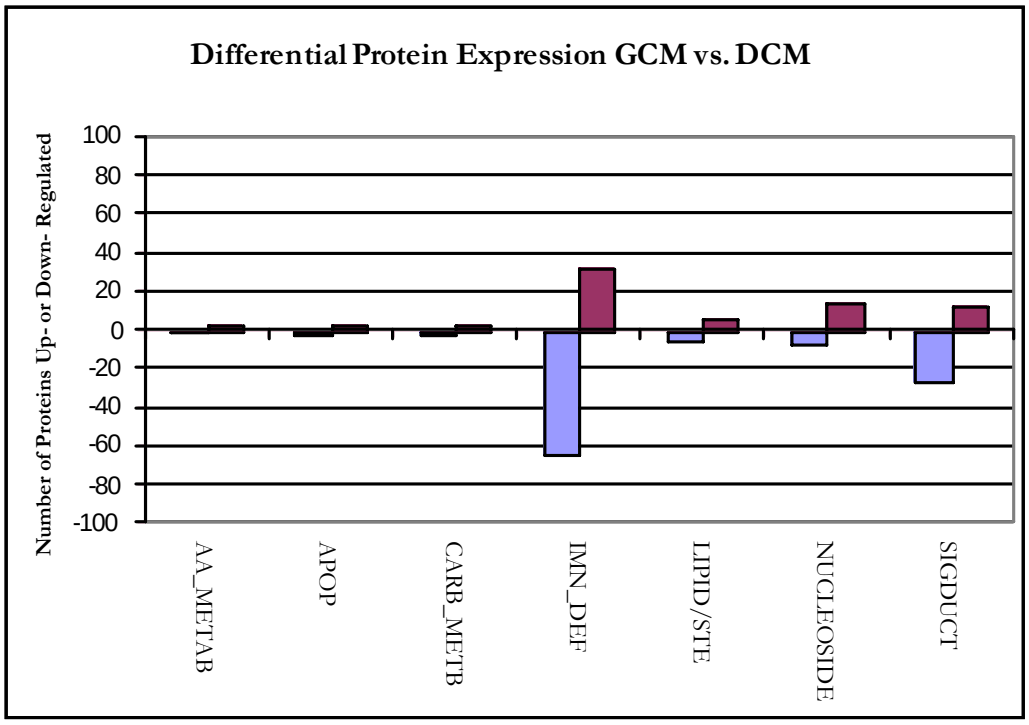
Pathway analysis was done using the MetaCore Pathway Analysis program. This is a web-based software for molecular pathway analysis. It allows for the analysis of proteomic software in the context of known molecular and metabolic pathways. Preliminary Pathway analysis of the differentially regulated proteins in acute vs. convalescent GCM was significant for the highlighting the immune defense pathways and the lipid metabolism pathways., suggesting that these pathways might play a significant role in the pathogenesis of giant cell myocarditis.

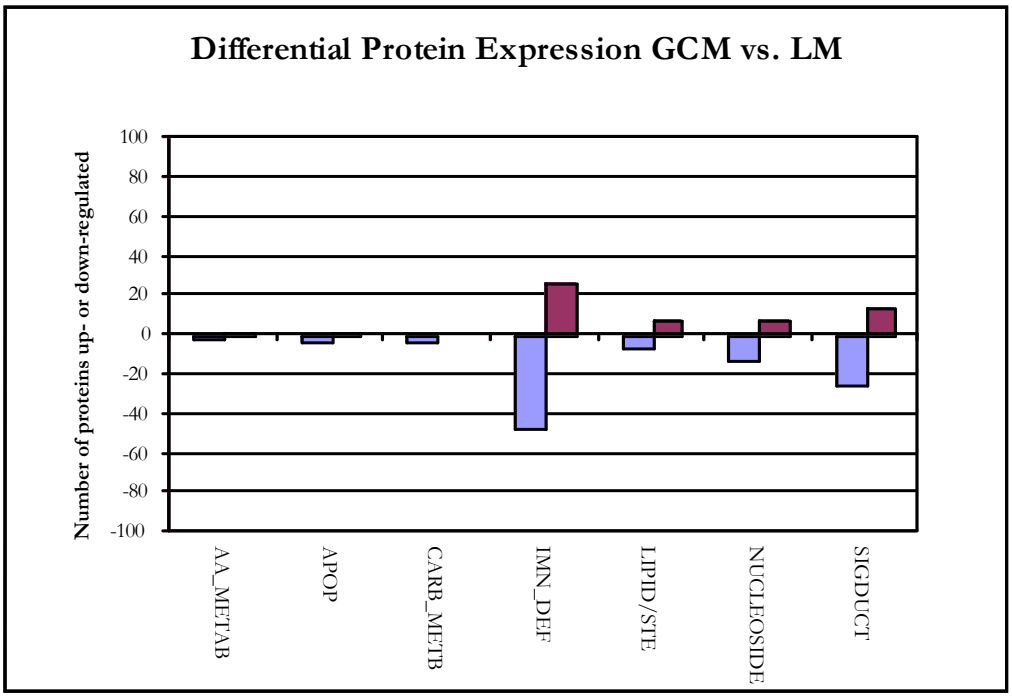


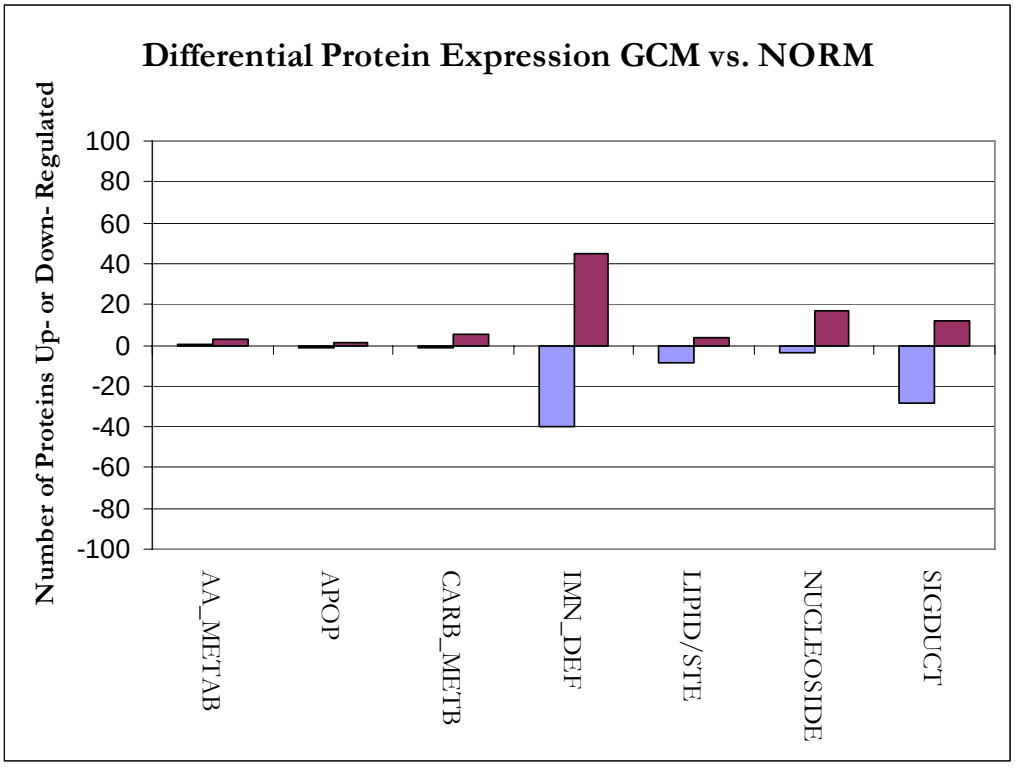
GCM Acute Vs. Convalescent : Network2 -- GRB14, CRP, DDR1, EDD, Ubiquitin

PHASE II - GCM VS.LM VS. NORMAL VS. IDCM (Preliminary Results)

In our phase II experiments, we undertook a fourplex comparison of Giant Cell Myocarditis (GCM), Lymphocytic Myocarditis (LM), Idiopathic Dilated Cardiomyopathy (IDCM) and Normal Controls. Having shown a difference between acute and convalescent GCM in the Phase I studies, the experimental design for this set of experiments was geared to identify a unique fingerprint, if present, for GCM as compared to other forms or etiologies of acute heart failure. Preliminary results for an N=3 are presented below (See figure annotation in Figure above)




Differential Protein Expression GCM vs. LM
Number of proteins up- or down-regulated
100
80
60
40
20
0
-20
-40
-60
-80
-100
AA_METAB
APOP
CARB_METB
IMN_DEF
LIPID/STE
NUCLEOSIDE
SIGDUCT


Differential Protein Expression GCM vs. NORM
Number of Proteins Up- or Down- Regulated
100
80
60
40
20
0
-20
-40
-60
-80
-100
AA_METAB
APOP
CARB_METB
IMN_DEF
LIPID/STE
NUCLEOSIDE
SIGDUCT

The preliminary results are significant for a variety in the protein expression profiles in the different disease profiles. For instance, the marked up-regulation in immune defense proteins observed in the phase one studies is muted in the comparisons between GCM and Normal, GCM and LM, GCM and DCM, suggesting that these conditions do indeed show unique protein profiles that can form the basis for discerning unique protein markers for DCM using the statistical tools described above. Furthermore, we expect to see a less pronounced difference in the protein expression profiles due to the influence of biologic variability.

Pathway Analysis GCM vs. DCM vs. LM vs. Normal Controls

Following analysis of the 6 experimental sets together, we obtained the following results: For each experimental set, we were able to identify

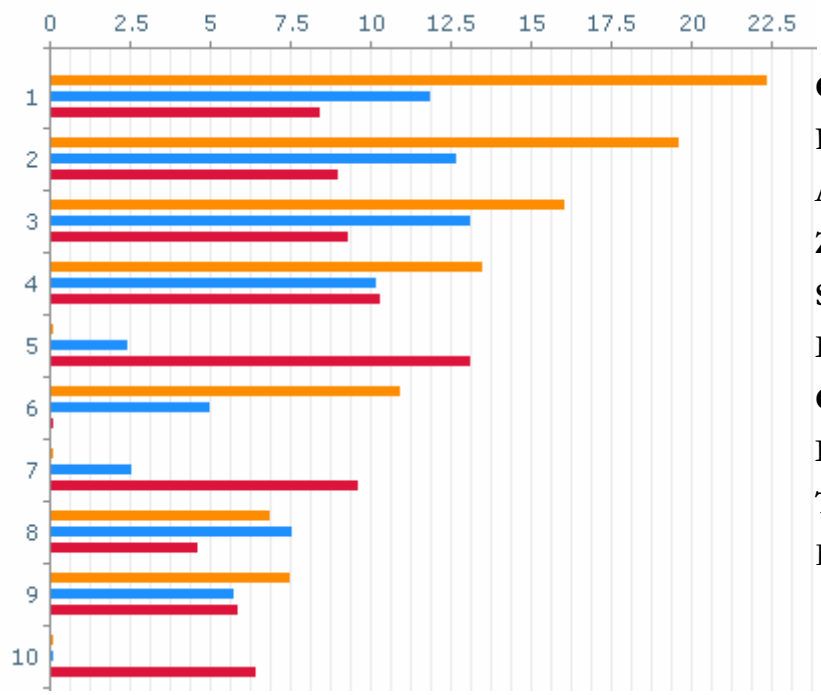
- 107,000 m/z peaks per MS run
- 2400 differentially expressed peptides
    - Redundant and single occurrence peptides removed

Using the Proquant analysis program, the differentially regulated proteins identified from these peptides were obtained and used as Metacore input.

Differentially Regulated Pathways - MetaCore

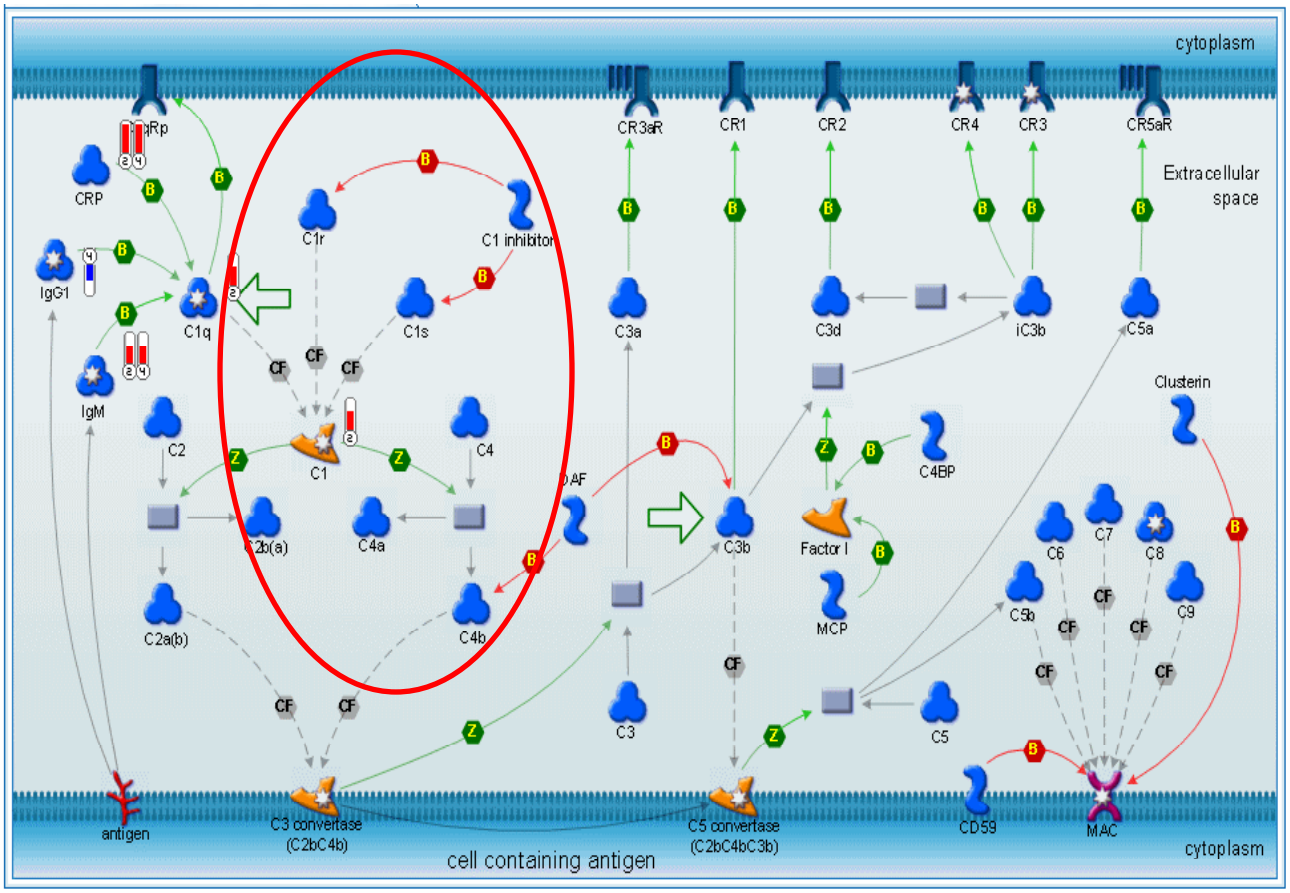
GCM vs LM	**36**
GCM vs DCM	**12**
GCM vs Normal	**14**

Top 10 Differentially Regulated Pathways GCM Vs. LM Vs. DCM Vs. Normal



Classic complement pathway
Plasmin signaling
Alternative complement pathway
ZNF202 in atherosclerosis gene expression
Slit-Robo signaling
Lectin induced complement pathway
Catherin mediated cell adhesion
ECM remodeling
TGF, WNT and cytoskeleton remodeling
Rho GTPase regulation of actin/cytoskeleton

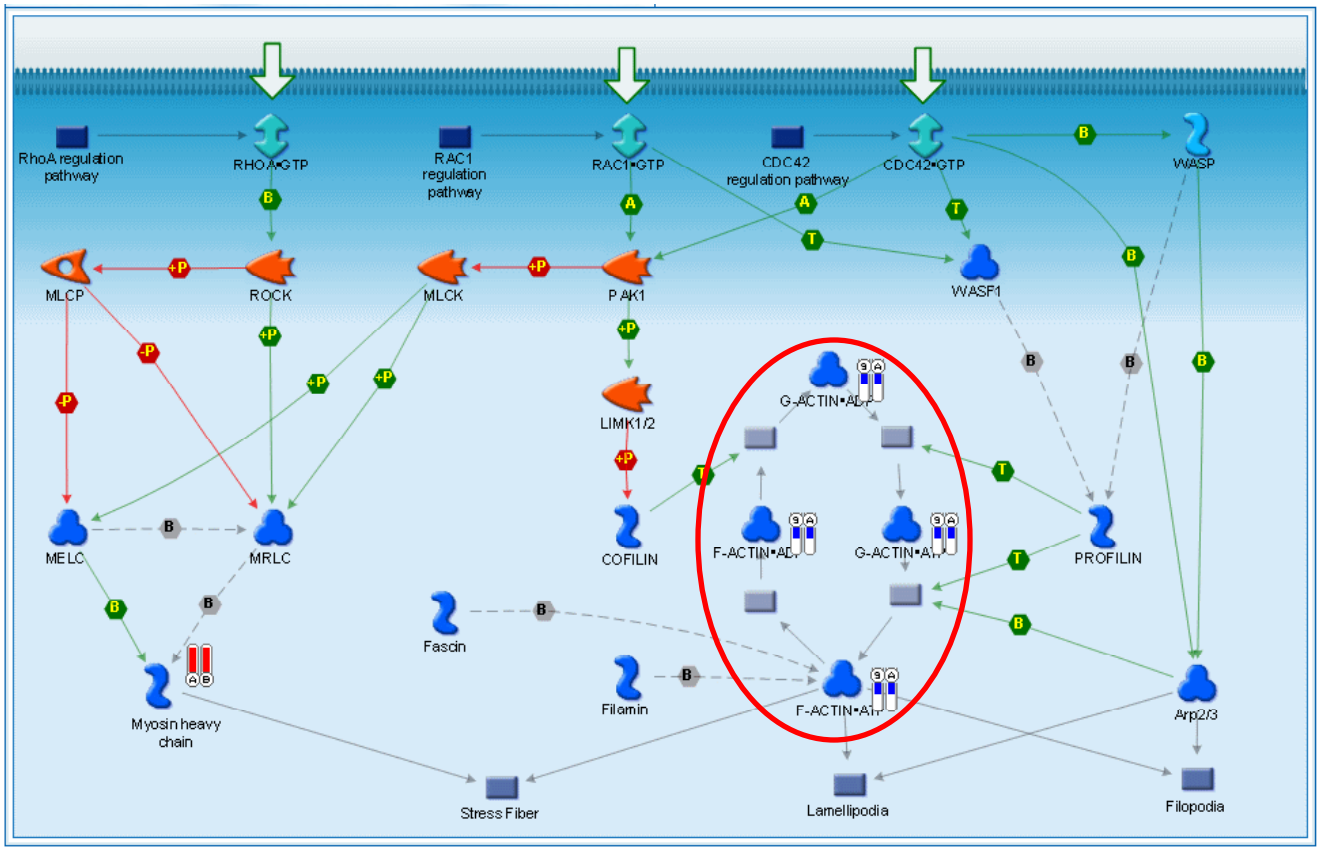
In particular, we focused on the differentially regulated pathways between GCM and LM, we see a statistically significant change in the classic complement pathway, Slit-Robo signaling pathway and the Rho GTPase regulation of the actin/cytoskeleton pathways (identified as pathways 1, 5 and 10 above). They are significantly upregulated in GCM as compared to LM. The upregulation of the immune response complements GCM in this comparison suggest the same type of immune dysregulation seen in our Phase I GCM Acute vs. Convalescent pathways. These pathways as shown below are biologically relevant in the pathogenesis of GCM.



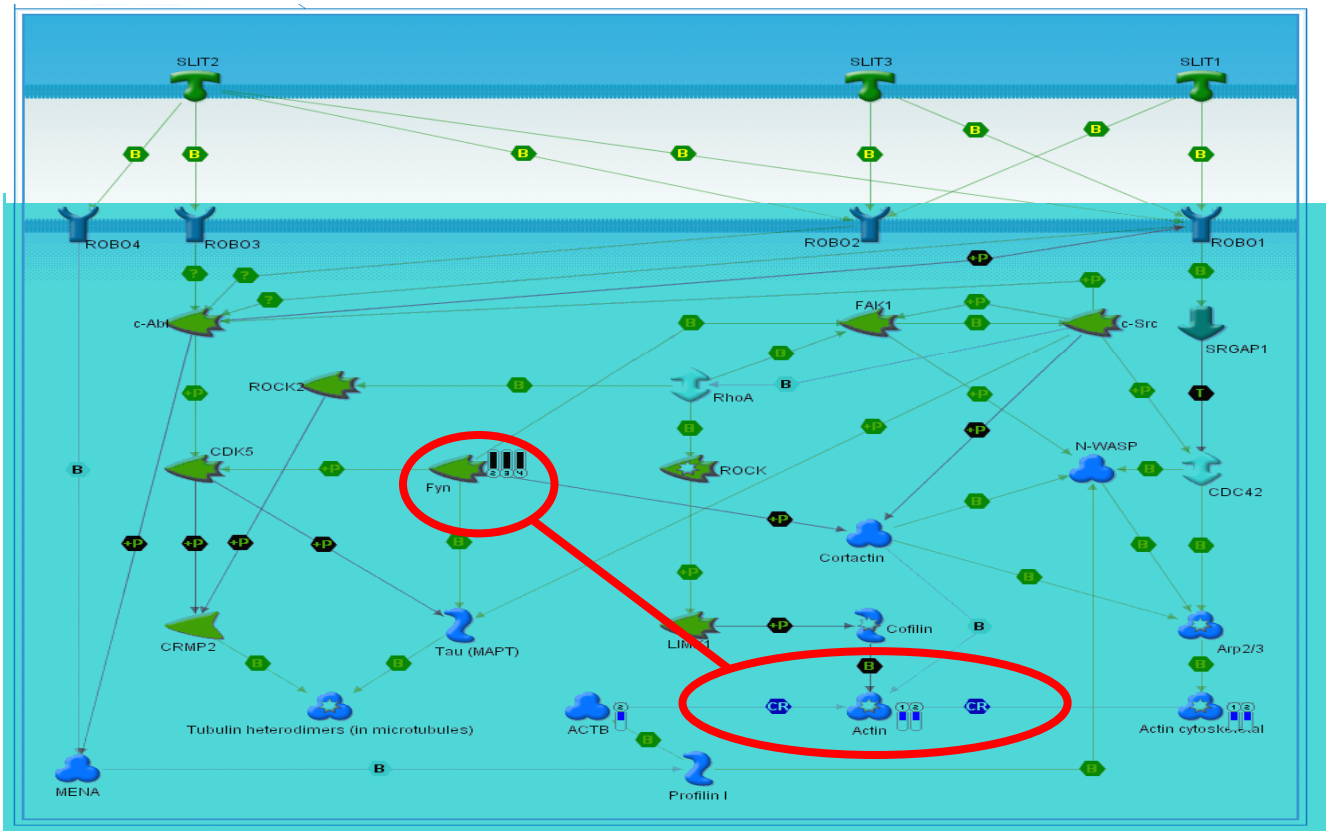
Classic Complement Pathway (GCM vs. LM)

In the Actin pathway, we see an upregulation of the cytosleletal proteins in GCM as compared with LM as shown in the pathway below. This again suggests a measure of cytoskeletal dysfunction occurs in the pathogenesis of GCM. However, when we examine the Slit-Robo Pathway, there seems to be a potential link, mediated by the fyn pathway between the immune dysregulation represented in the complement cascade and the cytoskeletal dysfunction in the Actin pathway. Of note, the fyn pathway also appeared to be significant in our analysis of GCM acute vs. convalescent.

GCM vs Lymphocytic Myocarditis Rho GTPases and the Actin Cytoskeleton

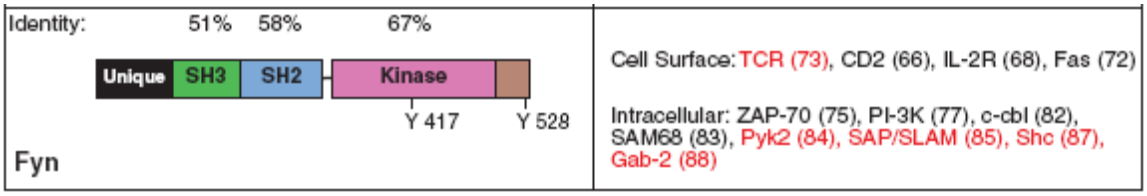


Slit-Robo Signaling (fyn pathway) in GCM vs. LM



Discussion:

Analysis of the differentially regulated proteins in both the GCM Acute vs. Convalescent studies and the GCM/LM/DCM studies was significant for the highlighting of the Fyn pathway. Fyn is 59kDa protein, a member of the src-family of tyrosine kinases which has a structure similar to the other family members: an N-terminal attachment site for saturated fatty acid addition, a unique region, a src-homology 3 SH3 and SH2 domain, a tyrosine kinase SH1 domain and a C-terminal negative regulatory domain. The domain structure allows for both tyrosine kinase function and an adapter function for larger macromolecules through the SH2 and SH3 domains as shown below.



The Fyn protein is expressed as two isoforms, one isoform Fyn T is expressed in T cells while the other Fyn B is expressed in brain and other tissues. This differential expression allows for further localization of the identified Fyn protein by its presumed origin. Thus the identification of Fyn-T in our sample is specific for augmented T cell activity in the samples from Giant Cell myocarditis.

Fyn has a variety of biologic functions, including T cell signaling, B cell development,

development of brain function, mitogenic signaling and cell adhesion mediated signaling. Of most biological relevance in this context is the role of fyn in T cell receptor signaling. Animal studies and cell line studies both indicate a critical role for fyn in proximal T-cell antigen receptor (TCR) signal transduction. It is implicated in pre-TCR signaling, positive selection, peripheral maintenance of naïve T cells and also in lymphopenia-induced proliferation[28]. Overexpression of the fyn(T) transgene is thought to produce an enhanced responsiveness to TCR signaling which might account for the increases susceptibly to autoimmunity observed in mouse models. When the fyn knockout is introduced into MPR/lpr mice, manifestations of autoimmunity are reduced and the mouse live longer[29]. Mutations in a number of signaling components in mice can lead to strong autoimmune phenotypes. Fyn-deficient mice exhibit a number of immunological abnormalities and also exhibit some autoimmunity[30]

Thus, it is likely that fyn is important in the pathogenesis of autoimmue myocarditis as manifested in GCM and might present an appropriate target for the development of diagnostic and treatment modalities.

The identification of the fyn pathway in our analysis represents one mechanism through which proteomic analysis can aid in the *a priori* identification of important mechanisms in disease and potential therapeutic and prognostic targets. However, further verification of the role of fyn in the pathogenesis of GCM needs to be carried out. We are currently working on this using in-situ hybridization to localize

the presence of fyn in GCM myocardial tissue as compared to normal myocardium to confirm that fyn is indeed a specific marker of possible T cell dysregulation in GCM.[31]

Strengths and Limitations:

Strengths:

The strengths to our approach lie in using proteomics to provide diagnostic and mechanistic insight into the pathophysiology of myocarditis. To our knowledge, this is a novel application of quantitative serum proteomics in the analysis of various histologic forms of myocarditis and dilated cardiomyopathy and it holds a lot of promise. In particular, it helps to examine the obvious, but as yet unproven, hypothesis that these distinct disease entities (by histopathology and etiology) will produce unique protein signatures. Clinically, it also provides an important novel approach to the care of the patient with acute cardiomyopathy by allowing for new diagnostic test panels based on the unique protein profiles that identify the specific histologic form, without the need for an invasive heart biopsy.

This approach also holds the potential for providing new mechanistic insights into the pathogenesis of myocarditis and cardiomyopathy since it allows for an unbiased look at the entire proteome to identify previously unknown protein changes in a disease or condition.

Limitations:

Given the high level of biological variability inherent in the plasma samples and technical variability in sample preparation and analysis, the ability to reproducibly measure a myriad of protein expression changes across numerous multivariate experiments is a key challenge in many proteomics experimental designs[32]. Specifically in the study of a relatively rare condition like myocarditis, sample size is limited by the number of available cases such that the ability to increase study power is also limited. However, in this project, we have addressed this variability using different techniques:

i) Biological Variability: We employed one-to-one matching of case to controls to reduce the confounding variables of age, sex, BMI and length of serum storage. Furthermore, in the GCM Acute vs. Convalescent studies, the use of self-matched controls in the study design minimized the level of biological variability in the comparison.

ii) Technical Variability:

   a. Randomization of matched sets to different isotope tags to prevent isotope-effect

   b. Independent validation of the proprietary protein identification and quantification platform Pro Quant using the i-Tracker system to search against the database.

iii) Statistical Variability:

a. Blinding of statistical analyst (not privy to information on diseases vs. non-diseased samples) during the clustering analysis reduces the risk of bias in the grouping of expression profiles by similarity.

6. HUMAN SUBJECTS

This project conforms to the Mayo Clinic Foundation guidelines relating to research risks when human subjects are involved. These risks are less applicable to this study since it is a minimal-risk protocol that involves the use of stored samples and no direct patient contact.

7. CONSENT

This study met the criteria for waiver of HIPAA authorization and informed consent. No consent forms were required.

8. CONCLUSION

Our results suggest that the identification of unique serum protein profiles associated with specific forms of acute cardiomyopathy is feasible using a strategy that does not require *a priori* identification of protein candidates. In our phase one studies, we have demonstrated a novel finding in possible role of altered immune response as well as altered lipid metabolism pathways in GCM. Furthermore, we

have demonstrated that GCM may be distinguished from LM by altered regulation of pathways involving the Fyn tyrosine kinase and actin cytoskeletal proteins and other immune regulatory pathways. However the further validation of protein candidates is required through biochemical assays, repeat proteomic assays and in the future, prospective in vitro studies is required in order to make any definitive conclusions about unique serum proteins in GCM. However, this implies that serum protein profiles may provide a unique protein signature for GCM. This strategy may ultimately allow for more selective use of EMB or for the noninvasive diagnosis of giant cell myocarditis.

9. FUTURE DIRECTIONS

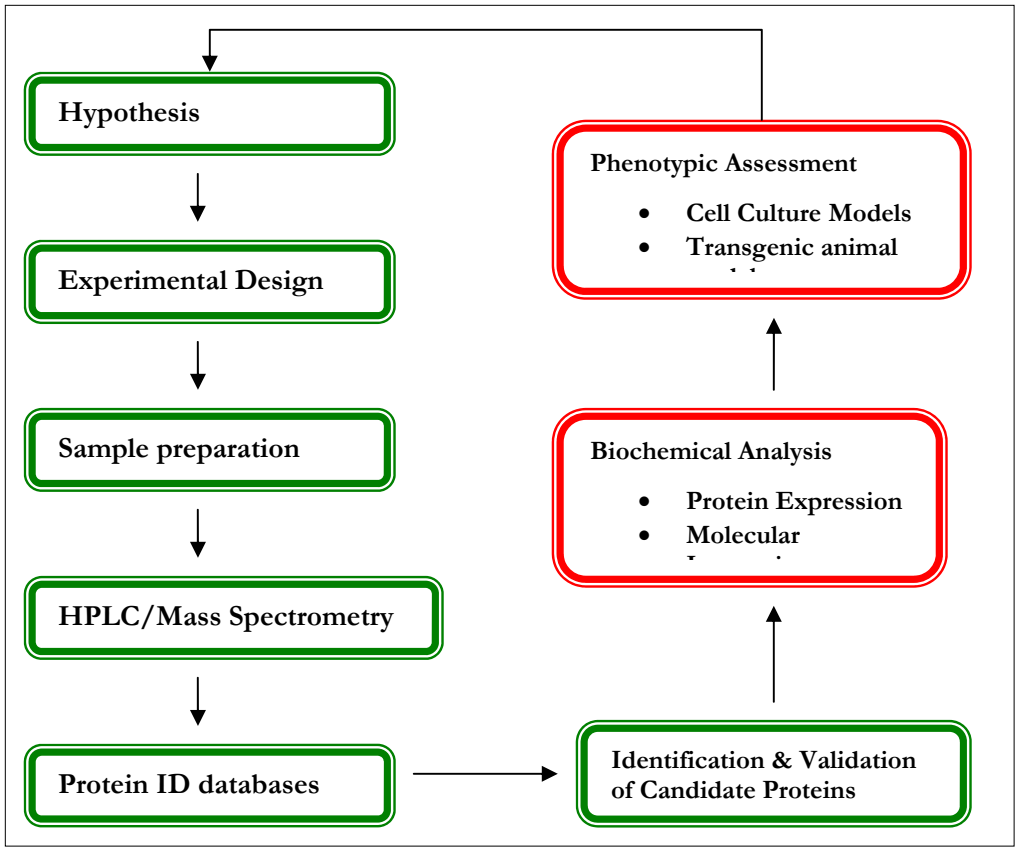


Figure 18: Strategies for hypothesis-driven proteomics

Future work on this project will involve the refining of the original hypothesis using the information on viable protein candidates obtained from this stage of the project. Thus far, we have worked on creating the hypothesis, experimental design, sample preparation, HPLC/Mass spectrometry and the identification and validation of protein candidates. To complete the flow of hypothesis-driven proteomics, there is a need to carry out further biochemical analysis and phenotypic studies. There is also a need to develop more robust statistical tools, especially for the non-linear analysis of

quantitative proteomic data. These are the directions that we hope to work on in the near future

10. LITERATURE CITED

1 Aretz HT, Billingham ME, Edwards WD, et al. Myocarditis. A histopathologic definition and classification. Am J Cardiovasc Pathol 1987;1: 3-14

2 Murphy JG, Franz RP. Endomyocardial Biopsy in Myocarditis. In: Cooper LT, ed. Myocarditis: From Bench to Bedside. New Jersey: Humana Press; 2003

3 Cooper LT, Giant Cell Myocarditis. In: Cooper LT, ed. Myocarditis: From Bench to Bedside. New Jersey: Humana Press; 2003

4 Edward D. Moloney, Jim J. Egan, Peter Kelly, Alfred E. Wood and Leslie T. Cooper, Jr., Transplantation for Myocarditis: A Controversy Revisited, The Journal of Heart and Lung Transplantation, Volume 24, Issue 8, August 2005,1103-1110.

5 PubMed Definition

6 J.W. Mason. Viral Latency :A Link between Myocarditis and Dilated Cardiomyopathy? Journal of Molecular and Cellular Cardiology (2002) 34, 695ÿ698.

7 Codd M, Sugrue D, Gersh B, Melton L. Epidemiology of Idiopathic Dilated and Hypertrophic Cardiomyopathy. Circulation 1989;80: 564-72

8 Felker G, Thompson R, Hare J et al. Underlying causes and long-term survival in patients with initially unexplained cardiomyopathy. N Engl J Med 2000; 342:1077.

9 Aretz HT, Billingham ME, Edwards WD, et al. Myocarditis. A histopathologic definition and classification. Am J Cardiovasc Pathol 1987;1: 3-14

10 Murphy JG, Franz RP. Endomyocardial Biopsy in Myocarditis. In: Cooper LT, ed. Myocarditis: From Bench to Bedside. New Jersey: Humana Press; 2003

11 Wu, LA, Lapeyre AC, Cooper LT. Current role of endomyocardial biopsy in the managements of patients with dilated cardiomyopathy and myocarditis. Mayo Clin Proc. 2001; 76:1030

12 Murphy JG, Franz RP. Endomyocardial Biopsy in Myocarditis. In: Cooper LT, ed. Myocarditis: From Bench to Bedside. New Jersey: Humana Press; 2003

13 Lauer B, Niederau C, Kuhl U, et al. Cardiac Troponin T in patients with clinically suspected myocarditis. J Am Coll Cardiol. 1997;30:1354

14 Caforio AL, Goldman JH, Haven AJ, et al. Circulating cardiac-specific autoantibodies as markers of autoimmunity in clinical and biopsy-proven myocarditis. The Myocarditis Treatment Trial Investigators. Eur Heart J1997; 18:270

15 Jaffe AS, Use of Cardiac Biomarkers for Detection of Myocarditis, In: Cooper LT, ed. Myocarditis: From Bench to Bedside. New Jersey: Humana Press; 2003

16 Anderson, Leigh, Candidate-based proteomics in the search for biomarkers of cardiovascular disease J Physiol (Lond) 2005 563: 23-60

17 Kittleson, M.M et al. Gene expression in giant cell myocarditis: Altered expression of immune response genes.

18 Granger CB, Van Eyk JE et al. National Heart, Lung, And Blood Institute Clinical Proteomics Working Group report. Circulation 2004 Apr. 13; 109 1697-703

19 Evans, G., Wheeler, C. H., Corbett, J. M. & Dunn, M. J. Construction of HSC-2D PAGE: a twodimensional gel electrophoresis database of heart proteins. Electrophoresis 18, 471–479 (1997).

20 Hanash S. Disease Proteomics. NatureVol 422. March 2003

21 Berhane, BT et al. Cardiovascular-related proteins identified in human plasma by the HUPO Plasma Proteome Project Pilot Phase. Proteomics 2005, 5, 3520-3530

22 Cooper LT, ed. Myocarditis: From Bench to Bedside. New Jersey : Humana Press; 2003

23 Betgovargez et al, Abundant-Protein Depletion Coupled with Multidimensional Fractionation of the Human Plasma Proteome. (www.beckmancoulter.com)-

24 Byrne C, Cagney B. ICAT and other labeling strategies for semiquantitative LC-based expression profiling. In: Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics: Wiley Interscience 2005

25 ProQuant Software Manual (www.abisystems.com)

26 i-Tracker: Tool for Quantitative Proteomics (http://www.dasi.org.uk/download/itracker.htm)

27 Fung, ET. Strategies in Clinical Proteomics. In: Conn MT, ed. Handbook of Proteomic Methods. New Jersey: Humana Press; 2003

28 Palacios et al. Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation. Oncogene (2004) 23, 7990-8000

29 Resh M.D., Fyn, a SRC family tyrosine kinase. Int. Journal of Biochem. & Cell. Bio. 30(1998) 1159-1162

30 Zamoyska, R et al. The influence of the src-family kinases, Lck and Fyn, on T cell differentiation, survival and activation. Immunological Reviews 2003. Vol 191: 107-108

31 Ibid.

32 Overcoming Variation in Quantitative Proteomics. Proteomics 2005, 3, 1912-1919