# Whole Genome Expression in Mice Containing a Human Mannose Binding Gene (hMBL): Examining the Immunological Role of Monoclonal Antibody (mAb) 3F8 In Attenuating Myocardial Ischemia-Reperfusion Injuries

**By**

**William Brian Gorsuch**

**A Dissertation Submitted to**

**Rutgers, The State University of New Jersey**

**School of Health Related Professions**

**In Partial Fulfillment of the Requirements for the Degree of**

**Doctor of Philosophy**

**Department of Health Informatics**

**Biomedical Informatics Program**

**October 2014**

RUTGERS
School of Health Related Professions

**Final Dissertation Approval Form**

Whole Genome Expression in Mice Containing a Human Mannose Binding Gene (hMBL):

Examining the Immunological Role of mAb 3F8 in Mediating Myocardial Ischemia-Reperfusion Injuries

BY

William Brian Gorsuch

Dissertation Committee:

Gregory L. Stahl, Ph.D., Professor, Brigham Women's Hospital, Harvard Medical School

Masayuki Shibata, Ph.D., Associate Professor, Rutgers, SHRP

John Quackenbush, Ph.D., Professor, Dana-Farber Cancer Institute & Harvard School of Public Health

Gwendolyn M. Mahon, Ph.D., Associate Dean, Associate Professor, Rutgers, SHRP

Approved by the Dissertation Committee:

_____ Date _____

_____ Date _____

_____ Date _____

_____ Date _____

_____ Date _____

_____ Date _____

**TABLE OF CONTENT**

**Whole Genome Expression in Mice Containing a Human Mannose Binding Gene (hMBL): Examining the Immunological Role of mAb 3F8 In Attenuating Myocardial Ischemia-Reperfusion Injuries**

William Brian Gorsuch

Rutgers Biomedical and Health Sciences

## ABSTRACT

During a myocardial ischemic event, acute occlusion sets in motion cell necrosis and myocardial tissue injury referred to as a myocardial ischemia-reperfusion (MI/R) injury. The resultant injury is triggered by an immunological response of which a major contributor involves the complement cascade of innate immune system involving mannose binding lectin (MBL). Few anti-complement therapeutics however have been approved for clinical use. Those studies to date have involved extensive whole genome expression in murine models of MI/R injury to assist in drug target elucidation. Studies performed have examined genomic traits and expression of mouse MBL (mMBL), which is not one hundred percent homologous to human mannose binding lectin (hMBL).

In this study, novel $hMBL^{+/+}$ mice treated with a novel mAb 3F8 were protected from MI/R injury as measured by area at risk and myocardial infarct staining when compared to control mice. Whole genome expression with the use of microarray was performed between hMBL mice undergoing MI/R treated with either a novel recombinant mAb 3F8 or mAb 1C10 as control. Mice treated with mAb 3F8 compared to mice treated with 1C10 revealed a significant down regulation in uncharacterized genes of the lncRNA family. Molecular modeling was used to study the three dimensional structural characteristics of mAb 3F8 recognition of hMBL. Within the hinge region of hMBL three possible locations were identified for the mAb 3F8 epitope for hMBL. These structurally similar locations offer possible insight into the ability of mAb 3F8 in protecting against MI/R injuries. These findings

will assist in better understanding the genomic role hMBL in MI/R, the ability of a novel murine mAb 3F8 to modulate those effects and aide in continued drug target elucidation.

**ACKNOWLEDGEMENTS**

I wish to thank a number of people who aided me in my research on mAb 3F8. I would like to give special mention to Dr. Masayuki Shibata and Dr. Greg Stahl. Dr. Greg Stahl from the Center for Experimental Therapeutics and Reperfusion Injury, Harvard Institutes of Medicine, Brigham and Women's Hospital Department of Anesthesiology, Perioperative and Pain Medicine, has been my research mentor from the beginning of my masters degree and all throughout my dissertation, serving as my dissertation advisor and committee chair. Dr. Stahl's enormous generosities in the use of his facilities have allowed me to gain experience in many laboratory techniques and learn from the many scientists on his research team. Dr. Stahl has been one of the most influential mentors and supporters of my academic and research work. The opportunities, generosity and support he has shown have changed my life.

Dr. Shibata from Rutgers Biomedical and Health Science School of Health Related Professions, Department of Health Informatics and Biomedical Informatics Program, has not only been my academic advisor and dissertation committee member but also a continued force of significant support and guidance during difficult times. Without Dr. Shibata's mentorship I would have never succeeded. He has challenged and taught me the true art of critical scientific thinking and problem solving.

I would like to acknowledge my other dissertation committee members. Dr. John Quackenbush and Dr. Gwendolyn Mahon. Dr. Quackenbush from the Computational Biology and Functional Genomics Laboratory at the Dana Farber Cancer Institute, Harvard School of Public Health, for his incredible generosity and time in mentoring the genomics component of my dissertation. His guidance was instrumental to my success as a doctorate candidate and researcher. Dr. Mahon, Associate Dean for Administration at Rutgers School of Health Related Professions for agreeing to serve on my committee and offering her experience in many areas.

I must also thank the researchers who sacrificed their schedules and time to assist me. Dr. Aedin Culhane from The Computational Biology and Functional Genomics Laboratory at Dana Farber Cancer Institute, whom always opened her door and time to assist me in my programming code writing and data analysis. Dr. Vasile Pavlov for his ongoing mentorship in the laboratory, surgical suite and sacrificing his own work to teach and assist me. Dr. Chenhui Zou for patiently mentoring me in my genotyping and polymerase chain reaction work. Lastly, but far from least Margaret Morrissey from the Stahl Lab. I am and will be forever grateful for Ms. Morrissey's kindness, generosity and mentorship in not only my lab research but in my entire research career in the Stahl Lab. She has been a friend, teacher and mentor.

A special mention must also be made for Dr. Syed Haque, Dr. Scott Magnuson and Karen Torres. Dr. Haque, Chair, Rutgers Health Informatics envisioned a doctorate program that was innovative and through his generosity, allowed me to be the first doctorate student to enroll in this new hybrid program. Dr. Magnuson and Karen Torres from GenUS BioSystems for assisting with developing our RNA protocols. Their assistance and expertise was greatly appreciated and invaluable to this study.

Many thanks to my family. My wife Lynne Gorsuch for sacrificing her own dreams and goals and for allowing me the time to pursue my dreams, the patience and understanding that comes with that, and for all the endless support in my undertakings. My daughters Emily and Kailey for allowing me to take time away from them to complete my PhD. and for teaching me that life is precious.

## LIST OF TABLES

## LIST OF FIGURES

## ABREVIATIONS

| | |
|---|---|
| $hMBL^{+/+}$ | human mannose binding lectin knockin mouse with mouse gene removed. |
| %I/AAR | percent infarct of the total area at risk |
| pH | measure of acidity/basicity of aqueous solution |
| C5b-9 | complement component terminal complex |
| C1q | complement component C1q |
| C3 | complement component C3 |
| B10A hearts | mice congenic for $H_2$ Halotype |
| MHC | major histocompatibility complex |
| IgGq | immunoglobulin G |
| IgG2b | immunoglobulin G |
| WT | wild type |
| IgM | immunoglobulin M |
| J-chain | joint chain polypeptide |
| RBCs | red blood cells |
| GI/R | gastrointestinal ischemia-reperfusion |
| $Rag1^{-/-}$ | recombination activity gene 1 |
| I/R | ischemia-reperfusion |
| CABG | coronary artery bypass grafting |
| PRIMO-CABG | Pexelizumab for the Reduction of infarction and Mortality in Coronary Artery Bypass Graft Surgery |
| SDS-PAGE | polyacrylamide gel eletrophoresis |
| VCAM-1 | vascular cell adhesion protein 1 |
| HUVEC's | human umbilical vein endothethial cells |
| CD93 | CD 93 molecule |

| | |
|---|---|
| KRT1 | keratin 1 |
| MAD2L1 | mitotic arrest deficient-like 1 |
| CALR | calreticulin |
| Lrp1 | low density lipoprotein receptor-related protein 1 |
| APCS | amyloid P component, serum |
| FGB | fibrinogen beta chain |
| LY96 | lymphocyte antigen 96 |
| PTX3 | pentraxin 3, long |
| TLR4 | toll like receptor 4 |
| YbX1 | Y boxbinding protein 1 |
| PTPRC | protein tyrosine phosphatase receptor type, C |
| SNP's | single nucleotide polymorphism |
| mRNA | micro Ribonucleic acid |
| IL1R2/IL | interleukin 1 receptor, type 2 |
| IL18RAP | interleukin 18 receptor accessory protein |
| MMP-9 | matrix metallopeptidase |
| R | R Bioconductor programming language |
| INK4 | cyclin dependent kinase inhibitor 2A |
| HDL | high density lipoprotein |
| BCL-2 | B-cell lymph |

# CHAPTER I

# INTRODUCTION

## *1.1 Statement of the Problem*

Epidemiological reviews reveal staggering statistics concerning some of the most prominent disease processes that affect both public health and healthcare costs in the United States. A significant number of deaths per year are attributed to coronary heart disease (CHD; 425,000 deaths/year). The cost of treating this disease has dramatically affected the delivery of quality health care to the United States' public, reflecting the total annual cost of $ 444.2 billion.[1,2,3] Two of the six leading causes of death in the United States are heart disease and diabetes mellitus, of which out of the six, heart disease is number one.[1] In just heart disease alone, 1.2 million people yearly will experience their first onset of an ischemic episode. Of these, 425,000 patients will die, and 75% will do so before ever leaving the hospital.[3]

The morbidity and mortality of heart disease encompasses all socioeconomic levels. Eighty three percent of those whom die are older than 65 years and more commonly men. The risk of death however increases in females linear to their age. Demographically, the highest populations at risk for heart disease are African Americans, American Indians, Hawaiians, and Asian Americans whom are natively removed by generations.[4]

Coronary heart disease often involves a "trigger" or is a result of an immunological response. A major contributor to the immunological response is the complement cascade of the innate immune system involving mannose binding lectin (MBL), and the myocardial ischemia-reperfusion (MI/R) injuries that ensue. Previous studies have demonstrated a role of MBL in heart disease and in a limited fashion identified differential gene expression, but have not involved extensive whole genome expression studies in murine models of myocardial ischemia-reperfusion. Those studies already performed examined genomic traits and expression of mouse MBL, which, is not 100% homologous to human MBL. This has

hampered and confused not only the development of novel therapeutics, but also the true knowledge of human MBL gene expression and its ability to alter and regulate molecular pathways in controlled study models of myocardial ischemic reperfusion.

### 1.2 *Background of the Problem*

Ischemia-reperfusion injuries are characterized by an initial deprivation of blood flow and/or oxygen supply to tissues/organs followed by restoration of blood flow/oxygen supply. The initial ischemic period results in cellular changes that alter subsequent signaling pathways and molecular expression.[5] The molecular pathways associated with heart failure as a result of CHD include abnormal calcium handling, oxidative stress, neurohormonal triggering and disturbance in the cytokine signaling cascade.[6] Early studies have contributed to additional gains in our knowledge of complement's role in MI/R, as well as its interaction with MBL-dependent lectin pathway.[7,8,9,10,11,12,13,14,15,16]

Furthermore, whole genome expression using microarrays historically have already been used to study heart failure in many studies.[17,18,19,20,21,22,23] However, whole genome expression analysis has much less been utilized in studying isolated MI/R as it pertains to the involvement of MBL. Those studies that do exist have been done so in animal models studying gene expression which is not one hundred percent homologous with human MBL. In mice, MBL is encoded in not one single gene, as in humans, but in two genes, *Mbl-a* and *Mbl-c*. These two murine genes are transcribed from two different chromosomes.[24,25,26]

Complement drug-target elucidation has attempted to either completely deplete complement components or to inhibit specific pathway components, such as MBL. Few anti-complement therapeutics however have been approved for clinical use. Part of this problem is in understanding not only the drug-ligand properties, but also the off target binding effects and subsequent unintended molecular pathways activated. These studies however do not allow for an expedited bench to bedside use of novel MBL therapeutics. The current era of –omics generated technologies is rapidly changing the face of knowledge discovery. Future studies

will need to have the ability to evaluate human genes within an animal vector/model system(s) for true drug-target elucidation, including off target binding complications, which can then be assessed for more rapid bench to bedside approval of safe therapeutics.

### *1.3 Goals, Objectives and Research Questions*

The goals and objectives of this dissertation are related to the study of one of the world's most influential, devastating and costly diseases, coronary heart disease (CHD) in the setting of ischemic-reperfusion injury as it pertains to the complement system component, MBL. The main goal will be to successfully use novel mice that are hMBL-$2^{+/+}$ in studying MI/R injuries. Whole genome expression utilizing gene microarrays coupled with genomic, molecular modeling, computational biology and bioinformatic techniques will aide knowledge discovery, drug-target elucidation and allow for direct gene expression profiling for not only identification purposes, but for gene function, protein network and pathway analysis. This study using a novel mouse model expressing hMBL will assess the actions of 3F8 to modulate mouse gene expression following MI/R and subsequent pathway involvement.

By following the past events in complement knowledge and discovery, as well as examining current and future trends, drug-target elucidation may be realized in hopes of inhibiting MBL protein function for better public health outcomes. This dissertation will seek to answer the following questions relating to the influence of hMBL protein function in mice and subsequent whole genome regulation through microarray analysis in the setting of using a monoclonal antibody (mAb) 3F8, in models of MI/R. The mAb 3F8 phenotype will be compared to SHAM, control and mice receiving mAb 1C10. The mAb 1C10 is an anti-MBL-2 antibody that binds to hMBL but does not inhibit its function.[27] Additionally, we will identify potential areas of interface between 3F8 and hMBL by using existing experimental data to build interacting segments of 3F8 on the segments of crystallized hMBL.

Questions to be addressed by this dissertation:

1. During MI/R what influence does hMBL in mice have on whole genome expression?

2. During MI/R what ability does mAb 3F8 have in inhibiting the mechanism of complement MBL protein function or by inhibiting genetic transcript to disrupt complement MBL-induced gene expression in a hMBL mouse model?

3. What difference is there in infarct size to area at risk (%I/AAR) in hMBL-$2^{+/+}$ mice in a MI/R model with the following treatments; control, mAb 1C10 and mAb 3F8.

4. Correlate microarray results after per for most highly expressed genes?

5. What are the 3D structural characteristics of mAb 3F8 recognition of hMBL?

6. Are calcium binding residues evolutionally conserved?

7. Is there any effect of calcium for peptide binding to mAb?

8. Does truncation affect the disulfide bridge of MBL fragments?

9. How is microarray raw data pre analyzed prior to expression analysis? Correspondence analysis and graphing? Hierarchical tree formation?

10. How is Bioconductor R script written to import Agilent single channel microarray raw data, process data, and analyze data?

11. Does polymerase chain reaction (qPCR) assist in interpreting and correlating microarray gene expression results?

## CHAPTER II

## LITERATURE REVIEW

### 2.1 *Ischemia-Reperfusion Injuries*

Ischemia-reperfusion injuries are characterized by an initial deprivation of blood flow to tissues/organs followed by restoration of blood flow. The initial ischemic period results in cellular changes that alter subsequent signaling pathways and molecular expression.[28,29,30] Glycolysis provides adenosine triphosphate, which during ischemia becomes depleted, and results in increased lactic acid formation and a decrease in pH. As this process continues phospholipase $A_2$ converts phospholipids in the cell membrane into arachidonic acid, which functions as a precursor for the biosynthesis of leukotrienes and prostaglandins.[31] Activation of polymorphonuclear leukocytes (PMN's), eicosanoids, cytokines, reactive oxygen species (ROS) and complement products are involved in this initial phase.[32] The intracellular and extracellular accumulation of these products triggers homeostatic pathways involving necrosis, apoptosis and possibly autophagy. The apoptotic response may then conclude with potential permanent tissue or end organ dysfunction.

A reduction in oxygen supply induces neutrophil adherence to endothelial cells.[5] A hallmark of the reperfusion period is characterized by increased leukocyte adhesion to the vascular endothelium.[33] Expression of leukocyte adhesion molecules during the ischemic period allows for their increased anchoring to the vascular endothelium by P-selectin and L-selectin.[34,35] Leukocyte accumulation during reperfusion induces significant increases in permeability of post capillary venues and toxic product deposition.[36]

### *2.2 Myocardial Ischemia-Reperfusion Injuries*

Thermal and chemical properties are needed for the normal heart contractile function. Those properties are derived from the oxidative metabolism of fatty acids, lactate and glucose. Energy and oxygen consumption are symbiotic, with energy being lost to both enthalpy and

cardiac muscle contraction. Myocardial ischemic reperfusion (MI/R) injuries are a prelude and present themselves as life threatening injuries due to a double mechanism of injury: myocardial ischemia followed by a reperfusion injury. Myocardial ischemia results in a heat efficiency of 5-10% compared to 25% in a non-ischemic heart. During myocardial ischemia, a rise in temperature results in accelerated heart rate and a resultant decrease in ventricular preload. This preload decrease results in a lower stroke volume; output then weakens due to weak contractile force. Lower temperatures also pose a detrimental physiological effect by increasing cardiac cell permeability. The aforementioned processes culminate in what is termed a myocardial infarction (MI).

A physiological adaptive process occurs during MI. The resultant altered myocardial cells preserve a substantial portion of the remaining healthy cardiac cells during a steady state. Myocardial cell necrosis can occur within 6 hours post MI. After 40 minutes, a substantial amount of subendocardial zone tissue necrosis occurs. Three hours after an MI, one third of the myocardial cells may die. Following this ischemic injury period is a period of reperfusion injury. This manifests as cellular edema, an excess of calcium and free radical release from ischemic tissue harboring neutrophils, and the activation of complement components.[37]

The major complement component, membrane attack complex (MAC), consists of C5b-9 to form a terminal complex which binds to the cell membrane and causes injury. The reperfusion injury also yields the complement components C3a and C3b, which assist in chemotaxis, local tissue injury and attraction of neutrophils.[38,39] This results in continued myocardial injury mediated by not only the duration of MI/R, but in the remaining metabolic demands of the myocardial cells at the site of injury.[40,41]

### 2.3 *The Complement System and Lectin Pathway*

The classical, lectin and alternative pathways interact and comprise the complement system (Figure 1). The initiation molecules of the classical, lectin and alternative pathways

are C1q, mannose-binding lectin (MBL)/ficolins/Collectin-11 and C3b, respectively. All three pathways converge at the activation and cleavage of C3 into C3a and C3b, via separate biochemical processes. The lectin pathway initiation complexes use three different lectin pathway specific serine proteases, termed MBL-associated serine proteases (MASPs), which according to the sequence of their discovery are called MASP-1, MASP-2 and MASP-3. Of those, MASP-2 fulfils analogous activities to the classical pathway key enzyme C1s, as like C1s, MASP-2 can cleave C4 and C4b-bound C2 to form the lectin pathway C3 convertase, C4b2a. In contrast to C1s, MASP-2 can drive lectin pathway activation in absence of any of the other two MASPs.

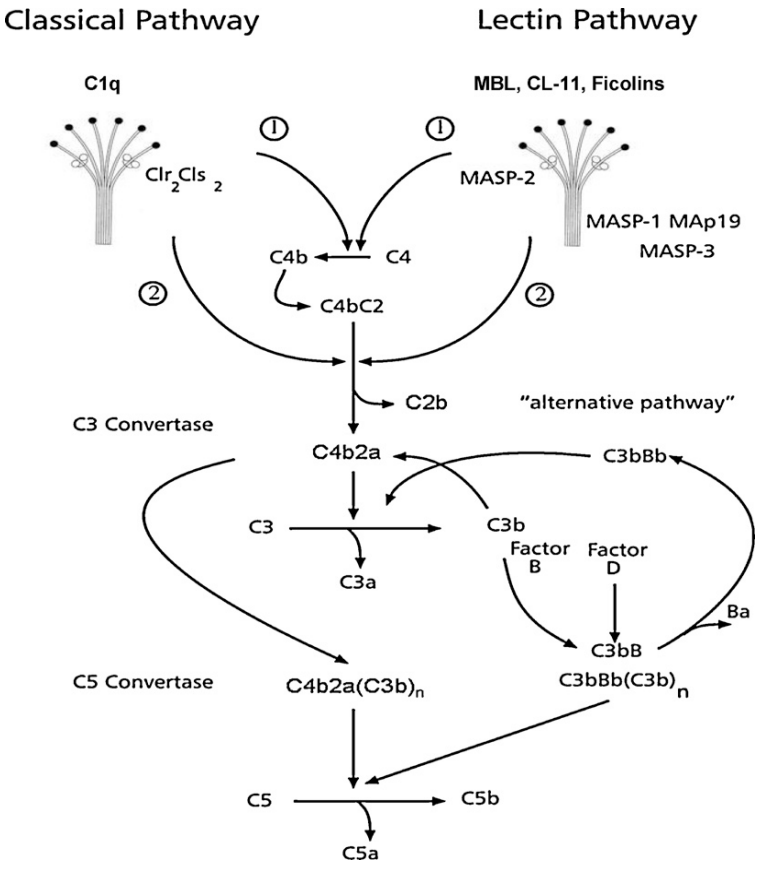


**Figure 1. Graphical Representation of The Complement System.** Adopted from Gorsuch et al. 2012.

Although MASP-1 can facilitate the process to form the C3 convertase complex by its ability to cleave C4b-bound C2, it is not capable to compensate for the loss of MASP-2 functional activity, since it cannot cleave C4.[15] This exclusivity underlines the essential role of MASP-2 in driving the lectin pathway of complement activation. Reports have indicated essential roles of the lectin pathway serine proteases MASP-1 and -3 in the initiation of the alternative pathway of complement activation.[42,43,44]

The C3 convertase complexes of the classical and the lectin pathway (i.e. C4b2a) and the alternative pathway (i.e. C3Bb) can switch their substrate specificities from C3 to C5 when multiple C3b molecules are covalently attached in close proximity to these complexes. These C5 convertase complexes, [i.e. C4b2a (C3b) and C3bBb (C3b)] finalize the enzyme mediated cascade activation events by the conversion of C5 into C5a and C5b. The anaphylatoxin, C5a, is a potent chemoattractant and triggers inflammation and activation of leukocytes, including PMNs. C5b assembles with terminal complement components C6, C7, C8, and C9 to form C5b-9. C5b-9 initiates cellular activation of nucleated cells, as well as lysis of anuclear cells.[45,46]

### *2.4 Historical Perspectives of MBL Research in MI/R Injuries*

The specific role of MBL and complement activation following oxidative stress in vitro and also in vivo were studied in multiple models.[27,28,47,48,49,50,51,52] Monoclonal antibodies that recognized MBL were developed and characterized for their ability to inhibit MBL binding and lectin complement pathway activation. Functional inhibition of MBL binding to stressed endothelium inhibited complement activation and subsequent C3b deposition(Figure 2).[27,28,47]

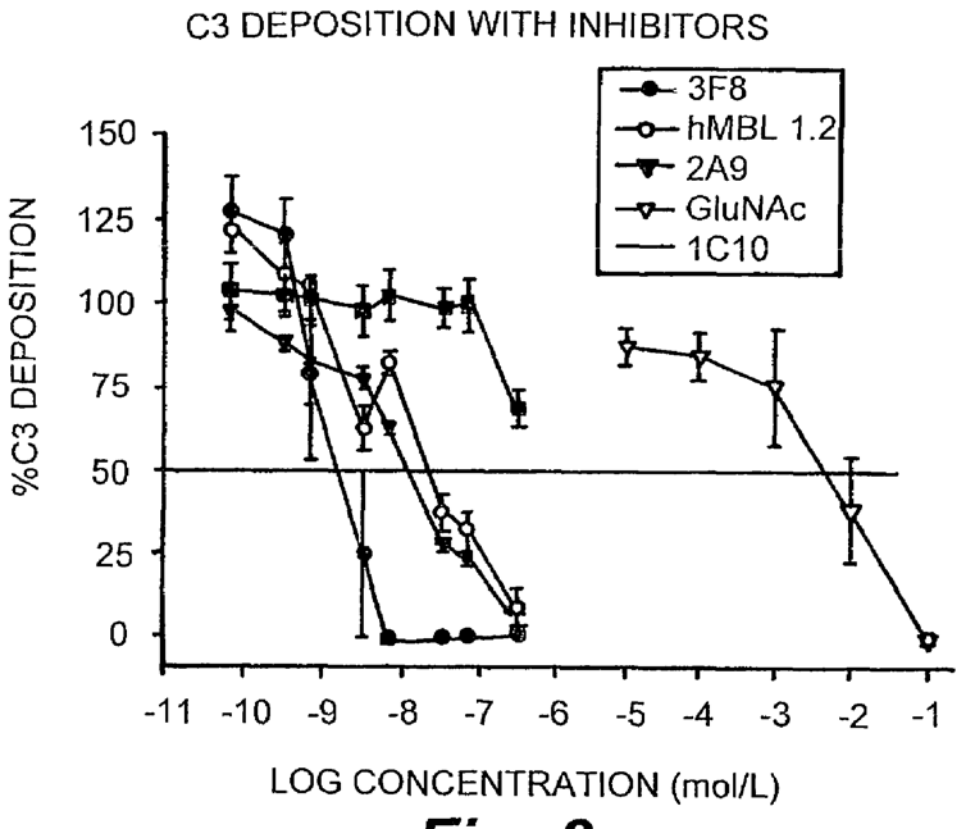


**Figure 2. Complement C3 Deposition on Myocardium**. The amount of complement C3 deposition on mannan-coated plates using complement inhibitor 3F8 and others. Re-print permission granted 9/25/12 (Collard et al., 2000).

Reperfusion of ischemic myocardium was necessary for not only MBL binding but also C3 deposition in the rat.[50] Preservation of rat myocardium from MI/R injury, inflammation and complement activation was observed with anti-MBL-A mAb treatment.[51] These early studies led to continued research on the role of classical, lectin and alternative pathways in I/R injury, as well as additional models of human disease.

### *2.5 MBL Knowledge Discovery in Current Models of MI/R injury*

Research into complement's role in the immunological response to MI/R has produced many studies in the literature. These early studies have also contributed to additional gains in our knowledge of complement's role in MI/R, as well as its interaction with

MBL-dependent lectin pathway complement activation.[9,10,11,12,13,14,15,16] Murata et al (2007) examined the role of MBL in humoral rejection of B10A hearts transplanted into immunoglobulin deficient (Ig-KO) mice. Ig-KO mice given monoclonal antibody to MHC class 1 antigens demonstrated no C4d deposition. Further, reconstitution with IgG1 or low dose IgG2b also did not deposit C4d. In contrast, IgG1 and IgG2b reconstitution demonstrated complement activation and C4d deposition on the coronary endothelium. Additional in vitro studies demonstrated that this complement activation was MBL-dependent and raised a novel hypothesis of non-complement activating antibodies and MBL in humoral rejection.[10]

The role of MASP-2 in the MBL complex in MI/R injuries has more recently been examined.[15] Novel MASP-2 $^{-/-}$ mice or an inhibitory antibody against MASP-2 were used in mouse models of MI/R. MASP-2 $^{-/-}$ mice were protected from MI/R injury of the myocardium compared to wild type (WT) mice. An additional line of evidence, which underlines a critical role of the lectin pathway component MASP-2 in human I/R comes from a recent clinical study showing that MASP-2 levels are significantly decreased in acute MI patients compared to control. Further, coronary circulation of MASP-2 is reduced following global ischemia in patients and correlates with plasma cardiac troponin I levels.[54] These data confirm the important role of the MBL complex in MI/R and suggest that functional inhibition of MASP-2 or MBL affords tissue protection following I/R.

IgM and complement components are co-localized in human infarcted myocardium.[55] While MBL and IgM do bind together, the J-chain interferes with MBL binding when IgM engages its ligand.[56,57] However, compared to pentameric IgM, the hexameric form of IgM is a more efficient complement activator.[58,59] *In vitro*, IgM binds MBL and activates complement on human endothelial cells and sensitized human RBCs.[60] In a model of GI/R, using mice that lack circulating IgM (sIgM) and MBL-A and -C, complement activation and tissue injury were present when both sIgM and MBL were reconstituted, in preference to each alone. Similar findings were also observed in a mouse model of MI/R using the sIgM/MBL null mouse line and reconstitution with MBL and/or IgM.[14] These data were also confirmed in a similar GI/R

model using Rag1$^{-/-}$ mice and confirmed co-localization of IgM and MBL, whereas C1q$^{-/-}$ mice were not protected from GI/R injury.[61] Combined, these findings suggest that IgM binds to the neo-epitope presented during I/R with subsequent MBL binding and complement activation.

### *2.6 Complement and Drug-Target Elucidation in MI/R Injuries*

Complement drug-target elucidation has attempted to either completely deplete complement components or to inhibit specific pathway components, such as MBL. Few anti-complement therapeutics however have been approved for clinical use. Early MI/R studies used cobra venom factor (CVF) to deplete complement to reduce inflammation and tissue injury.[62,63,64,65] However, because of immunogenic responses, CVF has not been used clinically. A humanized recombinant form of CVF, HC3-1496, was developed [66] and depletes C3, but does not form a functional C5 convertase. In a mouse MI/R model, HC3-1496 was equi-effective to CVF in reducing infarct size and preserving cardiac function following MI/R.[67]

The first complement specific inhibitor developed for future clinical use was complement receptor type one, sCRI (e.g., TP10). TP10 has been used in many models of I/R and was found efficacious in the vast majority of these pre-clinical studies.[68,69,70,71,72,73,74] However, the clinical use of TP10 has not met important endpoints to advance into further clinical trials.[75,76] C1INH complexes with C1 to inhibit C1 and components of the contact system proteases, including factors XIIa, XIa and kallikrein, as well as the lectin complement pathway.[77,78,79,80] Preclinical studies demonstrated the cardioprotective effects of C1INH in MI/R injury.[81,82,83,84] Clinically, C1INH preserved hemodynamic performance and resulted in lower serum troponin levels compared with placebo treated patients undergoing CABG.[85] Additionally, C1INH was used as “rescue therapy” for the treatment of MI/R injury in patients following failed percutaneous transluminal angioplasty (PTCA).[86] A recombinant form of human C1INH has been made and approved for clinical use in Europe, which also binds and inhibits MBL, suggesting that this form of C1INH may have additional advantages over plasma derived C1INH.[87]

There are tremendous numbers of antibodies and proteins with inhibitory actions against various complement components in the literature and this review did not include all of them.[88] Because of the many inflammatory and cellular activating properties of C5a and C5b-9, many biologics have been designed against C5 to inhibit formation of C5a and C5b-9 or to inhibit C5a function (e.g., the anaphylatoxin or its receptor). These biologics include but are not limited to the following: eculizumab, pexelizumab, Mubodina, Ergidina, TNX-558, and Neutrazumab.[89]

Administration of pexelizumab decreased overall patient mortality associated with acute myocardial infarction in the COMMA and COMPLY trials, but failed to meet the primary endpoint.[90,91,92] Similarly, patients requiring concomitant coronary artery bypass grafting (CABG) plus cardiopulmonary bypass (CPB) and treated with pexelizumab showed reduced myocardial injury and accompanying disorders during a phase IIa clinical trial.[93] Although the primary endpoint for this study was not reached, the study demonstrated an overall reduction in post-operative patient morbidity and mortality and resulted in another clinical trial, PRIMO-CABG II. PRIMO-CABG II did not meet the primary composite endpoint of death or MI.[94] Further investigation with pexelizumab in CABG or MI is not currently planned as of this writing.

Additional inhibitors of the MBL complex have been recently developed and used in I/R models. MBL/ficolin-associated protein-1 (MAP-1) also named MAp44, is an endogenous and natural complement inhibitor.[95,96] MAP-1, at pharmacologic concentrations, displaces MASP-1, -2 and -3 from the MBL complex and significantly inhibits inflammation, complement activation, myocardial dysfunction and coagulation in mice following MI/R.[7] Polyman2 is a dendrimeric molecule comprised of multiple copies of synthetic mannoside, which is equi-effective as anti-MBL mAb in rat or mouse stroke models. Polyman2 or anti-MBL mAb demonstrated significant reductions in neurological deficits and infarct volumes when therapy was given up to 18 hours after cerebral ischemia/reperfusion.[24] These novel complement

inhibitors extend our knowledge about the importance of complement and support/extend the importance of the MBL complex and it's associated MASPs in the sequelae of I/R injury.

### *2.7 mAb 3F8*

Monoclonal antibody 3F8 inhibits MBL in what is thought to be through binding to a discontinuous epitope. This epitope is located in the hinge region of MBL; LAVCEFPI and MARIKKWLTFSL. Such binding causes a conformational change which disrupts the ability of MBL to bind to its substrate, mannose.[97] This inhibition by conformation change was further explored in the ability of the MBL pathway to generate terminal complement components. Harboe et al (2009) inhibited MBL with 3F8 which prevented the generation of the terminal attack complex C5b-9.[98]

A novel mouse mAb 3F8, like the one used in this study which contains an MBL binding CDR3 region of mAb was selected amongst others from eight parent hybridoma clones which recognized human MBL (hMBL). Fragment antigen binding regions (Fabs) were derived by placing all mAbs in digestive enzymes and placed on SDS-PAGE. Surface plasmon resonance (SPR) was used to detect mAb and Fab fragment interactions to full length recombinant MBL (rMBL). Humanized framework (FR) regions were replaced by homologous mouse FR regions. In vitro efficacy was demonstrated by the percent of C3 deposition in human sera using mAb 3F8 and by the ability of 3F8 to inhibit VCAM-1 expression of hypoxic and re-oxygenated HUVECs.[27] A proposed MBL recognition site for 3F8 was derived using molecular modeling (Figure 3).[99]

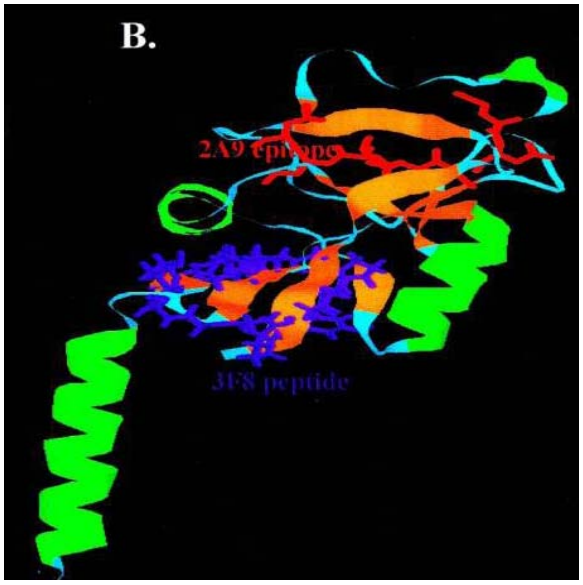


**Figure 3**. **Structural Alignment of mAb 3F8 and MBL.** Colors represent green for helices, orange for beta sheets, cyan for loop structures, blue for flitrix peptide recognized by monoclonal antibody (mAb) 3F8 and red for flitrix peptide recognized by mAb 2A9 on the crystal structure of human mannose binding lectin (hMBL) PDB code (1HUP)( Zhoa et al. 2012).

### *2.8 MBL Genetics*

Mannose binding lectin evolved 565 million years ago with homologs having been found in man's closest vertebra relative, the ascidians.[100] Various peptide regions have been evolutionarily conserved within MBL. Human MBL is encoded by a single gene.[24,25] The human MBL gene is termed *MBL2* or collectin subfamily member 2 (*COLEC2*). Humans also possess an *MBL1* pseudogene (*MBL1P1*) which is thought to have risen from MBL evolutionary gene duplication.[101] It is speculated that *MBL1* became inactive during human evolution. Both *MBL2* and *MBL1P1* are positioned close to chromosome 10 and are comprised of four exons interrupted by three introns. Encoded in exon 1 is a cysteine rich domain (1-21) containing a signal peptide with seven repeated GLy-Xaa-Yaa motifs (22-81) seen commonly in collagen structures.[24]

A glucocorticoid responsive element is located in the promoter region.[100] This process of motif repetition continues 12 more times in exon 2. A neck region (82-115) is encoded in

exon 3. Exon 4 contains a carbohydrate recognition domain (CRD)(116-228). The CRD domain recognizes microbial carbohydrates such a mannose and N-acetylglucosamine sugar motifs. The carbohydrate specificity is correlated within specific tripeptide signature sequence motifs. For mannose recognition these include Glutamic acid 192 (E), Proline 193 (P), Asparagine 194 (N), Tryptophan 211 (W), Asparagine 212 (N), and Aspartic acid 213 (D). Galactose recognition lies within Glutamine 185 (Q), Proline 186 (P) and Aspartic acid 187 (D).[102] The mannose binding pocket within the CRD is aided by hydrogen stabilization. This hydrogen stabilization is achieved by forming two coordinate bonds with calcium and four hydrogen bonds with glutamic acid 185 (E) plus asparagine 187 (N) of tripeptide EPN, and asparagine 205 (N) of WND.[103] Molecular surfaces on apoptotic host cells such as nucleic acids, metalloproteases, meprin alpha and beta are also recognized by calcium dependency.[104,105,106]

The majority of gene expression of *MBL2* is in the promoter sequence which contains several consensus elements. There is an exon 0, which is approximately 1kb upstream of exon1 and is an extra alternative exon which may also initiate *MBL2* gene transcription.[107] Exon 0 is however not translated into the post translational protein, but encoded is a MBL polypeptide which is identical to that of the dominant transcript for functioning as the alternative transcription start site for *MBL2*. Exon 0 may be responsible for initiating 10-15% of liver produced MBL.[100,108] Rhesus monkey sera have been shown to contain both MBL1 and MBL2. Humans and chimpanzees however only contain MBL2 and lack the MBL1 product, MBLA.[100,107]

Fifteen proteins interact with *MBL2*; CD93, KRT1, MASP-1, MASP-2, MAD2L1, CALR. CALCR, LRP1, APCS, FGB, LY96, PTX3, TLR4, YBX1 and PTPRC (www.genecards.com). The CD93 molecule is a receptor for complement component C1q and pulmonary surfactant protein A (SPA).[110] Keratin (KRT1) regulates activity of kinases by binding to integrin beta-1 (1TB1) and the receptor of activated protein kinase C (RACK1/GNB2L1).[111] Mannan binding lectin serine peptidase 1 (MASP-1) is the C4/C2 Ra-reactive factor (RaRF) activating

component and utilizes sugar molecules in the recognition of pathogens.[112] Mitotic arrest deficient-like 1 (MAD2L1) is a spindle component that prevents anaphase until chromosomes are aligned properly.[113] Calreticulin (CALR) is a chaperone, promoting folding through molecular calcium binding. Calcitonin receptor (CALCR) is a receptor for calcitonin.[114] Protein tyrosine phosphatase receptor type C (PTPRC) is required for antigen mediated T cell activation.[115] Finally, mannose binding lectin serine peptidase 2 (MASP-2) is required for interaction with substrate C4.[116]

The *MBL2* gene is associated with the orthologs of many species; chimpanzee [117], rat [118], cow [119], chicken [120], mouse [121], opossum, platypus, lizard, hbl4 gene of zebra fish, C4QN22_SCHMA gene of schistosome parasite, A7T2B8_NEMVE gene of sea anemone and unnamed genes of the sea squirt and common water flea.[122] In mice, MBL is encoded in not one single gene, as in human, but in two genes, *Mbl-a* and *Mbl-c*. These are transcribed from two different chromosomes.[24,25,26] The two forms of MBL possess different ligand specificities and are expressed in different tissues. Mbl-a and Mbl-c also form higher oligomeric structures and both are capable of activating complement.[24] In mice, only Mbl-a and Mbl-c as collectins can activate the lectin pathway.[123] Like human MBL2, both MBL forms in mice are produced in the liver. However, Mbl-a is expressed in the lung, kidney and testis while Mbl-c is expressed in the kidney, thymus and small intestine.[44,123,124] Functionally, Mbl-a is more similar to human MBL while structurally Mbl-c is more genetically homologous. Both human MBL and Mbl-a in mice are acute phase proteins, whereas Mbl-c is constitutively expressed.[25]

Approximately 5-10% of the Caucasian population display MBL deficiencies, which are due to homozygosity in one of three SNPs in the *MBL2* coding region. The occurrence of MASP-2 deficiency is about 6 in 10,000.[96] Deficiencies also exist in the classical pathway, alternative pathway, C3, terminal complement components, and regulators. Approximately 75% of classical pathway deficiencies are due to C4 or C1 complex which can result in systemic lupus erythema (SLE). C3 deficiency is rare but associated with severe life threatening bacterial infections.[125] These studies not only reveal the importance of

understanding polymorphisms and deficiencies in complement, but are also applicable in the development of study designs and focus for drug target elucidation in MI/R injuries.

The general genome underlying MBL, complement, acute hyperglycemic vasculopathy and the MI/R transcriptome of such processes must simultaneously be understood. Patients undergoing cardiopulmonary bypass had RNA isolated and globin mRNA depleted from whole blood postoperatively which revealed a gene regulatory network of 50 genes. The pro-inflammatory and protective pathways associated with MI/R injury initiated by CPB included up regulation of TLR-4 and -3, IL1R2/IL1RAP, IL6, Il18RAP, MMP-9, HGF/HGFR, Calgranulin-A/B, and coagulation factors V and XII.[30] It is currently not known if these gene transcripts are a "result of" or "assist in" initiation of complement lectin MI/R injury. When these clinical studies are more clearly defined, then specifically directed therapies can be developed.

### *2.9 Whole Genome Microarrays*

The *in vitro* and *in vivo* molecular techniques that have been the mainstay of complement MI/R research are now being aided by a new era of –omics based research. At the center is the use of whole genome microarrays. This is more of a pipeline process than an isolated comparison of probe set genes to experimentally derived samples. Five key processes involved include the microarray study design, preprocessing of data, inference, classification and validation of results.[126] The data derived must undergo normalization such as quantile normalization, then further analyzed by hierarchical clustering, gene set enrichment analysis (GSEA) [121,127,129] and linear models for microarray data (Limma) [130] or similar program processing. The resultant generation of the most individual significant genes of interest may then be explored as to their molecular pathway involvement, interaction of pathways, protein-protein interactions, gene-gene similarities and drug-target potentials.

Clustering in microarray experiments attempt to group genes based on similar behavior across a range of experiments or conditions. Those genes which are deemed to

share common expression patterns are assumed to also share common functions, cellular origins and regulatory roles.[131] Hierarchical clustering computes a pair-wise distance matrix between all genes, which in turn groups those genes into clusters based on genes that are nearest. As clusters are formed those undergo agglomeration which determines their distance from other clusters. The final result being an endogamy made of multiple similar clusters as nodes forming a hierarchical tree.[132]

Gene set enrichment analysis is comprised of three elements; calculation of an enrichment score (ES), estimation of significance level of ES, adjustment for multiple hypothesis testing. This method determines gene sets based on prior biological knowledge by assessing whether genes in a set (s) occur near to the top or bottom of genes in another list (l).[133] If so, the particular gene set is said to be correlated with the phenotypic class distinction.[128] The ES uses a Kolmogorov-Smirnov-like statistic to assess the degree a gene set (s) is overrepresented at the top or bottom of entire ranked gene list (L).[133] The estimation of significance level of ES is performed by first calculating a nominal P value, then permuting phenotype labels to obtain a recomputed ES for the permuted data which results in a null distribution. The nominal P value of the observed ES is calculated based on the previously derived null distribution.[129] Adjustment for multiple hypothesis testing consists of calculating a normalized enrichment score which represents the normalized ES for each gene set to account for the set size. False positive proportions are controlled by calculating the false discovery rate (FDR), which is the probability that a set (s) with a given NES represents a false positive.[127]

The advantages of GSEA are; it only requires membership from gene sets to compute ES, assess all genes in ranked list, reflects real biology by maintaining the gene-gene dependency.[134,135] The specific knowledge based approach used by GSEA can be seen in other programs such as GoMine, FatifGO, GoSurfer, EasyGo and David. These programs however differ from GSEA in that they only consider those gene sets on the top of list (L), which allows for subtle signals to be missed. Those programs also differ from GSEA in that

they produce over optimistic results by assuming gene independence. Programs other than GSEA either utilize varying statistical and ranking methods (GSA, SAFE, Catmap, ErmineJ, SAM-GS, PROPA) or limit gene-gene interactions to those which are already known as in a network-based approach (FunNet, PARADIGM, COFECO).[134]

Once the gene sets of most interest based on ES are isolated, GSEA calculates a leading edge subset. This subset assists the researcher in understanding which genes in set contributed the most to the sets ES, thus identifying them as the most important to focus on.[129] The GSEA program does this by identifying those genes in set (S) that appear on list (L) at or before the point in which the running sum reaches its maximum deviation from zero.[128] The researcher may then interpret this core set of genes within the original set as the most biologically important, as well as further grouping between gene sets on the basis of shared genes within the leading edge subsets.[135] This further allows for identification of common genes and gene sets in shared or distinct biological processes.[127]

Linear Models for Microarray Data (limma) is a bioinformatic package for gene expression analysis derived from microarray techniques. The limma package assesses differentially expressed genes by use of linear modeling.[136,137] The basic and most useful aspects of limma are in that it contains many so called “functions” called by R script code. Called functions such as *eBayes*, *topTable* and *lmfit* allow for the use of empirical Bayes statistics and output of the most differentially expressed genes in table format.[136] Replicate spots can be corrected prior to analysis with *duplicateCorrection*.[138] The end user has many choices for performing between array normalization by calling the functions *read.maimages* and *normalizeBetweenArrays*.[139] An important pre-processing step in microarray data analysis is in correcting for background probe fluorescence noise, which is carried out by using limma function *normexp*.[140]

The microarray pipeline culminates in utilizing those genes and gene sets isolated to provide for applicable molecular, pharmacological, and disease process knowledge discovery. Phenotypes are defined from molecular networks captured through global gene expression

profiling. Identifying features which discriminate between phenotypes is complicated in that many cellular functions are shared between different cell types.[141] The biological discovery derived from the expression profiles of microarray results are often used in predicting binding sites, predicting protein interactions and function, predicting functionally conserved modules and reverse engineering of gene regulatory networks.[142]

Networks of interacting genes, not individual genes or products, define phenotypes by their interactions. In doing so a collection of known gene interactions from open source material is used along with gene expression data to infer interacting human disease networks as utilized in the Predictive Networks program (PN).[143] The ability to gain access to open source data containing experimentally derived genes and diseases is imperative for gene-disease networks to succeed. One such solution is in the development of "Common" repository environments. One such example is Synapse program by Sage Bionetworks in which the curator, biologist and modeler can all interact. Synapse is a community based genomics analysis environment providing molecular disease models and the underlying datasets used to construct those.[144]

GeneSigBD which is a manually curated database and resource for gene expression signature analysis is also for open source published literature and provides a standardized resource for cancer gene signatures.[145] An important part of those curated databases used in building networks is in the ontological methods used. Nested Expression Analysis Systematic Explorer (nEASE) is one such method for gene ontology sub classification of gene expression data. Traditional gene ontology enrichment techniques are enhanced by nEASE which determines statistically enriched ontology sub terms based on co-annotation within a list of genes.[146]

Disease-gene networks allow the researcher to investigate whether or not a specific human disease and gene are related at a higher organism and cellular level.[147] Such methods led Mar et al (2011) to conclude that expression variance profiles are not randomly distributed across cell signaling networks. Rather, higher constraint and low expression variance genes

were significantly more connected to other networks and functioned more as core signal transduction members. In contrast, lower constraint and higher expression variance genes had fewer network connections and localized to the periphery of the cell.[141] The culmination of developing and curating human disease and gene networks for success in drug-target elucidation by combining drug and gene similarity measures.[148] Thus, similar diseases may potentially be treated with the same drug. This is achieved through network based elucidation of disease similarities revealing common functional models finalizing in a comprehensive disease associated gene database.[149]

Real time polymerase chain reaction (RTPCR) may be performed following final microarray derived gene expression findings for correlation of results.[150] It is important however to consider the effect of using single versus multiple housekeeping genes for normalization which may result in large errors during validation.[151] The process entails sample collection, RNA extraction and purification, analysis of nucleic acid quantity and quality, selection of a reference gene, mRNA quantification, and expression level calculations. The resultant data can then be used in the researchers preferred program mode of statistical analysis.

### *2.10 Whole Genome Microarray Use in Cardiovascular Disease*

There have been few studies examining the role of MI/R using whole genome microarray use. Those studies focused on MI/R in the setting of anesthetic preconditioning.[152,153] No study exists using whole genome microarray in studying either mouse or human MBL or any therapeutics to inhibit MBL pertaining to MI/R. Most studies to date have relied on using polymerase chain reaction (PCR) for a biased look at specific 'hand picked' genes.

Single nucleotide polymorphisms (SNP) can be of relevance in studying whole genome expression in regards to gene up or downregulation in cardiovascular disease. Such an example is in the SNP located on chromosome 9p21.3 which is associated with risk of coronary artery disease (CAD) and myocardial infarction.[154] Platelet function and platelet-

leukocyte interaction, both involved with cardiovascular pathogenesis, were also noted to be associated within this chromosomal region. More recently, a non-coding RNA region, antisense RNA in the INK4 locus (ANRIL) was found to overlap the 9p21.3 region. Those patients homozygous for SNP's associated with cardiovascular risk in this region also had an increase expression of ANRIL. Since the SNP's in region 9p21.3 are known to be associated with not only atherosclerosis, but also alter the binding domain of STAT1, which mediates response to inflammation.[155]

The importance of the p38-MAPK pathway and gene expression as pertaining to the expression effects of hyperglycemia on vascular tissue and thus atherosclerotic disease was examined in a model of diabetic induced cardiomyopathy.[156] The expression of ATP-binding cassette transporter A1 gene (ABCA1), which is responsible for regulating lipid influx from cells, was found to be down-regulated in induced hyperglycemia. Also noted was that hyperglycemia upregulated the expression of p38-MAPK, which in turn may play an active role in downregulation of ABCA1 and thus mediating atherosclerotic disease.[157] Hyperglycemia was also found to increase the expression of p38-MAPK while decreasing the expression of hepatic scavenger receptor class B type 1 (SR-BI), which binds HDL particles to mediate reversal of cholesterol transport and thus lower atherosclerotic disease risk.[158] No such study has focused on whether there is a link to the hyperglycemic mediated expression of the pro-death effector MAPK pathway (p38-MAPK) resulting in upregulation of BCL-2 and thus anti-apoptotic signaling and the complement lectin pathway.

### *2.11 Epitope Knowledge Discovery*

As part of drug discovery the development of immunological biologics has significant importance including the proper identification of the antigen-antibody epitope.[159] The foundation of drug design begins with the understanding of the epitope molecular recognition.[160] In doing so the most accurate means of determining and measuring the epitope is in solving it's three dimensional structure.[161] If no such three dimensional structure

exists between the antigen-antibody complexes then phage display libraries may be used to map the epitope location. Random peptides are generated and those with high affinity for binding to the antibody are selected. These peptides mimic similar physiochemical properties and spatial constraints as the native antibody epitope. Thus, when the linear regions of such peptides resemble that of the epitope the location is considered genuine. Epitopes are possibly conformational, being joined by the folding of discontinuous regions within the specific antigen.[159]

The use of linear random peptides has inherent problems. Linear peptides offer a large degree of conformational space to which many possibilities for varying conformations exist. This problem is solved by constraining the random peptides into more controllable and predictable regions, such as in the use of the Thioredoxin protein active loop. This more accurately assures only the most favorable conformations will be adopted. The resultant constraint forces those peptides to inherit a tertiary structure more homologous to that of the conformational epitope.[162]

### *2.12 hMBL Epitope Discovery for 3F8: Review of Stahl Lab Research*

The Stahl lab carried out a series of experiments aimed at identifying hMBL recognition sites for novel inhibitory antibodies, including monoclonal antibody (mAb) 3F8.[97] Trypsinized recombinant mannose binding lectin (rMBL) was used to identify the region within hMBL that mAbs recognize. The rMBL was derived from Chinese hamster ovary cells and *E.coli* cells. Three sample groups included rMBL plus dithiothreitol (DTT), trypsinized rMBL and rMBL without DTT. The rMBL plus DTT group underwent reduction and alkylation of Cysteine residues thus preventing peptides from forming disulfide bonds. The second group, trypsinized rMBL had cleaved peptides at the carboxyl side of amino acid Lysine and Arginine, except when followed by Proline. The last group had rMBL without DTT and thus preserved disulfide bonds with no carboxyl cleavage. The smallest rMBL fragment recognized was a 14-kDA fragment under non reduced conditions. The epitope was then demonstrated to be within

the carbohydrate recognition domain (CRD) region of hMBL by using N - terminus deletion protein sequencing (Figure 4).

Truncated MBL (trMBL) was used to generate N and C terminal deletion constructs. Two monoclonal antibodies, 2A9 and 3F8, and one polyclonal antibody, R2.2, were used for recognition testing. The R2.2 recognized C1-C4 and N1-N3 truncations. Monoclonal antibody 2A9 recognized all N truncations and all C except trMBL C2 which, has the peptide sequence DCVLLL removed. This information led to the conclusion that DCVLLL was important for 2A9 recognition. Monoclonal antibody 3F8 did not recognize any C truncations, indicating the possible importance of peptide sequence LAVCEFPI. 3F8 only recognized trMBL N1 indicating the possible importance of peptide sequence MARIKKWLTFSL for 3F8 recognition. The peptide sequences LAVCEFPI and MARIKKWLTFSL are located in close proximity to each other in native hMBL and form a hinge region within the CRD. These findings suggested discontinuous epitope recognition for 3F8 in hMBL.

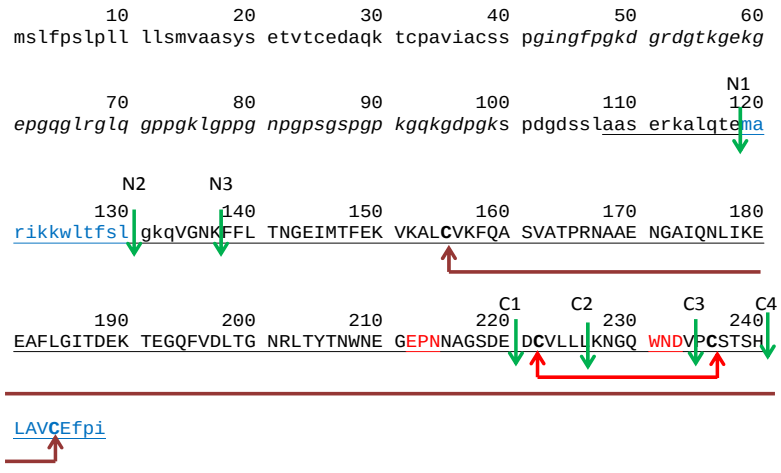


**Figure 4. hMBL FASTA Sequence With N and C Deletion Constructs**. Upper case letters represent C-type lectin domain residues. Lower case letters in italics represent collagen like domain residues. Underline residues represent crystal structures of hMBL. Green arrows show Stahl et al truncation experiments. Red arrows show disulfide bonds. Red EPN and WND residues are mannose binding residues. Blue residues represent suggested discontinious epitope by Stahl et al. (Gorsuch. W. B., inpublished data).

### *Statement of Hypothesis*

The hypothesis of this study is first, if a novel mAb 3F8 (Stahl et al., 2007) can inhibit hMBL in a novel mouse model binding to its substrate, then $hMBL^{+/+}$ mice treated with mAb 3F8 undergoing MI/R should be protected from myocardial injury as measured by whole genome expression, area at risk and infarct staining compared to control and sham mice treated with mAb 1C10. Whole genome expression should reveal; preservation of cardiac mitochondrial function, adaptive translational and transcriptional gene expression, up regulation of ubiquitin dependent proteolytic genes, altered involvement of apoptotic and autophagy pathways, reversal of cardiac contractile dysfunction genes, and continued expression of gene transcript involved in intracellular trafficking. Furthermore, molecular modeling using three dimensional structural comparison of the flitrix random peptide recognized by mAb 3F8 and hMBL should reveal the appropriate 3F8 epitope for hMBL and thus it's ability to attenuate MI/R injuries.

## CHAPTER III

## METHODS AND MATERIALS

### *3.1 Mice*

All mice were between the ages of 8-10 weeks old weighing 20-30g. Procedures were reviewed and conducted in accordance to the institutional Animal Care and Use Committee of The Brigham and Women's Hospital and Harvard Medical School (IACUC # A3431-01). Experiments were performed according to the standards and principles set forth in the National Institutes of Health (Guide for the Care and Use of Laboratory Animals-DHHS publication o.85-23, revised 2011).

Mice developed for this research were novel *hMBL-2* cDNA knock in mice in the *mMbl-1* locus using homologous recombination in mouse C57BL/6 embryonic stem cells and subsequent blastocyst injection of the appropriate targeted ES cells to create the gene targeted mice. DNA sequences were retrieved from the Ensemble database and used as reference in this project. The mouse RP23-427C7BAC DNA was used as template for generating the homology arms for the gene targeting vector, as well as the southern probes for screening targeted events. The 5' homology arm (~5.5 kb, containing exon1 and partial exon 2) and the 3' homology arm (~3.3 kb, containing partial exon 5) were generated by PCR using high fidelity Taq DNA polymerase (Stahl lab, unpublished data)(Figure 5)

The human *MBL2* cDNA ORF sequence (~0.75 kb) was amplified from human *MBL2* cDNA clone (GenBank accession #: NM_000242). These fragments were cloned in the FtNwCD or pCR4.0 vector, and were confirmed by restriction digestion and sequencing. The final vector was obtained by standard molecular cloning. Aside from the homology arms and cDNA, the final vector also contains Frt sequences flanking the Neo expression cassette (the neo cassette was used for positive selection of the electroporated ES cells), and a DTA expression cassette (for negative selection of the ES cells) (Figure 6).

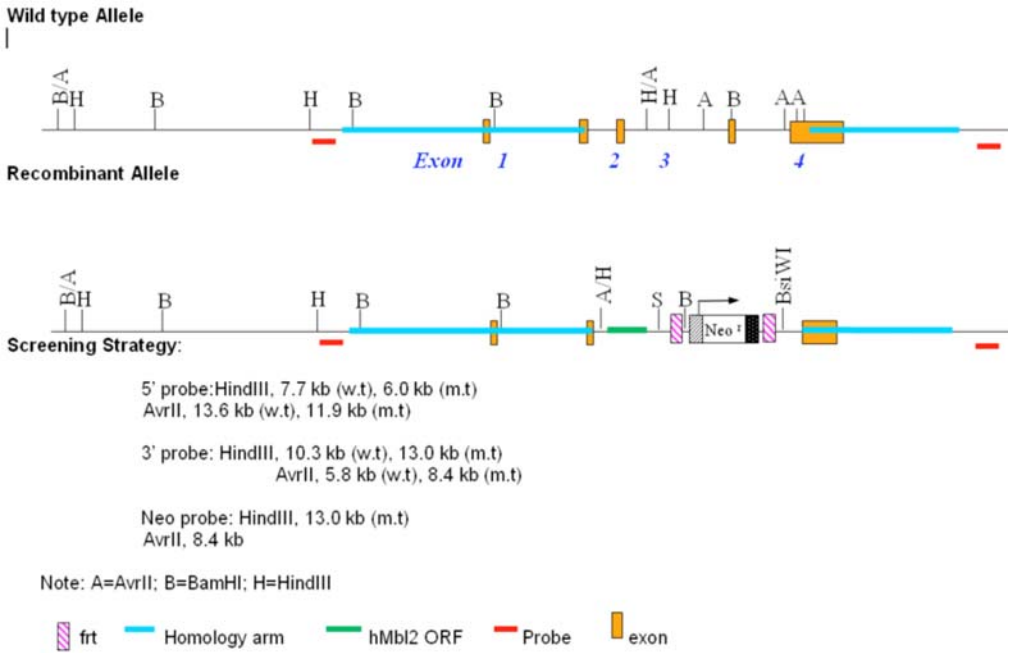


**Figure 5. Screening Strategy for Harvard Institute of Medicine-mMbl1**. Printed with permission from Stahl, G. L., on 5/1/13. Unpublished data

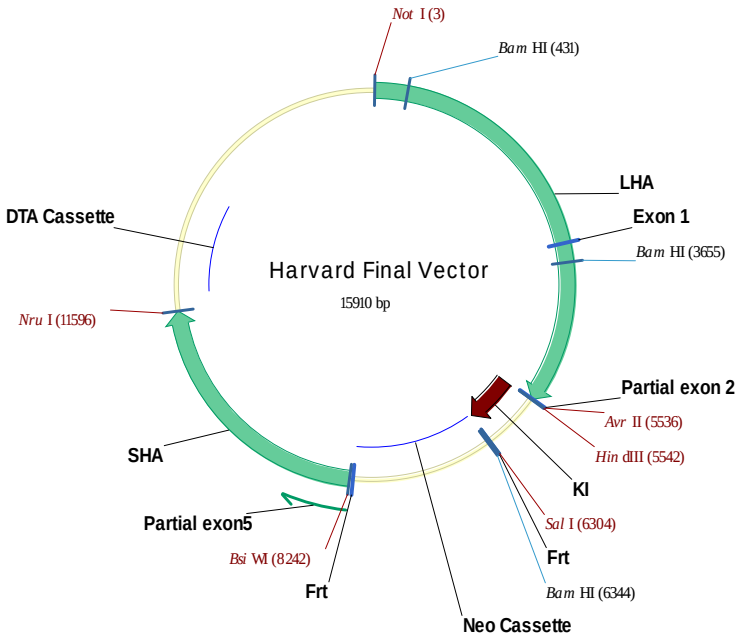


**Figure 6. Final Vector Map For Harvard Institutes of Medicine – mMbl1**. Printed with permission from Stahl, G. L., on 5/1/13. Unpublished data.

The final vector sequence for mMbl1 (homology arms in green, knock in region in red, Frt sites in italics) (Stahl lab, unpublished data):

```
   1GCGGCCGCCA ATGACCAGGG TACACAGTTG CTCATCTTTA GCTCTTCACA GACCTAGTTA
  61 TCTCCATTGT TTCTGTGTAT GGAAGACTGA GAATGTTTGG TACAACAACA TCTATAAATG
 121 TTTGTTATGA AAAGAAGTAT CTTGGAAGTC ACCTGTGAGC AGCTGTGCCT GCCGCATCCG
 181 GATCACTCAG GCCTCTGGGG TGGTCGAAAA GGAGGAGAAG GAAGAGGAGG AGGAAGAGGA
 241 TGATGATGAG GAGGAGGAAG AGGATGATGA TGATGAGGAG GAGAAGGATG ATGATGAGGA
 301 GGAGGATGAT GAGGAGGAGG TTGAGGAGGA GGAGGAGGAG GAGGAGGAGG AGGAGGAGGA
 361 GGAGGAGGAT CAGGAGTCGC ATCCTCCGCA CTGCTGCCCA CCCCGGCTGG GTTATTAGCA
 421 GAAGATGGTG GATCCCCTGG CAAACAGTGA AGCACATACT AGATGCATAG GTCCTGTGGA
 481 GAGCTGTTTT GGAGCAGCTG GTCAATCTTG ACCATCCCTG GATGGTGCTT ATTGGAGAAG
 541 GAGTATTGAC TAAGCTGTAC AGAAAGAAGC CTAAAGCAAC GCAGTCATTC TTATATGATA
 601 CGCTTGTATT TGGCAATATT GTTATCCAGA AGAAAAAAAA AATACAACAA ACAACATATT
 661 ATTCCCTCGG AAAATGTCAC CATTGATTCC ATCAAAGATG AAGGGGAATT ACAATTACGG
 721 GATGGTTGGC TTATTAAGAC ACCGACTAAG TCGTTTGCAG TTTATGCTGC CACTGCCATG
 781 GAGAAGTCAG AGTGGTGAAT CACATCAATA AGCGTGTCAC TGATTTACTC TCCTAAAGTG
 841 GGAGAACGCC TAGCTGCCGT CTGGGTTCCT GACTCTGAGG CCACCGTGTG TATGCATTGT
 901 CAGGAAGCAA AATTCACACC AGTGGATTGG TGGCACCGTT GCCGCAAATG TGGCTTTGTT
 961 GTTTGTGGTC CCTGCTCTGA AAAGAAATTT CTTCTTCCCA GCTAGTCTAA GCATGTACCG
1021 ATTTGCGACT TCTGCTATGA CCTGCTTTCC ACTAGGGACA TGATCACATG TCAGCTGACT
1081 AGATCAGACT CTTACAGTCA GTCGTTAAAG TCTCCTTTAA ATGACGAATC TGATGATGAT
1141 GACGACGCTG ATAGCAGTGA CTGAGGACAT GCTGTGAAAT ATTTAGTTGA GTCTGACTGC
1201 CTGAGAATCA ACTTTAGGGG ACGTGGAAAA TTCTGGTTTC TCTCACTTTT GCTTTAGCCA
1261 TGACTGCCTG AGAAGGTCTT ACCCTCTGTG TCTCTCTGTA TCCTATAGAA AGCAGCTCTG
1321 CGCTTTCTGC TCCCCTCTGT ACTCTCACAG AGACAAACTG CTGCAAATAC ATCTGATAGG
1381 GTCTTGGCTT CCTACTCTCC TAGTTTCTAG ATTATTTTCT TATAAGTTGG AAAAGATGTT
1441 TATTTAAGAG GTCTTGTATT ACATTTTCCC CCTGATGTGG AAAAACTACA AGCACAGAAT
1501 GATGGAACAA GAATATGATG TAACTGATAA CCGAGGCTTT CCCCCAAGCA TTCTATGTCA
1561 TGGTTAAGAC ACTGGTGACA GAACACGGTT TGGAAATTCT TTATTGTCTT TTCTCAATAC
1621 CAAGTGGCAG CATTGCCACT GAAATAAACA TAAGCCCTAC CCTCTTACTT CTGTGACTGA
1681 TTTAAATATA CTGGTGCTCA TATCAGTGGA GAATGAGTGT TTGATGGATG TAGGTAAGAA
1741 TATTTATTTG ACAGCCACAA GGAACTAAAT TTAGAGGTAT CTTTACGCTT ACATGACTAC
1801 ATATTTTTCA TTAAATAGTT GTCGTTATAT TTTTACCTCT TTTAGGCACC AAGTTATAAT
1861 GGTCACATAT ATGTTTAAAA TTAGCCAGTT GCCACTCTTA GCTTGATTTG AAATTCATCA
1921 CTAGTACAGA AGTGGCACAT TCATTAGCTT CTACTTTCCC AAGATATTTT ACATATAACT
1981 TTGTTTGCTT TACATTATTT CTAAAAGGAA AAAAATCTGA GGAAGCCCCC CCCCTTTTCT
2041 TTTCTTTTCT TTTCTTTTCT CTTCTTTTCT TTTCTTTTCT TTTCTTTTCT TTTCTTTTCT
2101 TCCCCCCCCC CCCACACACA GGGTTTCTCT TGTGTAGCTT TGCTTGTCCT GGAACTCAAT
2161 CTGTTGACCA GGCTGGCTTG GAATCCTCAG AACTCCACCT GGCTCTGCCT CCCCAGGCAC
2221 CAGTGCCACC TGGCAAGAAT CAAAAGTTTT AAGAAAATTC TCACAGTAGA ACATGGAGTC
2281 TTATAGATCT GGAAGAGATC AGAGTTTGAC TAGATTCCCT GTCTGCAAGC ATCGTCAATA
2341 CTTTTATGAA ATAACCTTGT TTGACAGCAT GCTGATACAT TTTACAGAGT TTATGAAGCC
2401 CCAAACAAAT TTTATTTTTG TAAAACAATT GTATCTTTAA CCTGTGTCTG AAATCATCTT
2461 GCAATATATG GAAATAGAAT AGCCATTGCT ATTTCTAAAT TGTGACCACC GGTCAAGATT
2521 GCTTTGAGAA CTGGTAGTTA TTCTCTATGA TTGGAGGTAG CCATTAGGCT GTGGGCCCAC
2581 TCAATGCACA CTGGATGTGC CGTGTACATT CTGTCCACTA GCATACTGTA GTAACTAGGT
2641 TTTCAAATGT GGGAAACCTG CAGGCTGACC ATTCTCTTTT GTTTTGTCAC AGTTCTTTTC
2701 AAACCCAGCG CTGCCTTTTC AAAATTTAAA ACTTGCATAA TTCCGCAGAT GTTCTGGGAT
2761 TTGATTCAGT CACTTGTGGA TGGAGAAAGA GTCATTGGTA TTTCACTATG TGGTTACAGT
2821 ATGAATTTTA CTTTTCTGCT CAGTTTTGAA AGTGTACGAG TTATGTGGGT TTCCCTGTAG
2881 GAAATGTGAA AAAAAGCACA ATAGAACATG GTCCTTGGGC TTTGTTCCTG CTCTCTGTTT
2941 TGGTTGGTTT TACTCTCAGA CGTATTTGTG TTTGTTTCAA AGCAGCATAT ACATGCAGTT
3001 AGTAAGCATG TTTGAAAGGG CCACAAAAAT CGGAGAAATG CTGCTGGGAA CTTGGACGGT
3061 CAGGAACCAG GGGGAAATGT CCTGAGATAA CATTTACATT GTCGATCTCT ATTGATTGTT
3121 TTTACAGTAA AAATATTTTA CAGCTTTTTT CTTTTTTTTT TTTTAAAGAA GTGTCTTGAT
3181 GGAGCCATGC CAAGAACCTG TGGCTATAGA GGACTGGAGT GTCTAAGCTA TATGGATGGA
```

3241 CCTCTGGGAG GAGAGAACAC AGTGGTAGGA AGATTGTAGA CTGCTAAAGC AGAGAGACTC
3301 AGCACTGAGC AGAGATCCCA CCAGCAGCTC TGAACAGCTT CCCAATTACA ACAGCTCCAA
3361 GGATAGTTGC TTATTAAACT GGGGTTGAAG GCCAGTCCAG CTT**GGTTCCC GGTCACCAGG**
3421 **CTAATGGCCC TGTCATTGGG GAGGATACTA CCAGAAGCTG GACTCGAGAC ATAGTTTCTC**
3481 **TTCCACTGCT CCTTTACTCT AAAGAAACCC TA**GTAAGTTG TGGTATGCAC TTTGGGGTCC
3541 AGCTGTGGTT CTGTGTCCTC TGCGTCCTTC CCCTCAGGGT GTTTAGATGC GGATGTACAC
3601 CAGTCTGGGA GGCTAAAGTG GAGCCCGGGA TGGGTGGAGT TCTGAGTTCT GCAGGATCCC
3661 GGTCAGGCTG GGAGTGAGTG CTGGGGCAAG AGGGTGGGGG GCGGGGGAGG GGGGTGTTGT
3721 GGAGAAGGCC TTTTGGAGGA TCTTAATGTT ACATCTCTTT TAGACAAAGG CCTTCTCCAA
3781 TGATCTTCAC TAATACAGCT CTGAGAGATC TTAGCTTGCA CATATTTCTT TATTTGGGGT
3841 GGGCTTGTAT GCATACACTT TATTGGGGTA AACTGAGGGT TATATAGTTT TGGGGAAGGA
3901 GGGGAGAATA TCCAGGGTGA GCAAAGGTCA TTGGCTGAAG GCTAAGGTAC TGAGATGCCT
3961 AATTAGCAAG GGGAGACATT CCAGGAATCT CAGGTACAGG ACTACTGACA TTCCTCAGGT
4021 GCGGACCAGG AACCCCCAGG TAGAGAAGAG GAAGCCCTGC TGACAAAGTG AGCTCAGGGC
4081 AGTTTGGGTA ATGTCAGACT TAGCAGCAAG GTTTCTTGAC AGTAAATCCT CTCAGGATGA
4141 AAGTGGTGGG TCTTGGGTCA GGAGTCTGGT CTCTCAGCTT TAACCTTGTA AGGGCCAGGC
4201 AAGGGTCTAA TGGGACAGCC CATGTAGGCC TGTATTTCTC ATCTGTGCCT TTCCTAGATT
4261 GAATTGACCA TATGCAAACT AAAGGACAGT CATTTCCAAT GGGAAAAGTA GGAGTGGATG
4321 GTGGGGAGGT GCCACCAACA AGTAACCACT CATGGGTAAC CGTGAACGGT CCTCAGCACG
4381 ATCGGAGGAG TTTTAAAAAT ATGTACAAAA GTTAGAAAGA GGAAAGATAG AGAAGGATGT
4441 AATCATGGGC GATGCCTCCC AACTTGTCAT CTCTTGCTGT TCGCTAGAGA AAGAAAAAAA
4501 ATTATGGGAA GTGGCAGAGC AGATTAAGCC AAGCCAAGTG CAACAGCTGA TCAGCCAGGA
4561 CTAAGGACAC TAGAATTAGG TCAAAGCAGA GCGACAACCC TGGGAGGCTG GTGGCTCAGG
4621 GTAGGACCTG GATAGCCAGA TTGTTCTGCT GGGGGAAGCA TGCTGAGACA CACCTCGGCC
4681 TGCATGAGCA GGCATGGTAG GAGCCTGAGC TTTTTCCTAG TAGAAATGAC CTTTTGGTGA
4741 GATTTAAAAG CAAGAAAAGT CTCCCATGGA AAGGATCAAT GAACCAAGAA GTCCCCACTT
4801 CTTTGATTGT GGTCTGCTCC CAACAAATTC GTATCACAAA ATAGAGCCTT TGAACATGTC
4861 ATTGACCAGA CTTTGAGAAA ATAACCACAA CTCTGAAGTA GGTGAGACCT CACCATGTAT
4921 AATAGGCATC TTTACTTGAA ATATGAGCCA TTGGTGGCTG AGAAATAGGT CTTGTGGGTT
4981 CTTAGGGGTT AAGGACCTAT GCAGATGGGG AAAAGTAGGA GACTGGGACT CCTGTGAAAT
5041 AGAAGAGGAA CCAAGAAAGA GAGATCCAGG GTCCGTTGGG TGCTAAGCAG TCTGCATTCC
5101 TGCCAGACCA CCACAAGCAC AGCTGGTTTA TAATTTTCAG AATTCGTAGG GTCTCGGGGT
5161 GACACAAGCT AAGCCTTTAA GCATCATGCT CACAGGGACT CTGGCCCCCT TGGGAAAACA
5221 AGCTCCCTTC CTCTTGCATC TTTCTTGGCT TTATCCCTGA TCCAACTCCC ATTTTTACTT
5281 CAGTAGCCTT GGTTGTGATG TATCTTCAGT AGCCTTGGTT GGGATGATGT ACCTTCAATA
5341 GCCTTGGTTG GGATGTACCT TCAGTAGCCT TGGTTGGGAT TATGTACCTT CAGTAGTTTG
5401 GAGTCAGAGG GTATGATATC CCTCACATCT GCTGTTCCCA GCCAGGGGTG CAGATCGCTA
5461 TAGTTAAGAA CAAAAGCTAT AAGCCAAGTG AGATGGAATT TGTCTCATTG ACTTTCTCCC
5521 CTCTCAG**GTA AGGACC**TAGG AAGCTTACCA TGTCCCTGTT TCCATCACTC CCTCTCCTTC
5581 TCCTGAGTAT GGTGGCAGCG TCTTACTCAG AAACTGTGAC CTGTGAGGAT GCCCAAAAGA
5641 CCTGCCCTGC AGTGATTGCC TGTAGCTCTC CAGGCATCAA CGGCTTCCCA GGCAAAGATG
5701 GGCGTGATGG CACCAAGGGA GAAAAGGGGG AACCAGGCCA AGGGCTCAGA GGCTTACAGG
5761 GCCCCCCTGG AAAGTTGGGG CCTCCAGGAA ATCCAGGGCC TTCTGGGTCA CCAGGACCAA
5821 AGGGCCAAAA AGGAGACCCT GGAAAAAGTC CGGATGGTGA TAGTAGCCTG GCTGCCTCAG
5881 AAAGAAAAGC TCTGCAAACA GAAATGGCAC GTATCAAAAA GTGGCTCACC TTCTCTCTGG
5941 GCAAACAAGT TGGGAACAAG TTCTTCCTGA CCAATGGTGA AATAATGACC TTTGAAAAAG
6001 TGAAGGCCTT GTGTGTCAAG TTCCAGGCCT CTGTGGCCAC CCCCAGGAAT GCTGCAGAGA
6061 ATGGAGCCAT TCAGAATCTC ATCAAGGAGG AAGCCTTCCT GGGCATCACT GATGAGAAGA
6121 CAGAAGGGCA GTTTGTGGAT CTGACAGGAA ATAGACTGAC CTACACAAAC TGGAACGAGG
6181 GTGAACCCAA CAATGCTGGT TCTGATGAAG ATTGTGTATT GCTACTGAAA AATGGCCAGT
6241 GGAATGACGT CCCCTGCTCC ACCTCCCATC TGGCCGTCTG TGAGTTCCCT ATCTGATAGT
6301 AAGTCGAC*GA AGTTCCTATA CTTTCTAGAG AATAGGAACT TC*GGATCCAC GATTCGAGGG
6361 CCCCTGCAGG TCAATTCTAC CGGGTAGGGG AGGCGCTTTT CCCAAGGCAG TCTGGAGCAT
6421 GCGCTTTAGC AGCCCCGCTG GCACTTGGCG CTACACAAGT GGCCTCTGGC CTCGCACACA
6481 TTCCACATCC ACCGGTAGCG CCAACCGGCT CCGTTCTTTG GTGGCCCCTT CGCGCCACCT
6541 TCTACTCCTC CCCTAGTCAG GAAGTTCCCC CCCGCCCCGC AGCTCGCGTC GTGCAGGACG
6601 TGACAAATGG AAGTAGCACG TCTCACTAGT CTCGTGCAGA TGGACAGCAC CGCTGAGCAA
6661 TGGAAGCGGG TAGGCCTTTG GGGCAGCGGC CAATAGCAGC TTTGCTCCTT CGCTTTCTGG
6721 GCTCAGAGGC TGGGAAGGGG TGGGTCCGGG GGCGGGCTCA GGGGCGGGCT CAGGGGCGGG
6781 GCGGGCGCGA AGGTCCTCCC GAGGCCCGGC ATTCTCGCAC GCTTCAAAAG CGCACGTCTG

6841 CCGCGCTGTT CTCCTCTTCC TCATCTCCGG GCCTTTCGAC CTGCAGCCAA TATGGGATCG
6901 GCCATTGAAC AAGATGGATT GCACGCAGGT TCTCCGGCCG CTTGGGTGGA GAGGCTATTC
6961 GGCTATGACT GGGCACAACA GACAATCGGC TGCTCTGATG CCGCCGTGTT CCGGCTGTCA
7021 GCGCAGGGGC GCCCGGTTCT TTTTGTCAAG ACCGACCTGT CCGGTGCCCT GAATGAACTG
7081 CAGGACGAGG CAGCGCGGCT ATCGTGGCTG GCCACGACGG GCGTTCCTTG CGCAGCTGTG
7141 CTCGACGTTG TCACTGAAGC GGGAAGGGAC TGGCTGCTAT TGGGCGAAGT GCCGGGGCAG
7201 GATCTCCTGT CATCTCACCT TGCTCCTGCC GAGAAAGTAT CCATCATGGC TGATGCAATG
7261 CGGCGGCTGC ATACGCTTGA TCCGGCTACC TGCCCATTCG ACCACCAAGC GAAACATCGC
7321 ATCGAGCGAG CACGTACTCG GATGGAAGCC GGTCTTGTCG ATCAGGATGA TCTGGACGAA
7381 GAGCATCAGG GGCTCGCGCC AGCCGAACTG TTCGCCAGGC TCAAGGCGCG CATGCCCGAC
7441 GGCGAGGATC TCGTCGTGAC CCATGGCGAT GCCTGCTTGC CGAATATCAT GGTGGAAAAT
7501 GGCCGCTTTT CTGGATTCAT CGACTGTGGC CGGCTGGGTG TGGCGGACCG CTATCAGGAC
7561 ATAGCGTTGG CTACCCGTGA TATTGCTGAA GAGCTTGGCG GCGAATGGGC TGACCGCTTC
7621 CTCGTGCTTT ACGGTATCGC CGCTCCCGAT TCGCAGCGCA TCGCCTTCTA TCGCCTTCTT
7681 GACGAGTTCT TCTGAGGGGA TCGATCCGTC CTGTAAGTCT GCAGAAATTG ATGATCTATT
7741 AAACAATAAA GATGTCCACT AAAATGGAAG TTTTTCCTGT CATACTTTGT TAAGAAGGGT
7801 GAGAACAGAG TACCTACATT TTGAATGGAA GGATTGGAGC TACGGGGGTG GGGGTGGGGT
7861 GGGATTAGAT AAATGCCTGC TCTTTACTGA AGGCTCTTTA CTATTGCTTT ATGATAATGT
7921 TTCATAGTTG GATATCATAA TTTAAACAAG CAAAACCAAA TTAAGGGCCA GCTCATTCCT
7981 CCCACTCATG ATCTATAGAT CTATAGATCT CTCGTGGGAT CATTGTTTTT CTCTTGATTC
8041 CCACTTTGTG GTTCTAAGTA CTGTGGTTTC CAAATGTGTC AGTTTCATAG CCTGAAGAAC
8101 GAGATCAGCA GCCTCTGTTC CACATACACT TCATTCTCAG TATTGTTTTG CCAAGTTCTA
8161 ATTCCATCAG AAGCTGACTC TAGATCCTGC AGGAATTC*GA AGTTCCTATA CTTTCTAGAG*
8221 *AATAGGAACT TC*CTCGAGAC CG**TAC**GGAAA CGAGTGCCTC CATATTCTCC TTGCCTCCTC
8281 TCTGGACTCT CACTTGCTTC CAAAGAAAAT TCAGTACTTG TTTCTCAAGC CGAGCATTAC
8341 GTGATTCTTT TGAGGGGAGA GGATATATTT GGTTGTGGCA TGAGGACATG AATGGAAGCT
8401 GACCATGAGG CACCAGGGTC TGGTTGAGCA CAGAGCAAAG GTCACACCCG TTTTGCTAGG
8461 AATACAGCAA GGAATCGTCA ATAGAGAACC TACAATAAAT AGCCAGCCAC CCCGCCTTTT
8521 CACCAAAGGA TCAAAGGCTG GCCATACCTG TGCCCGAGTG GTAGCTGATC TCAGCCATAA
8581 AAGCATCCAA TTTCCTTTCT CCGTGAATTG TCACTAAGAG GTACCAAACT TGCCTGTTTG
8641 AAGCCATCCT CTGTTTATAC GGCGTTCTTG CCTATTGTGT GTTTCTTTAC ATGTATCTAT
8701 GTTTGGTGGA GAATGGAAGA GAGCTAACAC TTCTGGATAC TTACTATACC CCAGGCATAG
8761 GCTATCTAGA TCCTCTCTGT GACTGAGCTG TTGCTTTTAG GCTCCACATA ATGCCCCAGG
8821 AGCCCTCTGT CTACCTGACT GTAAAGAAGG ACAGCAACCT TTGTGAGTAA ACACACTCAG
8881 TCCTAAGAAC TCCCTGCCAA GGAACGGTGT CCCCCCTTTA CTCACTCTGT ACCCTAAGAA
8941 TGAAATGCTG CCTTTTCGTC CTTCATTGAT ATGTGAGATT GATGTGATAT TGTGTTCTCA
9001 CAACATAGAA ATATGCAGAT GGTCATTCAG GGGGTGTCCA TCTCACAATG TTAGTCCCCA
9061 CTATCCAGCG TCTCTTCCTT TTACTGATCA TCTTATATGG ATCACTGAGC CCATTATCAG
9121 TTCATAGAGA GGAAAGTCCT GTAGGTAGCT TTTCTCATAA GTAGTTAAAT GATGCTACCT
9181 CAGCGGCCAG GACTGGACTT CCCTCCTTCA CAGTTCTGGT TCTTTCTTGG TCCAAGGTTT
9241 CGATCTATTG TGGCCATGAC GTTTTTGTTT TCCACTTATC CGAGGGTAAG TTGAGCTGAG
9301 ATTCAATAGC ACAATGAGTC AAGCCGGAAG AAGGTGAATT CTGGCCCCAG CTGCTTGCTG
9361 TTCACAAGTG AGATTGGAAA ACATCTTATG AATGAGCTAA CAGCTAACAG CTAAAGCTCC
9421 TCACTGAGGA GGCTTTGTTC ACTTCCATTT GTCTTTTACT TCTCCCAAAC TGCCAGGGAA
9481 AGCACTCTGG AGTCATTTGA TGGCCGATGG ATATTTGTTC AATGGATACA GCAAACATTT
9541 ACAAATGGAT GGATAAGTTG ACTACTAGGC AGTCAAGGTC AGTCTGTGTG CTGCTTCCAC
9601 AAAACTTTCC CAGCCGGTGT GACAAGTAGA GTCCCAGCCG GTGTGACAAG TAGAGTGGGA
9661 CCTTCAGAAG GCACCAGGGC CAGCAGAGTG CAGGTGGCAG GACCATGTGC CTTGGTTTCT
9721 AGCACACCTG TATTGTCACA TTACTGAACG TTCTTCAGAT AACACACACT TGTGATATAG
9781 AACAGAATTA TTCACTGGTT CATTTTGAAT CCAGATGTTT AAATACAGGC ACTGGGGATG
9841 GCTTAAGATA CAGAGTACGT CTACAGGATA AATTAGAATT ATAAATGTAG ATTATCCAAG
9901 AAATGGGGAC TTGGGTTTAA ATTATGTCTG TGAGTGAGTT TCCTTTGGTA GAAATGCAAG
9961 ACACAGCTAT TGGGGGGGGG GGTCATAGTT TGTTTCATGT GAAAAATTTT TAAAAACTGA
10021 TATTCAAAGA CCGAGAGACT CATGATCCAA GTTCCAGCCT GAATTCTGTA TCCTATGACT
10081 GAAGGGCCTG CTCATCAGGA CAGTCTTTCT CTGGACTCAG ACACCTCTGC TGTTGATATT
10141 GCCAGCTGTG TTCCAGCCAG AGGTAGGAAC AAAGGTTTCT AGTCAAAGGC TTGTCACTTA
10201 TACTAGATAA ATAACATCCT TCCTGGGACC AGGCAGCCTC CTTTCTCTGT GGCTGTCAGA
10261 GGTCAACCAG CTCAAAGGGC CAGGCCATAG TCTTGACCTT TGGGCATGGT TTCACTGTAG
10321 AGCCTTTGCC CTAAGACTGT TTGCTTGCTG TTATGTAACA GATCTGTACT TTGGTAATTG
10381 TCAGGAATTT GGGGTTTTTC CACATACGAA TTAGCCAGGA AAGGATCAGT TGTCATCTGA

```
10441 GGCCCAGGCT AGAGATGCCA GGAAGTTGGT TTTAAAAAGC CATGTTACTG TGGTAACCCT
10501 AAGTGGCTGA TGAGTTACTG AAGAGAGAGA GAGAGAGAGA GAGAGAGAGA GAGAGAGAGA
10561 GAGAGAGAGA GAGGTGTTCC TGACAAGAGA CAAGCTGGGA AGCAGTGACC TGTGTACCAA
10621 GTACGAACCA GTCTGCTCTT AGCTGATACA ACTACCACAT TCAAAAGCGA TAGAATTCAG
10681 TTAGGCATCT ACTAAGGAAA CACAAGCCTT AAGCTAGGTT AGCCTCTGGC AGCATGTGAT
10741 CTAGCTTGTG TTGTCCTGGC AGTCCATGTG TATTCTATGT AGCCCTTTAG AGCAGTGGCT
10801 GGCAGAATTG ACTAGAATCA TGACATTGTC TAAAATCTCT GGTCCTGTTA AATTTTTCTT
10861 AAAAGCTCCC ACCATTGTGA GGTTCTCAGG TGATCAAGTT GTTTGGCAAG GTGTCTGGAA
10921 ATCACCATCC CAAGAAATAA GTCACCCTGG TGACTTGTGT ACATCTTTTA CCCATGCCAG
10981 AAGGCAGACT TTCTTCATTT TCTGGAGGCC TCCCATGATT AAAGGGGTTC TGCTTTATTT
11041 ATAGTGTCTG TCCATTAACA CTCTCTTCCA ACTACCCCCT ACATATAACC TGTTGCTATC
11101 CAAACCCATC AGTACACATC CTAAGGTCTT AAAATCTCCC ATGACTTTAG AGAATAACAG
11161 AGCTAGGACT GAAAGGCCCT GAGGAGTTGT CAGAGTATGA GAGGTGCTGA GTTCATGAAG
11221 TCTCATGAAG GCACTTGTGT GACCAGGACT CGTACTGTTG TACATATTTT TGAGCACATA
11281 CATGTGTCTG TATGGCAATC AACAGACCCA AAGCCTAAAG CTCCCCGCAC GTCGTTGCCA
11341 GCAACTCATC CCAGCCTTCT TCCATCTGTG CCCAGGTCTC TTCTTCTCTG GTGAATATTC
11401 CAGAGGCTCT AGTCGAAGAG TATGCATGTT GGTCATGGAC AAACTGTCCA CAAATGGCCC
11461 TAGATGTGTT GAGGAGGCTG TGGAACCAGC AATGACTGCT CAGGCACTGG TATAACATTG
11521 AACAAACTTT AGATCTTGCC AATATTGTCT TTATCCAACT CATCTGACCA GTAAATCTAA
11581 GTTGGTCCCA GCTCGCGACG CGTGCTAGCC CGGGCGCTAG TTATTAATAG TAATCAATTA
11641 CGGGGTCATT AGTTCATAGC CCATATATGG AGTTCCGCGT TACATAACTT ACGGTAAATG
11701 GCCCGCCTGG CTGACCGCCC AACGACCCCC GCCCATTGAC GTCAATAATG ACGTATGTTC
11761 CCATAGTAAC GCCAATAGGG ACTTTCCATT GACGTCAATG GGTGGAGTAT TTACGGTAAA
11821 CTGCCCACTT GGCAGTACAT CAAGTGTATC ATATGCCAAG TACGCCCCCT ATTGACGTCA
11881 ATGACGGTAA ATGGCCCGCC TGGCATTATG CCCAGTACAT GACCTTATGG GACTTTCCTA
11941 CTTGGCAGTA CATCTACGTA TTAGTCATCG CTATTACCAT GGTCGAGGTG AGCCCCACGT
12001 TCTGCTTCAC TCTCCCCATC TCCCCCCCCT CCCCACCCCC AATTTTGTAT TTATTTATTT
12061 TTTAATTATT TTGTGCAGCG ATGGGGGCGG GGGGGGGGGG GGGGCGCGCG CCAGGCGGGG
12121 CGGGGCGGGG CGAGGGGCGG GGCGGGGCGA GGCGGAGAGG TGCGGCGGCA GCCAATCAGA
12181 GCGGCGCGCT CCGAAAGTTT CCTTTTATGG CGAGGCGGCG GCGGCGGCGG CCCTATAAAA
12241 AGCGAAGCGC GCGGCGGGCG GCTGCGACCT GCAGGTCCTC GCCATGGACC CTGATGATGT
12301 TGTTGATTCT TCTAAATCTT TTGTGATGGA AAACTTTTCT TCGTACCACG GGACTAAACC
12361 TGGTTATGTA GATTCCATTC AAAAAGGTAT ACAAAAGCCA AAATCTGGTA CACAAGGAAA
12421 TTATGACGAT GATTGGAAAG GGTTTTATAG TACCGACAAT AAATACGACG CTGCGGGATA
12481 CTCTGTAGAT AATGAAAACC CGCTCTCTGG AAAAGCTGGA GGCGTGGTCA AAGTGACGTA
12541 TCCAGGACTG ACGAAGGTTC TCGCACTAAA AGTGGATAAT GCCGAAACTA TTAAGAAAGA
12601 GTTAGGTTTA AGTCTCACTG AACCGTTGAT GGAGCAAGTC GGAACGGAAG AGTTTATCAA
12661 AAGGTTCGGT GATGGTGCTT CGCGTGTAGT GCTCAGCCTT CCCTTCGCTG AGGGGAGTTC
12721 TAGCGTTGAA TATATTAATA ACTGGGAACA GGCGAAAGCG TTAAGCGTAG AACTTGAGAT
12781 TAATTTTGAA ACCCGTGGAA AACGTGGCCA AGATGCGATG TATGAGTATA TGGCTCAAGC
12841 CTGTGCAGGA AATCGTGTCA GGCGATCTCT TTGTGAAGGA AACCTTACTT CTGTGGTGTG
12901 ACATAATTGG ACAAACTACC TACAGAGATT TAAAGCTCTA AGGTAAATAT AAAATTTTTA
12961 AGTGTATAAT GTGTTAAACT ACTGATTCCT AATTGTTTGT GTATTTTAGA TTCCAACCTA
13021 TGGAACTGAT GAATGGGAGC AGTGGTGGAA TGCAGATCCA CTAGGATCTA ACTTGTTTAT
13081 TGCAGCTTAT AATGGTTACA AATAAAGCAA TAGCATCACA AATTTCACAA ATAAAGCATT
13141 TTTTTCACTG CATTCTAGTT GTGGTTTGTC CAAACTCATC AATGTATCTT ATCATGTCTG
13201 GATCGTAGTT CTAGAGCGGA CCCGCCTTAA GTGAGTCGTA TTACGGACTG GCCGTCGTTT
13261 TACAACGTCG TGACTGGGAA AACCCTGGCG TTACCCAACT TAATCGCCTT GCAGCACATC
13321 CCCCTTTCGC CAGCTGGCGT AATAGCGAAG AGGCCCGCAC CGATCGCCCT TCCCAACAGT
13381 TGCGCAGCCT GAATGGCGAA TGGCGCTTCG CTTGGTAATA AAGCCCGCTT CGGCGGGCTT
13441 TTTTTTGTTA ACTACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG
13501 TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT
13561 GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT
13621 TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT
13681 AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG
13741 CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTCT CCAATGATGA GCACTTTTAA
13801 AGTTCTGCTA TGTGGCGCGG TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCG
13861 CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT
13921 TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC
13981 TGCGGCCAAC TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA
```

```
14041 CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT
14101 ACCAAACGAC GAGCGTGACA CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACT
14161 ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC
14221 GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA
14281 TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG
14341 TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG
14401 AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA
14461 AGTTTACTCA TATATACTTT AGATTGATTT ACCCCGGTTG ATAATCAGAA AAGCCCCAAA
14521 AACAGGAAGA TTGTATAAGC AAATATTTAA ATTGTAAACG TTAATATTTT GTTAAAATTC
14581 GCGTTAAATT TTTGTTAAAT CAGCTCATTT TTTAACCAAT AGGCCGAAAT CGGCAAAATC
14641 CCTTATAAAT CAAAAGAATA GCCCGAGATA GGGTTGAGTG TTGTTCCAGT TTGGAACAAG
14701 AGTCCACTAT TAAAGAACGT GGACTCCAAC GTCAAAGGGC GAAAAACCGT CTATCAGGGC
14761 GATGGCCCAC TACGTGAACC ATCACCCAAA TCAAGTTTTT TGGGGTCGAG GTGCCGTAAA
14821 GCACTAAATC GGAACCCTAA AGGGAGCCCC CGATTTAGAG CTTGACGGGG AAAGCGAACG
14881 TGGCGAGAAA GGAAGGGAAG AAAGCGAAAG GAGCGGGCGC TAGGGCGCTG GCAAGTGTAG
14941 CGGTCACGCT GCGCGTAACC ACCACACCCG CCGCGCTTAA TGCGCCGCTA CAGGGCGCGT
15001 AAAAGGATCT AGGTGAAGAT CCTTTTTGAT AATCTCATGA CCAAAATCCC TTAACGTGAG
15061 TTTTCGTTCC ACTGAGCGTC AGACCCCGTA GAAAAGATCA AAGGATCTTC TTGAGATCCT
15121 TTTTTTCTGC GCGTAATCTG CTGCTTGCAA ACAAAAAAAC CACCGCTACC AGCGGTGGTT
15181 TGTTTGCCGG ATCAAGAGCT ACCAACTCTT TTTCCGAAGG TAACTGGCTT CAGCAGAGCG
15241 CAGATACCAA ATACTGTTCT TCTAGTGTAG CCGTAGTTAG GCCACCACTT CAAGAACTCT
15301 GTAGCACCGC CTACATACCT CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC TGCCAGTGGC
15361 GATAAGTCGT GTCTTACCGG GTTGGACTCA AGACGATAGT TACCGGATAA GGCGCAGCGG
15421 TCGGGCTGAA CGGGGGGTTC GTGCACACAG CCCAGCTTGG AGCGAACGAC CTACACCGAA
15481 CTGAGATACC TACAGCGTGA GCTATGAGAA AGCGCCACGC TTCCCGAAGG GAGAAAGGCG
15541 GACAGGTATC CGGTAAGCGG CAGGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG
15601 GGAAACGCCT GGTATCTTTA TAGTCCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA
15661 TTTTTGTGAT GCTCGTCAGG GGGGCGGAGC CTATGGAAAA ACGCCAGCAA CGCGGCCTTT
15721 TTACGGTTCC TGGCCTTTTG CTGGCCTTTT GCTCACATGT AATGTGAGTT AGCTCACTCA
15781 TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT ATGTTGTGTG GAATTGTGAG
15841 CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC TACGTAATAC
15901 GACTCACTAG
```

The final vector was confirmed by both restriction digestion and sequencing analysis. The 5' and 3' external probes were generated by PCR and were tested by genomic Southern analysis for screening of the ES cells. The probes were cloned in the pCR4.0 backbone and confirmed by sequencing (Stahl Lab, unpublished data).

### *3.2 Experimental Myocardial Ischemia-Reperfusion (MI/R) Model*

Mice in a blinded fashion were chosen for either the sham treated 1C10 (n=5), sham treated mAb 3F8 (n=5) or full procedure group treated with 1C10 (n=5) or mAb 3F8 (n=5). Mice were induced with sodium pentobarbital (60-70 mg/kg) and a tracheal incision was made for direct tracheal exposure to assist in endotracheal intubation with a 20g Angiocath (BD Angiocath, Franklin Lakes, NJ). Mice were administered in a blinded fashion either 300µg of

mAb 3F8 or 300μg of 1C10 intravascular (IV) immediately after endotracheal intubation and just prior to performing sternotomy. Ear punch specimens were collected and stored for later genotyping confirmation. Positive pressure ventilation with a SAR-830 small animal ventilator (CWE Inc., Colorado Springs, CO) 29 was initiated. Anesthesia was maintained with isoflurane 0.2-0.5 monitored anesthesia care (MAC) and discontinued during the reperfusion period. Electrocardiograms (Astro-Med, Mentor, OH) were performed before, during, after induction of MI/R and at the conclusion of the reperfusion period. A median sternotomy was performed and sternal edges retracted with 5-0 nylon sutures (Ethicon, Summerville, NJ). The pericardium was divided and left pleural space violated to assure no residual pneumothorax. After direct visualization an 8-0 nylon (Ethicon) suture was placed around the left anterior descending coronary artery (LAD) 0.3mm inferior to the left atrial appendage. The suture in the full procedure group was tightened over a 0-0 piece of nylon (Deknatel, Mansfield, MA) to occlude the artery and induce ischemia for 45 minutes (Figure 7).

Mice that failed a first attempt at suture occlusion were excluded to avoid the possibility of preconditioning. The reperfusion period was 1 hour following MI/R and in the full procedure group after removal of LAD ligature. The heart was explanted and placed in a sterile dish with 8ml of RNAlater for the dissection of the right ventricle and septum. The resected left ventrical was placed in a 1.6ml Cryogenic Vial (Corning Inc., Corning, NY) tube with 1ml RNAlater and stored at -80$^{o}$C.

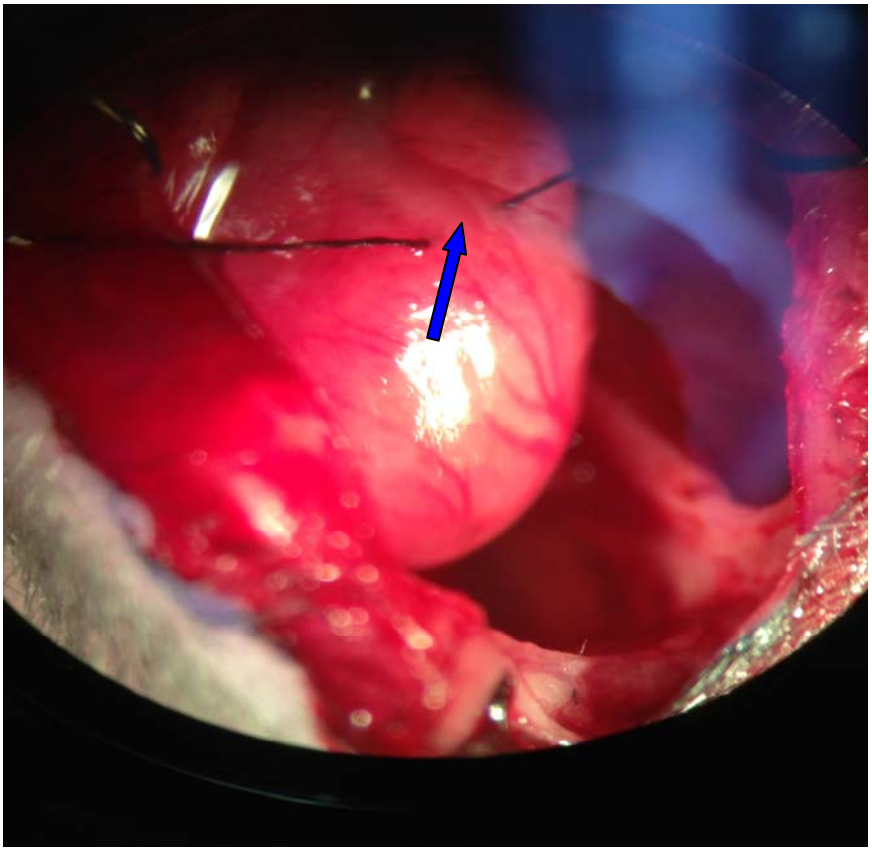

**Figure 7. Mouse Heart Coronary Artery Ligation Technique**. Mouse heart from a hMBL$^{+/+}$ mouse with suture placed under left anterior descending (LAD) coronary artery, indicated by blue arrow, prior to ligation for induction of ischemia (Gorsuch, WB unpublished data).

### *3.3 Measurement of Infarct Size and Area at Risk (AAR)*

Two groups of hMBL mice, control (n=2) and mAb 3F8 treated (n=2) in a blinded fashion were induced with sodium pentobarbital (60-70 mg/kg) and a tracheal incision was made for direct tracheal exposure to assist in endotracheal intubation with a 20g Angiocath (BD Angiocath, Franklin Lakes, NJ). Positive pressure ventilation with a SAR-830 small animal ventilator (CWE Inc., Colorado Springs, CO) 29 was initiated. Anesthesia was maintained with isoflurane 0.2-0.5 monitored anesthesia care (MAC) and discontinued during the reperfusion period. Electrocardiograms (Astro-Med, Mentor, OH) were performed before, during, after induction of MI/R and at the conclusion of the reperfusion period.

A median sternotomy was performed and sternal edges retracted with 5-0 nylon sutures (Ethicon, Summerville, NJ). The pericardium was divided and left pleural space violated to assure no residual pneumothorax. After direct visualization an 8-0 nylon (Ethicon) suture was placed around the left anterior descending coronary artery (LAD) 0.3mm inferior to the left atrial appendage. The suture in the full procedure group was tightened over a 0-0 piece of nylon (Deknatel, Mansfield, MA) to occlude the artery and induce ischemia for 45 minutes. Mice that failed a first attempt at suture occlusion were excluded to avoid the possibility of preconditioning. Just prior to removal of ligation suture 100μg of mAb 3F8 was administered to the treatment group intravenously (IV). A two inch 20g catheter was placed as a mediastinal chest tube. The sternum was re-approximated with 5-0 nylon sutures in figure eight pattern. The skin was sutured with 5-0 nylon and Veterinarian Dermabond was applied to prevent a pneumomediastinum while the chest tube was suctioned and removed. All mice underwent four hours of reperfusion to allow full development of infarcted myocardium.

Mice were re-anesthetized with sodium pentobarbital (60-70mg/kg). A median sternotomy was performed, and the LAD ligation suture tightened. The inferior vena cava was isolated through a right oblique abdominal incision, and 200μg of heparin injected. To allow exsanguinations, the vena cave was transected. All great vessels of the aortic arch were ligated excluding the ascending aorta. After transecting the descending aorta, a piece of polyethylene 10 tubing was inserted into the vessel lumen, and 100-200μl of 1% Evans Blue dye (Acros, Morris Plains, NJ) was injected for antegrade perfusion. Hearts were then excised and cross-sectioned into 1-mm slices using a coronal acrylic matrix (Roboz, Gaithersgurg, MD). Sections were placed into 6-well plates (Sigma Aldrich, St. Louis, MO) and incubated in 1% triphenyl tetratrazolium chloride (TTC) (Acros, Morris Plains, NJ) for 20 minutes. Heart slices were imaged and measured using a Nikon SMZ800 stereoscopic zoom microscope and SPOT™ imaging software (Diagnostic Instruments, Sterling Heights, MI). The size of the

infarction was determined by calculating the total areas of the left ventricular infarct, non-ischemic tissue, and AAR (Figure 8).

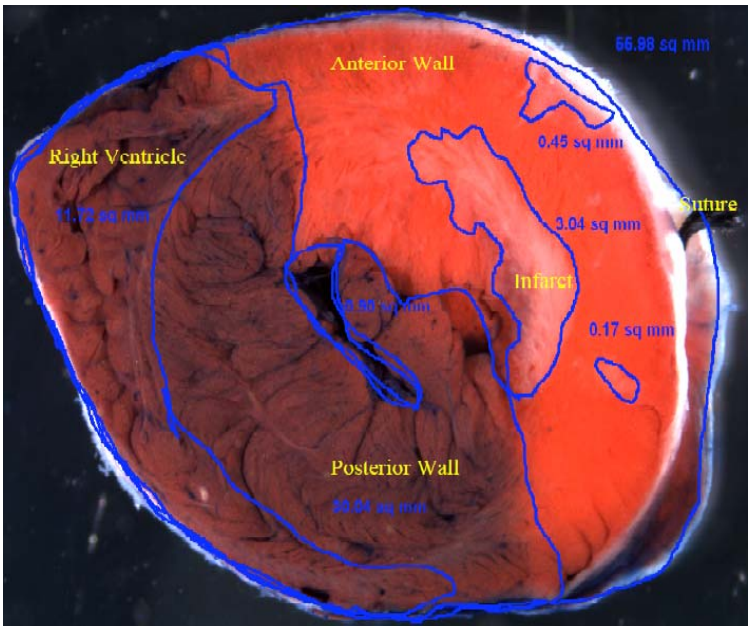


**Figure 8. Planimetry Measuring Technique**. Illustration of Area at risk (AAR), percent left ventricle (%LV) and percent infarct (%I) planimetry example. Bright red indicates viable myocardium which was not perfused with Evans Blue staining due to coronary ligation suture. White areas are infarct. Blue areas indicate non involved right ventricle. Adopted from Gorsuch et al., 2009.

Statistical analysis using Sigma Stat® software (Systat Software, San Jose, CA) was performed. All groups within the data analysis for AAR underwent normality and goodness of fit testing and displayed an accepted Shapiro-Wilk statistic with corresponding *p* values to reject the null hypothesis and accept group normality for measurements in AAR. One-way ANOVA was then performed and significance was accepted if $p \leq 0.05$. Infarct analysis on hearts having undergone MI/R utilized base to apex sections for planimetry measurements calculating infarct size percent to left ventricle area (%I/LV), AAR percent to left ventricle (%AAR/LV), and infarct size percent to AAR (%I/AAR) in Sham mice and mice treated with 1C10 and 3F8.

### *3.4 Real Time Polymerase Chain Reaction for Microarray Expression Analysis*

Genes selected for real time polymerase chain reaction (qPCR) were done so based on adjusted p values from the LIMMA Toptable results (adjusted p value < 0.05). Gene BC030870 was selected as the top most statistically significant gene between phenotypes 3F8 FULL and 1C10 FULL, which, was also a gene of interest belonging to the lncRNA family of genes. Gene Hspa1a, which, belongs to the heat shock family and was chosen between phenotypes 3F8 FULL and 3F8 SHAM to determine any expression changes within the same treatment phenotype but different operative models. The gene Xrl4a was chosen from the previous model comparison as it was an uncharacterized gene and further correlation was needed.

### *3.41 RNA Purification and Extraction*

The heart was explanted and placed in a sterile dish with 8ml of RNAlater (SIGMA-ALDRICH, St. Louis, MO) for the dissection of the right ventricle and septum. The resected left ventricle was placed in a 1.6ml Cryogenic Vial (Corning Inc., Corning, NY) tube with 1ml RNAlater and stored at -80$^{o}$C. All samples were placed on dry ice and shipped to GenUs Biosystems (Northbrook, IL) for RNA extraction, purification and construction of Agilent single channel microarray chips (Agilent Technologies, Santa Clara, CA).

### *3.42 RNA Quantity Measurement*

Individual RNA samples were measured for quantity by GenUs Biosystems (Northbrook, IL) prior to being processed on Agilent single channel microarray chips (Agilent Technologies, Santa Clara, CA).

### *3.43 Primers*

Primers for Agilent gene probe A_55_P2124582 (*Xlr4a*) and A_66_P126348 (BC030870) (Agilent Technologies, Santa Clara, CA) were built from Primer-Blast

(www.ncbi.nlm.nih.gov/tools/primer-blast/). Gene FASTA sequence was uploaded to the web based server. Specific criteria for input parameters required by BIO-RAD iQ SYBR Green QPCR (BIO-RAD, Hercules, CA) system required a product length between 72-200 base pairs, region to be amplified to have no or less than 4 long repeats of single bases, a GC content of 50-60% and no or minimum secondary structure. Massachusetts General Hospital (MGH) PrimerBank scientific web server (www.pga.mgh.harvard.edu/primerbank/) was then used to upload and order primers for delivery and use.

### *3.44 cDNA Library*

RNA samples were returned from GenUs Biosystems (Northbrook, IL). iScript Select cDNA Synthesis Kit (BIO-RAD, Hercules, CA) was utilized. Added in a 0.2ml RNAse free tube were nuclease free water, 5x iScript select reaction mix, Oligo $(dT)_{20}$ primer, RNA sample and iScript reverse transcriptase. Mixture ran on DNA Engine (MJ Research Inc.). Protocol consists of running samples for 90 minutes at 42$^{o}$C to activate transcriptase followed by incubation at 85$^{o}$C for 5 minutes to heat inactivate reverse transcriptase. Stored at -20$^{o}$C for future PCR amplification.

### *3.45 Quantitative Polymerase Chain Reaction*

The iQ SYBR Kit (BIO-RAD, Hercules, CA) was used for qPCR. Placed into ninety well plates were iQ SYBR Green Supermix (BIO-RAD, Hercules, CA), reverse primers, forward primers, sterile water and DNA template from previous step. Quantitative polymerase chain reaction was then run on BioRad iCycler (BIO-RAD, Hercules, CA). Data were then visualized and interpreted using CFX Manager Software (BIO-RAD, Hercules, CA) on a Macintosh Powerbook with processor OSX version 10.8.5, 2.7 GHz Intel Core i7 with 16GB 1600 MHz DDR3 memory (Apple Inc., Cupertino, CA).

### *3.5 Processing of Microarray Data and LIMMA*

Raw unprocessed data was received from GenUs Biosystems (Northbrook, IL) after being hybridized and read on Agilent single channel microarrays. The data were assessed and analyzed prior to statistical analysis through cross analysis using correspondence analysis and hierarchical cluster histograms. Data were structured and read into the Bioconductor Linear Models for Microarray Date (Limma) using R code script.[97] Using Limma, background signals were subtracted with the *normexp* method. Normalization between green channel arrays was achieved with quantile normalization. Log2 transformation of the green channel intensity values was performed. The Limma *RGList* from aforementioned steps was transformed into a *MAList* for further manipulation. Replicate spots were averaged and a 3 x 3 contrast design matrix was built for linear modeling functions. The Limma *lmFit* function was then used to estimate the variability in the data. The R code script developed for the aforementioned process is as follows:

```
>targets<-readTargets("targets.txt")
>x<read.maimages(targets,path="Y:/Brian_2012/dissertation/GenUsmicroarraydissert
atiodata/raw.txt data/raw_data", source="agilent", green.only=TRUE)
>y<-backgroundCorrect(x,method="normexp", offset=16)
>y<-normalizeBetweenArrays(y,method="quantile")
>y.ave$targets <-targets
>design<-model.matrix(~0 + targets$Merge)
>colnames(design)
>colnames(design) = c("1C10","3F8", "SHAM")
>fit<-lmFit(y.ave,design)
>fit$coefficients[1:2,]
>contrast.matrix<-makeContrasts(IC10-SHAM, 3F8-SHAM, 3F8-1C10,levels=design)
>levels(design)
>design[1:2,]
```

```
>contrast.matrix<-makeContrasts(1C10-SHAM, 3F8-SHAM, 3F8-1C10,levels=design)
>colnames(design) = c("Tr1C10","Tr3F8", "SHAM")
>colnames(design) = c("Tr1C10","Tr3F8", "SHAM")
>fit<-lmFit(y.ave,design)
>fit$coefficients[1:2,]
>contrast.matrix<-makeContrasts(Tr1C10-SHAM,Tr3F8-SHAM,Tr3F8
Tr1C10,levels=design)
>fit2<-contrasts.fit(fit, contrast.matrix)
>fit2<-eBayes(fit2)
>results<-decideTests(fit2)
>fit2[1:2,]
>fit2$coefficients[1:2,]
>topTable(fit2, coef=3, p.value=0.05)
>topTable(fit2, coef=1, p.value=0.05)
>topTable(fit2, coef=2, p.value=0.05)
```

Following data processing with Limma the Gene Set Enrichment Analysis (GSEA) program was downloaded from the Broad Institute website (www.broadinstitute.org/gsea/index.jsp). All Limma processed data was formatted into .gct data and .cls phenotype GSEA file structure requirements. Basic program settings we set as follows; false collapse, 1000 phenotype permutations, Agilent Mouse Genome chip platform, weighted enrichment statistic, sign2noise metric for ranking genes, real gene list sort mode, descending gene list order, maximum size 500 and minimal size 1. Advanced settings were set as; max_probe collapsing mode, meandiv normalization mode, no_balance randomization mode, omission of features with no symbol, mark detailed gene sets, non performance of median class metrics, 100 markers, 20 plot graphs, timestamp seed for permutation, no random values ranked and to create a zip file.

### *3.6 Molecular Modeling*

Molecular modeling will focus on understanding the molecular mechanism of mAb 3F8 recognition of target human MBL. The original research by Zhao et al. (2002) will be reexamined to interpret the results from molecular 3D modeling and the accuracy of such data in the setting of computational bioinformatic advancements, tools and databases. The main goal being to identify the 3D structure of the epitope or antigenic determinant and allow for such knowledge gained to assist in the design of small molecular inhibitors of human MBL in the future. The foundation of this molecular modeling will consist of homology modeling followed by three dimensional structure alignment. The aforementioned goals will be accomplished with the utilization of bioinformatic databases, modeling and visualization tools, 3D loop structure predictions, 3D structure comparisons and phylogeny of MBL from various species to identify functionally and structurally important residues. The conclusion will enable further post doctorate research in utilizing structure based drug design (SBDD) techniques for homology modeling and biological function prediction of those genes of highest interest.

#### *3.61 Homology Modeling*

Computational Thioredoxin loop generation with flitrix random peptide recognized by mAb 3F8 was performed. Thioredoxin from *E. coli*, as used by Zhao et al. (2002) during biopanning with flitrix random peptide display library (Invitrogen, Carlsbad, CA), was downloaded from PDB (www.rcsb.org/pdb/home/home.do) using PDB Blast similarity search. The resultant protein 2TRX was retrieved in FASTA sequence format. The PDB file was uploaded to molecular visualization desktop program Chimera version 1.8.1 (www.cgi.ucsf.edu/chimera) for visualization and identification of the Thioredoxin active loop.

The web based server SWISS-MODEL (www.swissmodel.expasy.org) was used to build an experimental three dimensional structure of Thioredoxin containing the 3F8 flitrix recognized peptide FGQHFDGLPTSA. The peptides between Cys32 and Cys35 of Thioredoxin were removed and inserted was the aforementioned flitrix peptide sequence. This

new Thioredoxin built flitrix active loop would serve as the target sequence and the original PDB Thioredxin file as the template for which homology three dimensional modeling would occur. Programs used by SWISS-MODEL which were used to assess the local quality of the predicted Thioredoxin replaced loop structure were Anolea (Laboratory of Molecular Bioinformatics Catolica, Chile) for assessing the packing of experimental models, GROMOS (Laboratory of Physical Chemistry, Zurich) for predicting structure quality and ProQRES (Stockholm Bioinformatics Center, Stockholm) for model accuracy. Procheck (European Bioinformatics Institute, Cambridge, UK) was used to perform stereo chemical assessment and protein structure verification. Ranking of alternative models was performed by QMEAN (Swiss Institute of Bioinformatics, Bael, Switzerland) and DFIRE (Howard Hughes Medical Institute, NY). Structural features were classified with DSSP and Promotif (University College, London).

Modloop (www.modbase.compbio.ucsf.edu/modloop/) online server was next used for further loop building and refinement. The new structure was uploaded to Rutgers University Sirius2 scientific server using open source program Cyberduck secure file transfer (http://cyberduck.io) version 4.3.3. SYBYL-X (Tripos, St. Louis, MO) was used to delete all peptides not in active loop between Cys32 and Cys45. This segment was termed flipeptide and used for further modeling.

### *3.62 Structure Alignment*

Three dimensional structure comparison was used to identify the mAb 3F8 epitope for hMBL. Interacting segments were built and compared. The flipeptide segment and hMBL PDB files (1HUP) were loaded into the MArkovian Transition of Structure evolution (MATRAS) protein 3D structure comparison online scientific server (www.stromp.protein.osaka-u.ac.jp/matras/). Parameters included full atom PDB files and sequence alignment with BLOSUM62 turned off. A second program was used, RCSB PDB Protein Comparison Tool (www.pdb.org/pdb/workbench/workbench.do?action=menu) to further assess 3D alignment

between flipeptide and hMBL. The RCSB PDB tool is based on the alignment algorithms of FATCAT and Combinational Extensions (CE). Parameters chosen included FatCat rigid and flexible alignment algorithm, fragment length of 4, RMSD cutoff of 3.0, AFP distance cutoff of 5.0 and maximum number of twists 0. Combinational Extension parameters included maximum gap of 10, fragment length of 5, CA distance and angle between side chains scoring function, RMSD of 10, open gap of 5 and a gap extension of 0.5. The RCSB PDB protein comparison tool also allows for structural alignment using local sequence alignment with the Smith Waterman algorithm.

## CHAPTER IV

## RESULTS

### *4.1 Mice Genotyping*

Confirmation of genotype was performed with Licor Odyssey Imager using Odyssey CLx manager software (Figure 9).

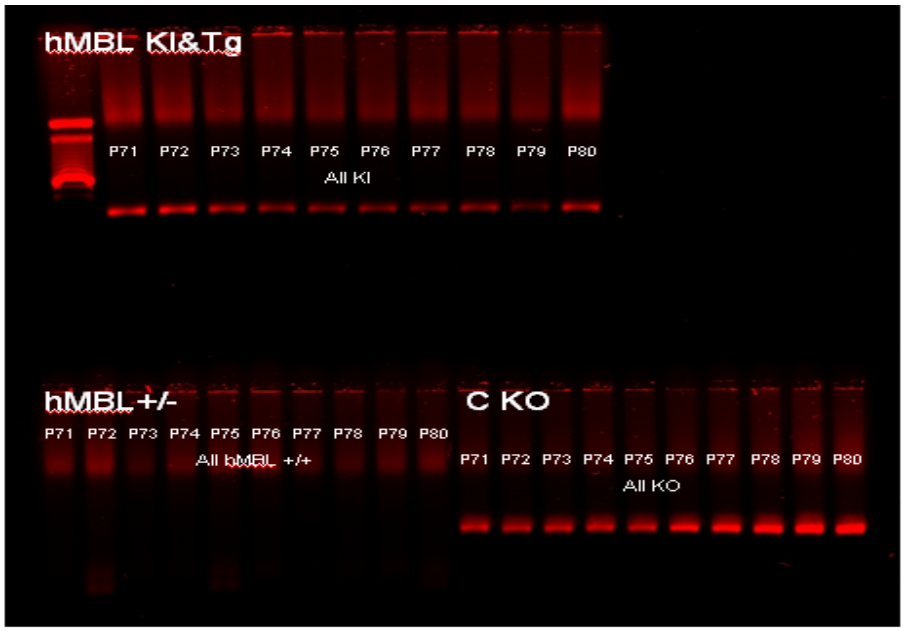


**Figure 9**. **hMBL Genotyping And C3 Expression Assay Results**. Genotpying performed on hMBL mice (a). Upper left confirming human mannose binding lectin (hMBL) gene being knocked into mouse. Lower right confirming same mice having mouse MBL knocked out. Left lower summarizing and confirmation of all mice being hMBL $^{+/+}$ (Stahl Lab unpublished data 2014).

### *4.2 Measurement of Infarct Size and Area at Risk (AAR)*

Mice in all groups were observed to have no significant differences in area at risk to percentage of left ventricle (Figure 10). Untreated mice had significantly greater percent infarct to area at risk (%I/AAR)(Figure 11) and infarct area to left ventricle (Figure 12). Planimetry was performed to visualize and measure the aforementioned statistics (Figure 13).

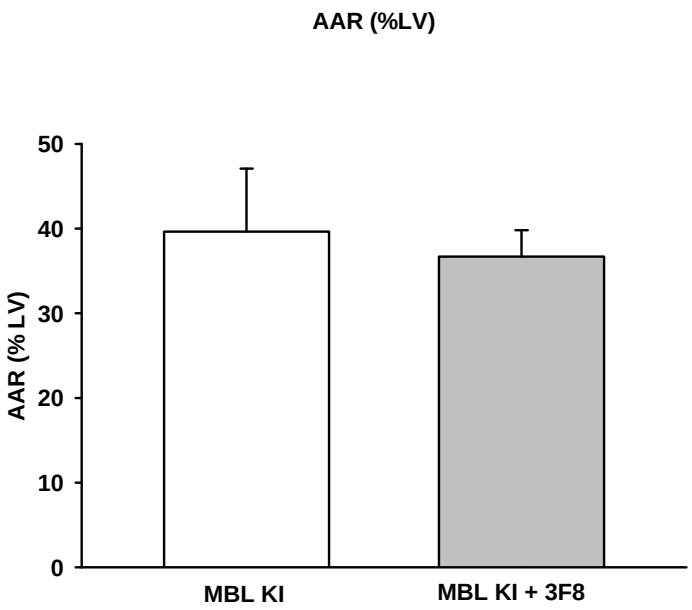


**Figure 10. Area At Risk As Percentage of Left Ventricle In hMBL KI Versus hMBL KI + 3F8 Mice**. Calculated percent area at risk to left ventricle area (%AAR/LV) for untreated hMBL knock in (KI) (N=2) mice and mice treated with 3F8 (N=2) undergoing myocardial ischemia-reperfusion injury (MI/R) measured by planimetry. Data represent mean +/- SD.

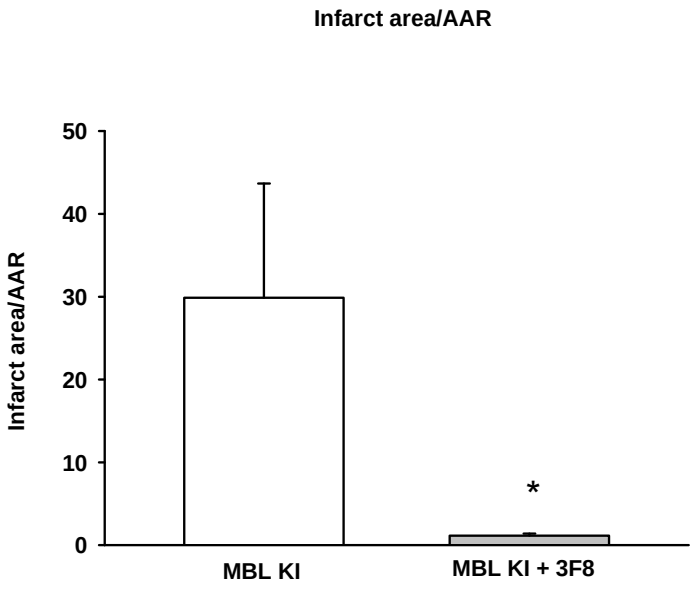


**Figure 11. Infarct Area As Percentage of Area At Risk In hMBL KI Versus hMBL KI + 3F8 Mice**. Calculated percent infarct to area at risk (I/AAR) for untreated hMBL knock in (KI) (N=2) and mice treated with 3F8 (N=2) undergoing myocardial ischemia-reperfusion injury measured by planimetry. Data represent mean +/- SE. * Significantly different from sham MBL KI, $p \leq 0.05$.

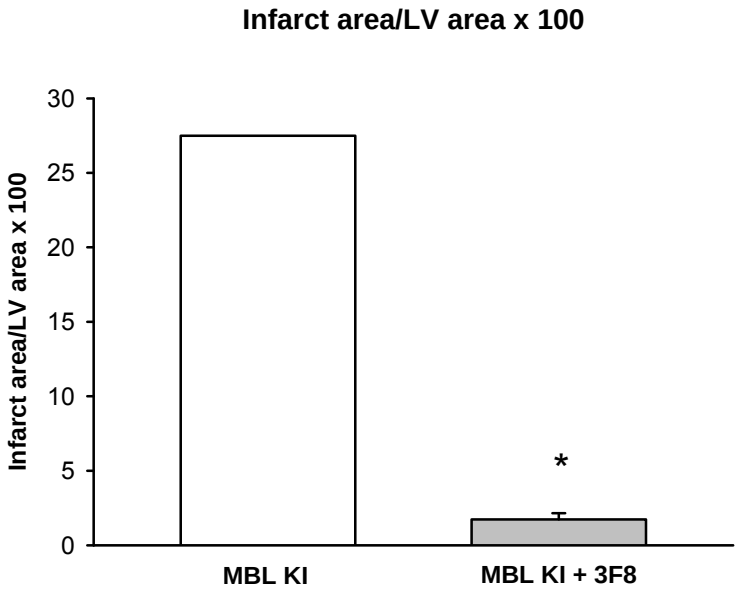


**Figure 12. Infarct Area As Percentage of Left Ventricle In hMBL KI Versus hMBL + 3F8 Mice**. Calculated infarct area to left ventricle area as percent (IA/LV) for untreated hMBL knock in (KI) (N=2) and mice treated with 3F8 (N=2) undergoing myocardial ischemia-reperfusion injury measured by planimetry. Data represent mean +/- SE. * Significantly different from untreated hMBL KI, $p \leq 0.05$.

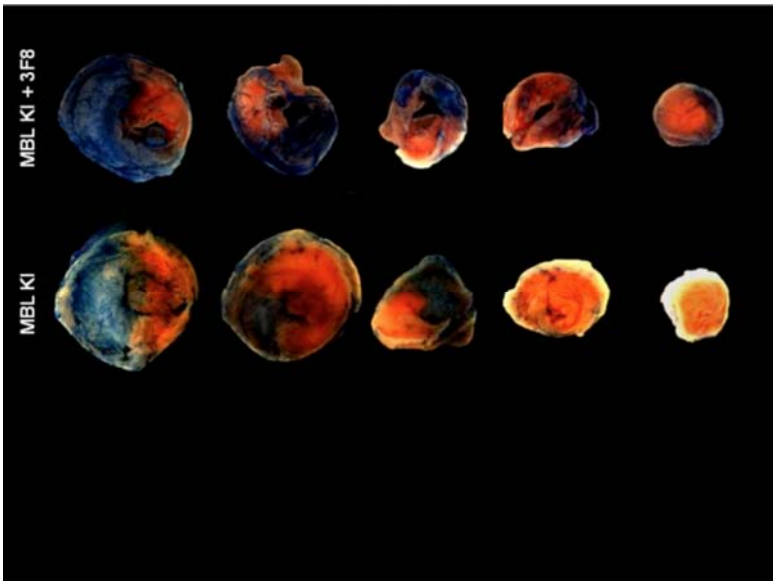


**Figure 13. Myocardial Staining**. Planimetric assessment of area at risk (AAR) and infarct areas (IA) after myocardial ischemia-reperfusion injury (MI/R) in human mannose binding lectin (hMBL) knock in (KI) mice compared to mice treated with monoclonal antibody (mAb) 3F8 undergoing MI/R. Bright red section indicates AAR that was not perfused with Evans Blue staining, secondarily to coronary ligation suture. White denotes infarct zone. Seen are sections of equal location in comparison between groups from apex to base.

### *4.3 Microarray Gene Expression Correspondence Analysis*

Raw unprocessed gene probe data underwent correspondence analysis which revealed significant variance in models 3F8 FULL:2 and 1C10 FULL:1 as shown in two dimensional graphical representation (Figure 14). Hierarchal tree reveals FULL-1C10:1 and FULL-3F8:2 furthest distance from respected similar phenotype pairs (Figure 15). All probe signal intensity values were normalized (Figure 16a-b). Three levels of treatment$Condition factor analysis were performed; 1C10 Full MI/R (n=3) vs. 1C10 SHAM MI/R (n=3), 3F8 Full MI/R (n=3) vs. 3F8 SHAM MI/R (n=3) and 3F8 Full MI/R (n=3) vs. 1C10 Full MI/R (n=3). The false discovery rate (FDR) of < 0.05 was selected within the Limma program. The 1C10 Full MI/R compared to the 1C10 SHAM group had significant gene expression upregulation of seventeen genes and downregulation of 3 genes. Upregulated genes consisted of Fos, Hspala1, Jun, Fosb, Fam7la, Ckmt1, Ler2, Egr1, Ler5, Rgs4, Mycn, Junb, Rnu3b1, Malat1, Mrg1 and 2 uncharacterized genes from Agilent mouse genome probe identifiers; A_30_P01031072 and A_55_P2021187. Downregulated gene expression was seen in *Zerb1* and two uncharacterized genes from Agilent mouse genome probe identifiers A_55_P2404554 and A_55_P2009994 (Table 1).

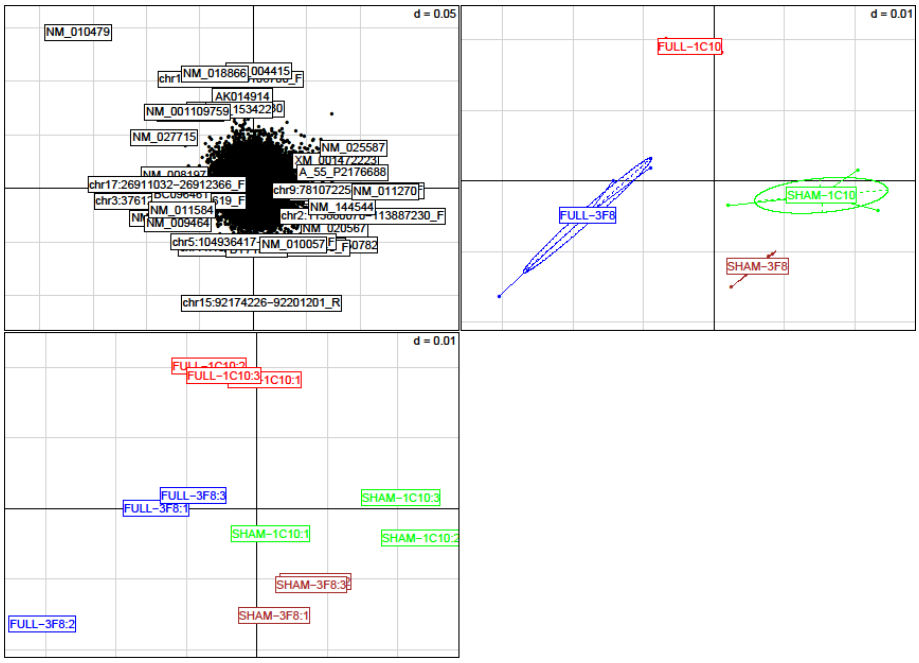


**Figure 14. Correspondence Graph of Gene Expression Data Analysis**. Two dimensional graphical representation of correspondence analysis performed after raw Agilent mouse genome probe identifier data underwent background correction, quantile normalization and averaging of replicate spots. Phenotype groupings include FULL 3F8 in blue, FULL 1C10 in red, SHAM 3F8 in purple and SHAM 1C10 in green. Upper left centroid distribution. Upper right and lower left clustering 2D distance graphical representation.

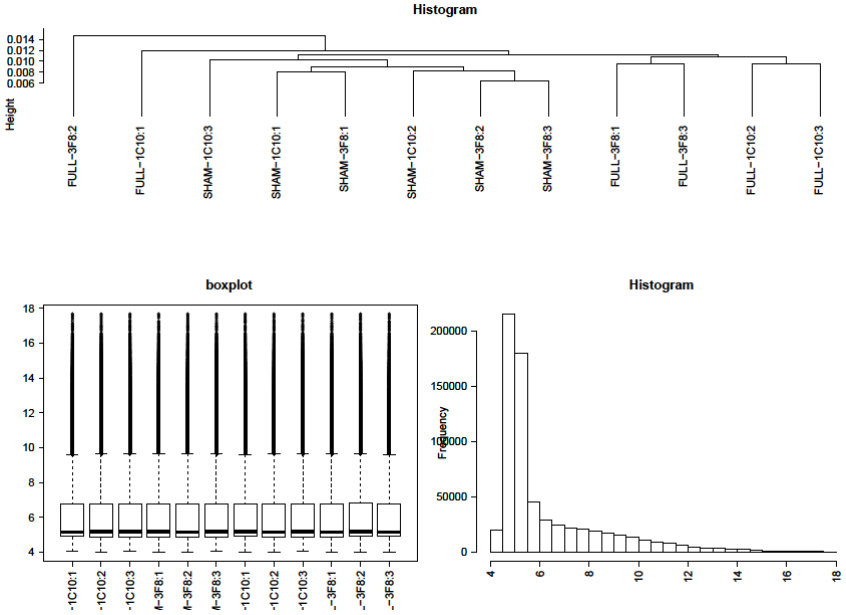


**Figure 15. Histogram and Hierarchical tree of Gene Expression Data Analysis**.

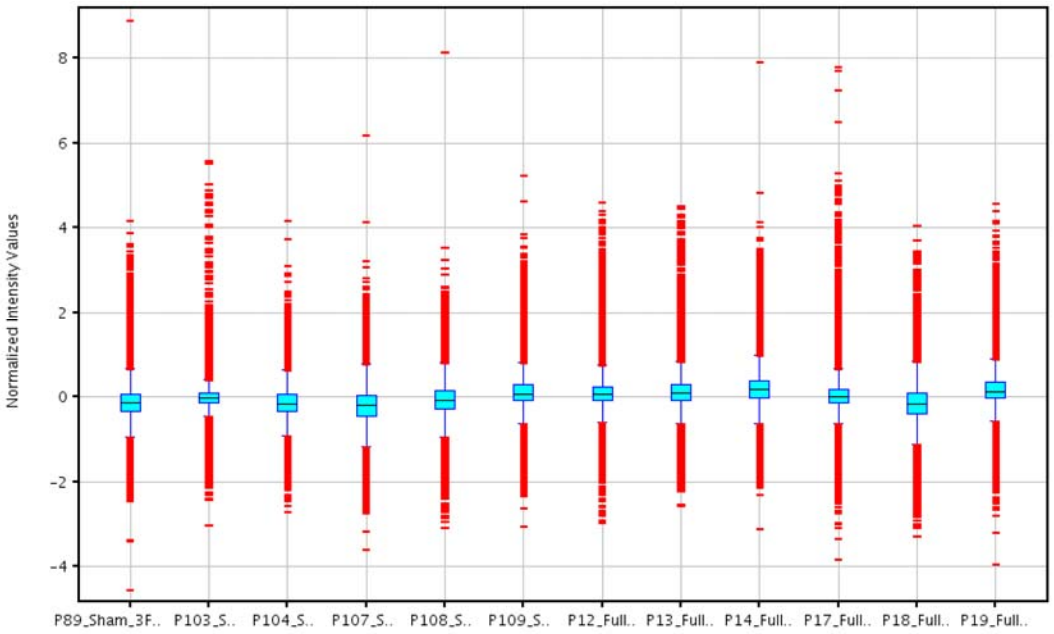


**Figure 16a. Post Normalization Graphical Representation of Gene Probe Intensity Values**. Graphical representation of Agilent mouse genome probe intensity signal values after normalization. Boxplot analysis of all categorical phenotypes. Normalized intensity values Y axis and individual animal identification number on X axis.

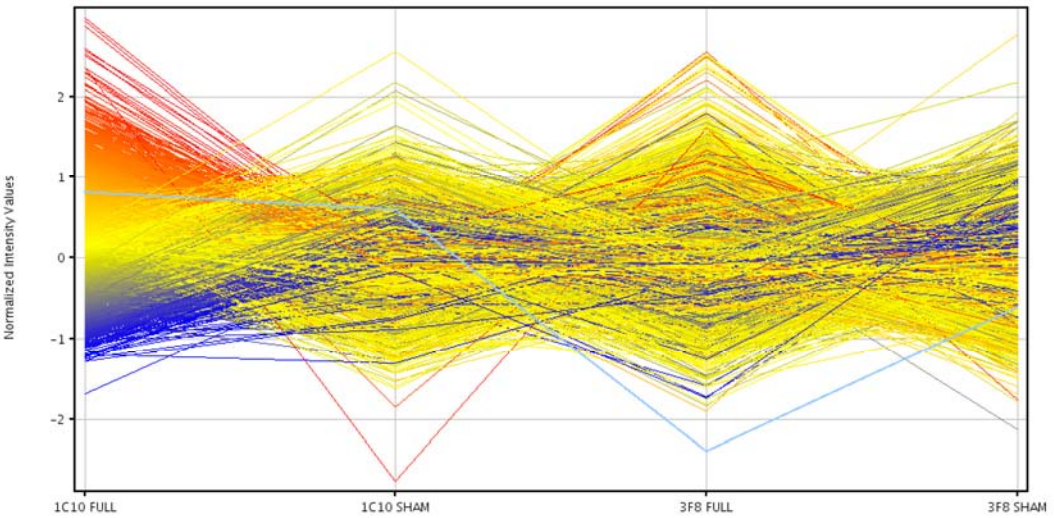


**Figure 16b. Post Normalization Graphical Representation of Gene Probe Intensity Values**. Normalized intensity color matrix showing blue line which represents most significantly downregulated gene *BC030870* in the 3F8 FULL phenotype group as resulted from linear models for microarray analysis (Limma). Model phenotype description X axis and normalized intensity values Y axis.

**Table 1. 1C10 FULL vs. 1C10 SHAM Mice Gene Expression Results**. Green and white represents upregulation and downregulation of genes in 1C10 FULL mice compared to 1C10 SHAM mice undergoing myocardial ischemia-reperfusion injury. Name column indicating gene set description. Size column indicating how many individual gene sets in particular database set. Results indicated by enrichment score (ES), normalized enrichment score (NES), nominal p value (NOM p-val) and false discovery rate for p value (FDR q-val). Significance calculated for adjusted p < 0.05.

| ProbeName | Gene | Gene Function | logFC | t | P.Value | adj.P.Val |
|---|---|---|---|---|---|---|
| A_52_P262219 | Fos | FBJ Murine Osteosarcoma Viral | 2.32733647 | 15.7477073 | 2.65E-10 | 1.48E-05 |
| A_55_P2068459 | Hspa1a | Heat Shock | 4.7796615 | 12.2168917 | 7.38E-09 | 0.00020602 |
| A_55_P2158990 | Jun | Proto-Oncogene | 1.74241846 | 11.6354312 | 1.38E-08 | 0.00025675 |
| A_55_P2404554 | | uncharacterized | -1.59268272 | -9.83617287 | 1.14E-07 | 0.00151103 |
| A_55_P2113051 | Fosb | FBJ Murine Osteosarcoma Viral Oncogene | 1.420649 | 9.70013829 | 1.35E-07 | 0.00151103 |
| A_55_P2147956 | Fam71a | Family With Sequence Similarity 71, Memb | 1.3813068 | 8.75008243 | 4.73E-07 | 0.00440258 |
| A_55_P2072656 | Ckmt1 | Creatine Kinase, Mitochondrial 1B1 2 | 1.18891031 | 8.59373394 | 5.87E-07 | 0.00468035 |
| A_55_P2163028 | Ier2 | immediate Early Response | 1.38525536 | 7.5021737 | 2.86E-06 | 0.01995137 |
| A_66_P114501 | Zcrb1 | Zinc Finger CCHC- | -1.009138 | -7.05769196 | 5.69E-06 | 0.03526557 |
| A_30_P01031072 | | uncharacterized | 1.20119845 | 6.94429692 | 6.80E-06 | 0.0377095 |
| A_51_P367866 | Egr1 | Early Growth Response | 1.13484866 | 6.8890067 | 7.43E-06 | 0.0377095 |
| A_55_P2021109 | Ier5 | Immediate Early Response | 1.41033315 | 6.78005727 | 8.85E-06 | 0.03898333 |
| A_55_P2009217 | Rnu3b1 | nuclear RNA | 1.55569506 | 6.73815701 | 9.47E-06 | 0.03898333 |
| A_30_P01020325 | Malat1 | met.assoc.lung adeno transcript 1 | 1.01003927 | 6.70008142 | 1.01E-05 | 0.03898333 |
| A_55_P2009994 | | uncharacterized | -0.99300826 | -6.67605425 | 1.05E-05 | 0.03898333 |
| A_55_P2026738 | Rgs4 | Regulator Of G-Protein Signalling | 1.24753129 | 6.50445708 | 1.39E-05 | 0.04841314 |
| A_55_P2021187 | | uncharacterized | 1.14852752 | 6.4508382 | 1.52E-05 | 0.04979088 |
| A_52_P536494 | Mycn | v-myc avian myelocytomatosis viral relatec | 1.24439268 | 6.37019921 | 1.73E-05 | 0.05287274 |
| A_55_P2011106 | Junb | Transcription Factor Jun-B | 1.22305133 | 6.34795141 | 1.80E-05 | 0.05287274 |
| A_55_P2116621 | Mrg1 | inhibitor of HIF1A induced genes | 0.86453464 | 6.28098881 | 2.01E-05 | 0.05619655 |

Mice in the 3F8 Full MI/R when compared to 3F8 SHAM group had a significant change in gene expression upregulation in eight genes and dowregulation of fourteen genes. Upregulated gene expression changes in the 3F8 Full MI/R group were *Otud1, Atf3, Hspala, Cyp2ab1, Btg1* and two uncharacterized genes from Agilent mouse gene probe identifiers; A_30_P01031719 and A_30_P01024899. Significant downregulation of gene expression in the 3F8 Full MI/R group was seen in *Xrl4a, Nrp1, Fbxw7, Zcchc11, Pcsk6, Lrrc14b* and seven uncharacterized genes from Agilent mouse genome probe identifiers; A_30_P01029859, A_30_P01029543, A_30_P01023257, A_55_P2096165, A_66_P106983, A_55_P2009994 and A_30_P01021666 (Table 2).

**Table 2. 3F8 FULL vs. 3F8 SHAM Mice Gene Expression Results**. Green and white represents upregulation and downregulation of genes in 3F8 FULL mice compared to 3F8 SHAM mice undergoing myocardial ischemia-reperfusion injury. Name column indicating gene set description. Size column indicating how many individual gene sets in particular database set. Results indicated by enrichment score (ES), normalized enrichment score (NES), nominal p value (NOM p-val) and false discovery rate for p value (FDR q-val). Significance calculated for adjusted p < 0.05.

| ProbeName | Gene | Gene Function | logFC | AveExpr | t | adj. P. val. |
|---|---|---|---|---|---|---|
| A_30_P01029859 | | uncharacterized | -1.522849 | 9.177356 | -10.79855 | 1.13E-07 |
| A_55_P2124582 | Xlr4a | uncharacterized | -1.588371 | 6.521829 | -9.668365 | 3.90E-07 |
| A_55_P2220937 | Otud1 | Deubiquitinating enzymes ( | 1.570303 | 8.253829 | 9.442381 | 5.07E-07 |
| A_52_P452689 | Atf3 | Cyclic AMP-Dependent Transcription Factor | 1.715114 | 10.64971 | 9.003451 | 8.53E-07 |
| A_30_P01029543 | | uncharacterized | -2.100547 | 7.867409 | -8.826045 | 1.06E-06 |
| A_30_P01023257 | | uncharacterized | -1.410527 | 9.973175 | -8.646087 | 1.32E-06 |
| A_55_P2124586 | Xlr4c | uncharacterized | -1.182609 | 5.672845 | -8.525022 | 1.54E-06 |
| A_55_P2046101 | Xlr4a | uncharacterized | -1.819503 | 6.488816 | -8.408188 | 1.78E-06 |
| A_55_P2068459 | Hspa1a | Heat Shock- Protein | 3.196788 | 9.148873 | 8.048635 | 2.84E-06 |
| A_55_P2021109 | Ier5 | Immediate Early Response Gene 5 Protein | 1.329004 | 9.634909 | 7.878182 | 3.55E-06 |
| A_51_P469285 | Nrp1 | VEGF 165 Recepto | -1.2991 | 7.70324 | -7.561299 | 5.45E-06 |
| A_30_P01031719 | | uncharacterized | 0.964883 | 5.347865 | 7.548025 | 5.55E-06 |
| A_51_P506683 | Fbxw7 | Ubiquitin Protein Ligase | -1.062578 | 8.063436 | -7.514147 | 5.81E-06 |
| A_51_P501844 | Cyp26b1 | Cytochrome P450 | 1.582436 | 6.790421 | 7.255618 | 8.33E-06 |
| A_55_P2128646 | Gmnn | Geminin, DNA Replication Inhibitor | -1.240759 | 7.762792 | -7.24747 | 8.42E-06 |
| A_52_P31543 | Btg2 | BTG/Tob family. | 1.263798 | 11.24992 | 7.146272 | 9.72E-06 |
| A_55_P2096165 | | uncharacterized | -1.073794 | 6.896048 | -7.047428 | 1.12E-05 |
| A_55_P2131766 | D4Wsu53e | | -1.136949 | 10.39996 | -6.920899 | 1.34E-05 |
| A_55_P2032695 | Zcchc11 | zinc finger | -1.17499 | 5.726346 | -6.851088 | 1.49E-05 |
| A_30_P01024899 | | uncharacterized | 1.020743 | 5.471744 | 6.644862 | 2.01E-05 |
| A_55_P1972991 | Pcsk6 | Subtilisin-Like Proprotein Convertase | -0.930064 | 8.093849 | -6.614094 | 2.11E-05 |
| A_66_P106983 | | uncharacterized | -1.176988 | 9.62836 | -6.598332 | 2.16E-05 |
| A_55_P2009994 | | uncharacterized | -1.01393 | 6.851189 | -6.523856 | 2.41E-05 |
| A_30_P01021666 | | uncharacterized | -1.0062 | 8.826594 | -6.478308 | 2.58E-05 |
| A_55_P2052824 | Lrrc14b | Leucine-Rich Repeat-Containing Protein | -0.926676 | 9.25111 | -6.466211 | 2.63E-05 |

Mice in the 3F8 Full MI/R compared to 1C10 Full MI/R group had five significant genes down- regulated and none significantly upregulated in regards to the adjusted p value ≤ 0.05. There were 5,000 gene probes out of 69,870 showing expression significance in regards to nominal p ≤ 0.05. Down- regulated genes with adjusted p value significance were *Isoc2a, Mpst*, *BC030870*, *LOC619975*, and one uncharacterized gene from Agilent mouse genome probe identifiers; A_55_P2057519 (Table 3).

**Table 3. 3F8 FULL vs. 1C10 FULL Mice Gene Expression Results**. Green and white represents upregulation and downregulation of genes in 3F8 FULL mice compared to 1C10 FULL mice undergoing myocardial ischemia-reperfusion injury. Name column indicating gene set description. Size column indicating how many individual gene sets in particular database set. Results indicated by enrichment score (ES), normalized enrichment score (NES), nominal p value (NOM p-val) and false discovery rate for p value (FDR q-val). Significance calculated for adjusted $p < 0.05$.

| ProbeName | GeneName | logFC | t | P.Value | adj.P.Val |
|---|---|---|---|---|---|
| A_66_P126348 | BC030870 | -3.105803 | -12.42482 | 2.92E-10 | 1.63E-05 |
| A_55_P2145531 | Mpst | -1.062593 | -7.654114 | 4.61E-07 | 0.012857 |
| A_52_P427759 | LOC619975 | -1.039707 | -6.723259 | 2.67E-06 | 0.030518 |
| A_55_P2057519 | uncharacterized | -1.107658 | -6.719822 | 2.69E-06 | 0.030518 |
| A_30_P01027439 | Isoc2a | -1.067711 | -6.710875 | 2.73E-06 | 0.030518 |

### *4.4 Gene Set Enrichment Analysis*

The 3F8 FULL versus 1C10 FULL phenotype top 100 differentially expressed genes with raw unprocessed expression values from microarray with $p \leq 0.05$ from Limma toptable were loaded into desktop server gene set enrichment analysis (GSEA). Broad Institute molecular signatures gene set databases (MsigDB) used for enrichment analysis consisted of the biological process gene sets, molecular function gene sets and immunological signature gene sets. During enrichment 815 out of the databases 825 gene sets were filtered out. Reported significance of enrichment in the biological process gene set database was zero gene sets with FDR ≤ 0.25 and zero gene sets with nominal p value ≤ 0.05. The molecular function gene sets database consisted of 396 gene sets of which 394 were filtered out and zero sets had a significant FDR ≤ 0.25 and zero gene sets with nominal p value ≤ 0.05. Seventy eight gene sets were upregulated in the MsigDB immunological signature gene sets dataset with no gene sets significantly enriched (Table 4). Gene ranking for upregulation and downregulation performed by GSEA (Table 5).

The 3F8 FULL versus 1C10 FULL phenotype differentially expressed genes with values from limma with $p \leq 0.05$ (n=5000) were uploaded to GSEA preranked mode maintaining their ranking order based on p value. Broad Institute MSigDB biological process gene sets database enrichment (n=825) resulted in the filtering out of 250 gene sets. There

**Table 4. GSEA Limma Toptable Results for Phenotype 3F8 FULL vs. 1C10 FULL**. Gene Set Enrichment Analysis (GSEA) between phenotypes 3F8 FULL (n=3) and 1C10 FULL (n=3). Biological process gene sets (a), molecular function gene sets (b) and immunological signatures gene sets (c). Name column indicating gene set description. Size column indicating how many individual gene sets in particular database set. Results indicated by enrichment score (ES), normalized enrichment score (NES), nominal p value (NOM p-val) and false discovery rate for p value (FDR q-val).

| NAME | SIZE | ES | NES | NOM p-val | FDR q-val |
|---|---|---|---|---|---|
| PROTEIN_MODIFICATION_BY_SMALL_PROTEIN_CONJUGATION | 1 | 0.90816325 | 1.0558443 | 0.49705306 | 1 |
| PROTEIN_UBIQUITINATION | 1 | 0.90816325 | 1.0558443 | 0.49705306 | 1 |
| UBIQUITIN_CYCLE | 1 | 0.90816325 | 1.0558443 | 0.49705306 | 1 |
| BIOPOLYMER_METABOLIC_PROCESS | 2 | 0.814433 | 1.000896 | 0.41060904 | 1 |
| POST_TRANSLATIONAL_PROTEIN_MODIFICATION | 2 | 0.814433 | 1.000896 | 0.41060904 | 1 |
| CELLULAR_PROTEIN_METABOLIC_PROCESS | 2 | 0.814433 | 1.000896 | 0.41060904 | 0.94793755 |
| BIOPOLYMER_MODIFICATION | 2 | 0.814433 | 1.000896 | 0.41060904 | 0.81251794 |
| PROTEIN_METABOLIC_PROCESS | 2 | 0.814433 | 1.000896 | 0.41060904 | 0.7109532 |
| PROTEIN_MODIFICATION_PROCESS | 2 | 0.814433 | 1.000896 | 0.41060904 | 0.6319584 |

(a)

| NAME | SIZE | ES | NES | NOM p-val | FDR q-val |
|---|---|---|---|---|---|
| SMALL_PROTEIN_CONJUGATING_ENZYME_ACTIVITY | 1 | 0.908163 | 1.06216 | 0.481781 | 1 |
| ENZYME_BINDING | 1 | 0.908163 | 1.06216 | 0.481781 | 0.611336 |

(b)

| NAME | SIZE | ES | NES | NOM p-val | FDR q-val |
|---|---|---|---|---|---|
| GSE10325_BCELL_VS_LUPUS_BCELL_DN | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE11864_UNTREATED_VS_CSF1_IFNG_IN_MAC_DN | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE11864_CSF1_VS_CSF1_IFNG_PAM3CYS_IN_MAC_DN | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE11864_CSF1_IFNG_VS_CSF1_PAM3CYS_IN_MAC_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE11924_TH2_VS_TH17_CD4_TCELL_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE14308_TH17_VS_NAIVE_CD4_TCELL_DN | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE1432_CTRL_VS_IFNG_6H_MICROGLIA_DN | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE1432_CTRL_VS_IFNG_24H_MICROGLIA_DN | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE1432_1H_VS_6H_IFNG_MICROGLIA_DN | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE1432_1H_VS_24H_IFNG_MICROGLIA_DN | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE1460_CD4_THYMOCYTE_VS_NAIVE_CD4_TCELL_CORD_BLOOD_U | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_CTRL_VS_LPS_6H_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_CTRL_VS_LPS_8H_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_CTRL_VS_POLYIC_1H_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_CTRL_VS_PAM3CSK4_8H_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_CTRL_VS_PAM3CSK4_12H_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_LPS_VS_POLYIC_8H_BMDM_DN | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_POLYIC_VS_PAM3CSK4_2H_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_POLYIC_VS_PAM3CSK4_4H_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_POLYIC_VS_PAM3CSK4_8H_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_POLYIC_VS_PAM3CSK4_24H_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_PAM3CSK4_VS_GADIQUIMOD_24H_BMDM_DN | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_0.5H_VS_8H_LPS_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |
| GSE17721_0.5H_VS_24H_LPS_BMDM_UP | 1 | 0.908163 | 1.067456 | 0.464066 | 1 |

(c)

**Table 5. GSEA Limma Toptable Ranking**. Top 25 Gene Set Enrichment Analysis (GSEA) ranking for upregulated genes (a), and top 25 GSEA ranking for downregulated genes (b) using Linear Models for Microarray Data toptable 100 genes, significance p < 0.05. Name column indicating Agilent probe identification number followed by GSEA ranking score.

| NAME | GENE_SYI | SCORE |
|---|---|---|
| A_30_P01028713 | null | 0.426708 |
| A_55_P2325758 | null | 0.373755 |
| A_30_P01021060 | null | 0.372187 |
| A_30_P01031719 | null | 0.362772 |
| A_55_P2102941 | null | 0.361503 |
| A_55_P2408478 | null | 0.356711 |
| A_55_P2322961 | null | 0.346592 |
| A_30_P01029428 | null | 0.335777 |
| A_55_P1996479 | null | 0.315402 |
| ATG3 | ATG3 | 0.310582 |
| A_30_P01018801 | null | 0.302255 |
| A_55_P2040485 | null | 0.29631 |
| A_30_P01017602 | null | 0.293402 |
| A_55_P2032388 | null | 0.286719 |
| A_55_P1962982 | null | 0.282018 |
| A_55_P2041320 | null | 0.253267 |
| AKR1B3 | null | 0.249434 |
| A_55_P2130178 | null | 0.243877 |
| A_55_P2239317 | null | 0.225295 |
| NMT1 | NMT1 | 0.225213 |
| A_55_P2046943 | null | 0.216448 |
| 1810045K07RIK | null | 0.214242 |
| A_66_P127484 | null | 0.210339 |
| 5730438N18RIK | null | 0.200752 |

(a)

| | | |
|---|---|---|
| A_55_P2299524 | null | -0.320562 |
| A_55_P2064906 | null | -0.321845 |
| A_55_P2069530 | null | -0.322519 |
| A_30_P01029238 | null | -0.323393 |
| A_30_P01024517 | null | -0.325805 |
| A_55_P2075894 | null | -0.329898 |
| A_55_P2146483 | null | -0.33038 |
| A_55_P2074864 | null | -0.331103 |
| 0610012H03RIK | null | -0.336475 |
| A_55_P2114994 | null | -0.336847 |
| A_30_P01025509 | null | -0.341158 |
| A_30_P01025256 | null | -0.354233 |
| A_55_P2082625 | null | -0.372696 |
| A_55_P2076134 | null | -0.379731 |
| BC048507 | null | -0.386304 |
| A_30_P01020024 | null | -0.38794 |
| A_30_P01019682 | null | -0.39027 |
| A_30_P01033218 | null | -0.391386 |
| A_52_P234910 | null | -0.400728 |
| A_55_P2057519 | null | -0.403222 |
| A_55_P2252426 | null | -0.403907 |
| A_30_P01029126 | null | -0.406705 |
| A_30_P01020022 | null | -0.4094 |
| A_55_P1980721 | null | -0.442038 |
| A_30_P01024143 | null | -0.585226 |
| A_66_P126348 | null | -0.790902 |

(b)

were 73 gene sets with significant FDR ≤ 0.05, 27 gene sets with nominal p ≤ 0.01 and 28 with nominal p ≤ 0.05 (Table 6). Leading edge analysis was performed on downregulated gene sets with FDR ≤0.05 indicating top genes contributing most to enrichment score (Figure 17). Broad institute MSigDB molecular function gene set database enrichment (n=396) resulted in filtering out of 111 gene sets. There were zero sets with FDR ≤ 0.05, 2 gene sets with nominal p ≤ 0.01 and 8 gene sets with nominal p ≤ 0.05 (Table 7). Broad Institute MSigDB immunological signatures gene set database enrichment (n=1910) resulted in zero gene sets being filtered. There were 428 gene sets with FDR ≤ 0.05, 136 gene sets with nominal p ≤ 0.01 and 372 with nominal p ≤ 0.05 (Table 8). Leading edge analysis performed on gene set enrichment analysis (GSEA) significant upregulated gene sets from MsigDB immunological

signatures gene sets database to show genes contributing the most to enrichment of leading edge sets for FDR ≤0.05 (Figure 18).

**Table 6. GSEA Prerank Mode Biological Process Gene Set Database Results**. Upregulation of gene sets for phenotype 3F8 FULL when compared to 1C10 FULL when analyzed by Gene Set Enrichment Analysis (GSEA) prerank mode using Broad Institute molecular signatures database (MsigDB) biological process gene sets (a). Downregulation of gene sets for phenotype 3F8 FULL when compared to 1C10 FULL. Name column indicating gene set description. Size column indicating how many individual gene sets in particular database set. Results indicated by enrichment score (ES), normalized enrichment score (NES), nominal p value (NOM p-val) and false discovery rate for p value (FDR q-val). Significance measured by false discovery rate ≤ 0.05 (b).

| NAME | SIZE | ES | NES | NOM p-val | FDR q-val |
|---|---|---|---|---|---|
| AXON_GUIDANCE | 2 | -0.766153 | -1.344894 | 0.109312 | 1 |
| PATTERN_SPECIFICATION_PROCESS | 1 | -0.948283 | -1.270197 | 0.082474 | 1 |
| REGIONALIZATION | 1 | -0.948283 | -1.26837 | 0.121272 | 1 |
| EMBRYONIC_MORPHOGENESIS | 1 | -0.948283 | -1.255556 | 0.125725 | 1 |
| EMBRYONIC_DEVELOPMENT | 1 | -0.948283 | -1.251142 | 0.097276 | 1 |
| POSITIVE_REGULATION_OF_I_KAPPAB_KINASE_NF_KAPPAB_CASCADE | 4 | -0.496008 | -1.182803 | 0.269311 | 1 |
| I_KAPPAB_KINASE_NF_KAPPAB_CASCADE | 4 | -0.496008 | -1.181978 | 0.279612 | 1 |
| MUSCLE_CELL_DIFFERENTIATION | 1 | -0.879977 | -1.178438 | 0.267932 | 1 |
| SKELETAL_MUSCLE_DEVELOPMENT | 1 | -0.879977 | -1.177239 | 0.247465 | 1 |
| ACTIN_FILAMENT_ORGANIZATION | 1 | -0.879977 | -1.175023 | 0.242308 | 1 |
| MYOBLAST_DIFFERENTIATION | 1 | -0.879977 | -1.174273 | 0.225806 | 1 |
| CELL_MATURATION | 1 | -0.879977 | -1.173434 | 0.215768 | 1 |
| REGULATION_OF_I_KAPPAB_KINASE_NF_KAPPAB_CASCADE | 4 | -0.496008 | -1.170664 | 0.270793 | 1 |
| DEVELOPMENTAL_MATURATION | 1 | -0.879977 | -1.170121 | 0.258427 | 1 |
| STRIATED_MUSCLE_DEVELOPMENT | 1 | -0.879977 | -1.163648 | 0.268571 | 1 |
| DI___TRI_VALENT_INORGANIC_CATION_TRANSPORT | 1 | -0.860656 | -1.153629 | 0.281059 | 1 |
| CALCIUM_ION_TRANSPORT | 1 | -0.860656 | -1.15162 | 0.269151 | 1 |
| TRANSMISSION_OF_NERVE_IMPULSE | 5 | -0.427734 | -1.134019 | 0.301741 | 1 |
| NEGATIVE_REGULATION_OF_GROWTH | 5 | -0.428533 | -1.129907 | 0.293204 | 1 |
| NEUROPEPTIDE_SIGNALING_PATHWAY | 1 | -0.851874 | -1.126159 | 0.313929 | 1 |
| GENERATION_OF_NEURONS | 3 | -0.498471 | -1.063572 | 0.370742 | 1 |
| CELLULAR_MORPHOGENESIS_DURING_DIFFERENTIATION | 3 | -0.498471 | -1.051603 | 0.387716 | 1 |
| AXONOGENESIS | 3 | -0.498471 | -1.048951 | 0.360784 | 1 |
| NEURITE_DEVELOPMENT | 3 | -0.498471 | -1.048743 | 0.394892 | 1 |

(a)

| NAME | SIZE | ES | NES | NOM p-val | FDR q-val |
|---|---|---|---|---|---|
| BIOPOLYMER_METABOLIC_PROCESS | 77 | 0.299227 | 2.63378 | 0 | 0.001573 |
| PROTEIN_METABOLIC_PROCESS | 54 | 0.304899 | 2.440317 | 0 | 0.007927 |
| BIOPOLYMER_MODIFICATION | 34 | 0.369961 | 2.379857 | 0 | 0.006902 |
| CELLULAR_MACROMOLECULE_METABOLIC_PROCESS | 53 | 0.301526 | 2.35819 | 0 | 0.005983 |
| CELLULAR_PROTEIN_METABOLIC_PROCESS | 52 | 0.297484 | 2.31178 | 0 | 0.007372 |
| PROTEIN_MODIFICATION_PROCESS | 33 | 0.360759 | 2.283221 | 0 | 0.009136 |
| ESTABLISHMENT_OF_LOCALIZATION | 37 | 0.337147 | 2.276889 | 0.002247 | 0.008035 |
| TRANSPORT | 32 | 0.364965 | 2.269476 | 0 | 0.007441 |
| NUCLEOBASENUCLEOSIDENUCLEOTIDE_AND_NUCLEIC_ACID_METABOLIC_PROCESS | 51 | 0.293559 | 2.229045 | 0 | 0.010285 |
| REGULATION_OF_METABOLIC_PROCESS | 31 | 0.352285 | 2.183469 | 0 | 0.015742 |
| REGULATION_OF_CELLULAR_METABOLIC_PROCESS | 30 | 0.349982 | 2.105155 | 0.004255 | 0.026624 |
| POST_TRANSLATIONAL_PROTEIN_MODIFICATION | 25 | 0.374863 | 2.072546 | 0.002179 | 0.031688 |
| TRANSCRIPTION_DNA_DEPENDENT | 21 | 0.390207 | 2.048709 | 0.00655 | 0.034349 |
| LIPID_METABOLIC_PROCESS | 18 | 0.418248 | 2.037865 | 0 | 0.034457 |
| RNA_BIOSYNTHETIC_PROCESS | 21 | 0.390207 | 2.034283 | 0.006928 | 0.033129 |
| REGULATION_OF_NUCLEOBASENUCLEOSIDENUCLEOTIDE_AND_NUCLEIC_ACID_METABOLI | 22 | 0.393323 | 2.025445 | 0.010799 | 0.032658 |
| REGULATION_OF_GENE_EXPRESSION | 27 | 0.342591 | 1.99611 | 0.004386 | 0.039535 |
| REGULATION_OF_TRANSCRIPTION | 20 | 0.388187 | 1.969418 | 0.011547 | 0.045701 |
| MEMBRANE_ORGANIZATION_AND_BIOGENESIS | 8 | 0.587362 | 1.956017 | 0.004292 | 0.047546 |
| NEGATIVE_REGULATION_OF_CELLULAR_PROCESS | 30 | 0.317921 | 1.951571 | 0.006757 | 0.047192 |
| VESICLE_MEDIATED_TRANSPORT | 12 | 0.482124 | 1.95149 | 0.004444 | 0.045099 |
| TRANSCRIPTION | 25 | 0.33535 | 1.93859 | 0.00211 | 0.046933 |
| MEMBRANE_LIPID_BIOSYNTHETIC_PROCESS | 5 | 0.708178 | 1.937867 | 0.004415 | 0.045169 |
| GLYCEROPHOSPHOLIPID_METABOLIC_PROCESS | 5 | 0.695814 | 1.927841 | 0 | 0.046462 |

(b)

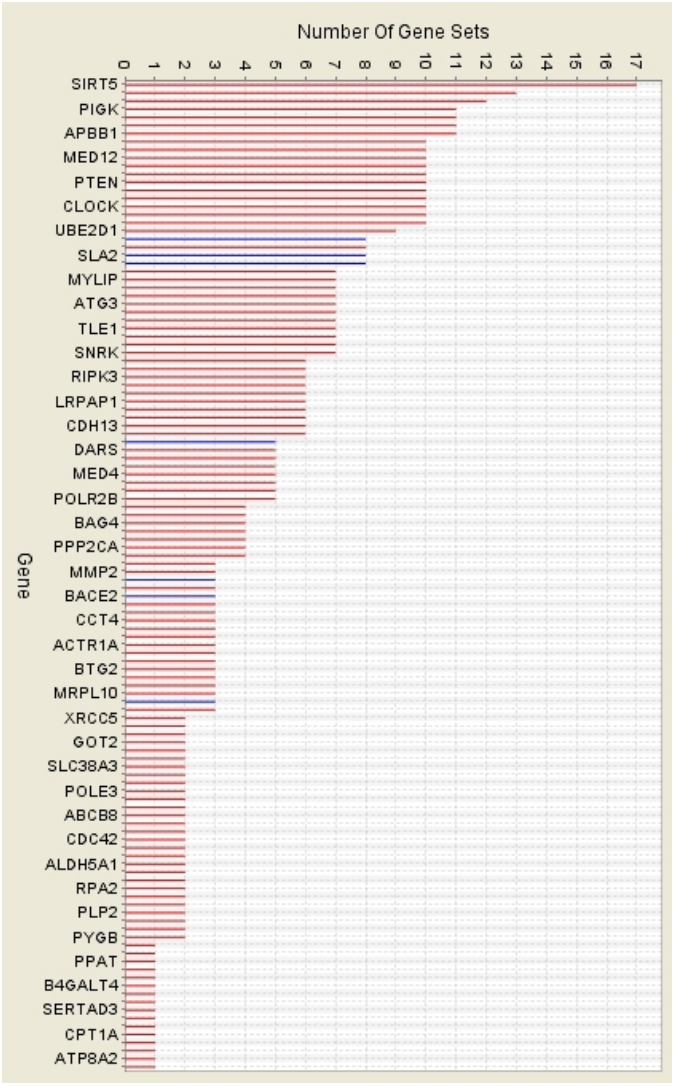


**Figure 17**. **GSEA Leading Edge Analysis For Prerank Biological Process Gene Set Database**. Leading edge analysis performed on Gene Set Enrichment Analysis (GSEA) significant downregulated gene sets from Broad Institute molecular signatures database (MsigDB) biological gene sets. Shown are genes on Y axis contributing most to enrichment of leading edge gene sets. Number of gene sets comprising dataset indicated on X axis. False discovery rate ≤0.05.

**Table 7. GSEA Prerank Mode Molecular Function Gene Set Database Results**. Upregulation of gene sets for phenotype 3F8 FULL when compared to 1C10 FULL when analyzed by Gene Set Enrichment Analysis (GSEA) Prerank mode using Broad Institute molecular signatures database (MsigDB) molecular function gene sets (a). Downregulation of gene sets for phenotype 3F8 FULL compared to 1C10 FUL showed no significant enrichment(b). Name column indicating gene set description. Size column indicating how many individual gene sets in particular database set. Results indicated by enrichment score (ES), normalized enrichment score (NES), nominal p value (NOM p-val) and false discovery rate for p value (FDR q-val). Significance measured by false discovery rate ≤ 0.05 (b).

| NAME | SIZE | ES | NES | NOM p-val | FDR q-val |
|---|---|---|---|---|---|
| RECEPTOR_ACTIVITY | 23 | 0.359018 | 1.909073 | 0.012848 | 0.491237 |
| TRANSFERASE_ACTIVITY_TRANSFERRING_GLYCOSYL_GROUPS | 7 | 0.530481 | 1.691443 | 0.024017 | 1 |
| SMALL_PROTEIN_CONJUGATING_ENZYME_ACTIVITY | 2 | 0.945345 | 1.64989 | 0.008097 | 1 |
| TRANSFERASE_ACTIVITY_TRANSFERRING_PHOSPHORUS_CONTAINING_GROUPS | 21 | 0.318041 | 1.639207 | 0.034325 | 0.820494 |
| HYDROLASE_ACTIVITY_ACTING_ON_ESTER_BONDS | 15 | 0.358514 | 1.607284 | 0.032587 | 0.814778 |
| PHOSPHORIC_ESTER_HYDROLASE_ACTIVITY | 12 | 0.397472 | 1.56766 | 0.060606 | 0.892631 |
| ISOMERASE_ACTIVITY | 2 | 0.868827 | 1.566162 | 0.041394 | 0.776324 |
| LIPID_BINDING | 5 | 0.578516 | 1.564846 | 0.051896 | 0.685119 |
| KINASE_ACTIVITY | 18 | 0.321918 | 1.552756 | 0.056277 | 0.660839 |
| PHOSPHORIC_MONOESTER_HYDROLASE_ACTIVITY | 9 | 0.441217 | 1.536999 | 0.054705 | 0.654998 |
| CATION_BINDING | 7 | 0.483121 | 1.527887 | 0.077088 | 0.627769 |
| NUCLEOBASENUCLEOSIDENUCLEOTIDE_KINASE_ACTIVITY | 3 | 0.685429 | 1.494372 | 0.057732 | 0.692695 |
| DNA_BINDING | 15 | 0.329133 | 1.479379 | 0.088795 | 0.69166 |
| TRANSMEMBRANE_RECEPTOR_ACTIVITY | 15 | 0.337524 | 1.478628 | 0.084821 | 0.644996 |
| TRANSFERASE_ACTIVITY_TRANSFERRING_HEXOSYL_GROUPS | 5 | 0.557068 | 1.478625 | 0.077895 | 0.601996 |
| TRANSFERASE_ACTIVITY_TRANSFERRING_PENTOSYL_GROUPS | 2 | 0.817295 | 1.47463 | 0.061224 | 0.575421 |
| TRANSFERASE_ACTIVITY_TRANSFERRING_ALKYL_OR_ARYLOTHER_THAN_METHYL( | 4 | 0.60828 | 1.473934 | 0.08243 | 0.544431 |
| GLUTATHIONE_TRANSFERASE_ACTIVITY | 4 | 0.60828 | 1.467607 | 0.08 | 0.531064 |
| ATPASE_ACTIVITY_COUPLED | 3 | 0.66927 | 1.440472 | 0.076321 | 0.577344 |
| HELICASE_ACTIVITY | 2 | 0.787234 | 1.434045 | 0.078029 | 0.566272 |
| DNA_DEPENDENT_ATPASE_ACTIVITY | 2 | 0.787234 | 1.427327 | 0.090535 | 0.556629 |
| STRUCTURAL_CONSTITUENT_OF_CYTOSKELETON | 4 | 0.595801 | 1.426341 | 0.090909 | 0.533296 |
| UNFOLDED_PROTEIN_BINDING | 5 | 0.52793 | 1.410863 | 0.103175 | 0.552155 |
| ATP_DEPENDENT_DNA_HELICASE_ACTIVITY | 2 | 0.787234 | 1.405388 | 0.099796 | 0.542025 |

(a)

| NAME | SIZE | ES | NES | NOM p-val | FDR q-val |
|---|---|---|---|---|---|
| TRANSMEMBRANE_RECEPTOR_PROTEIN_TYROSINE_KINASE_ACTIVITY | 2 | -0.755222 | -1.340089 | 0.123314 | 1 |
| ELECTRON_CARRIER_ACTIVITY | 3 | -0.604842 | -1.265476 | 0.176817 | 1 |
| CARBOXYLESTERASE_ACTIVITY | 1 | -0.915301 | -1.225888 | 0.175649 | 1 |
| MONOVALENT_INORGANIC_CATION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY | 1 | -0.913544 | -1.22523 | 0.186475 | 1 |
| HYDROGEN_ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY | 1 | -0.913544 | -1.215931 | 0.149789 | 1 |
| PHOSPHOLIPASE_A2_ACTIVITY | 1 | -0.915301 | -1.215831 | 0.191781 | 1 |
| CYTOCHROME_C_OXIDASE_ACTIVITY | 1 | -0.913544 | -1.195801 | 0.177515 | 1 |
| RIBONUCLEOPROTEIN_BINDING | 1 | -0.875878 | -1.175544 | 0.245614 | 1 |
| RNA_SPLICING_FACTOR_ACTIVITYTRANSESTERIFICATION_MECHANISM | 1 | -0.882123 | -1.172984 | 0.247335 | 1 |
| 3_5_CYCLIC_NUCLEOTIDE_PHOSPHODIESTERASE_ACTIVITY | 1 | -0.869633 | -1.160275 | 0.228916 | 1 |
| CYCLIC_NUCLEOTIDE_PHOSPHODIESTERASE_ACTIVITY | 1 | -0.869633 | -1.153878 | 0.269231 | 1 |
| AMINE_BINDING | 1 | -0.851874 | -1.150496 | 0.252918 | 1 |
| OXIDOREDUCTASE_ACTIVITY_ACTING_ON_CH_OH_GROUP_OF_DONORS | 3 | -0.533163 | -1.141975 | 0.274704 | 1 |
| EXTRACELLULAR_LIGAND_GATED_ION_CHANNEL_ACTIVITY | 1 | -0.851874 | -1.141276 | 0.289474 | 1 |
| LIGAND_GATED_CHANNEL_ACTIVITY | 1 | -0.851874 | -1.140576 | 0.304432 | 1 |
| OXIDOREDUCTASE_ACTIVITY_GO_0016616 | 3 | -0.533163 | -1.134718 | 0.306306 | 1 |
| EXCITATORY_EXTRACELLULAR_LIGAND_GATED_ION_CHANNEL_ACTIVITY | 1 | -0.851874 | -1.12145 | 0.318 | 1 |
| OXIDOREDUCTASE_ACTIVITY_ACTING_ON_THE_CH_NH_GROUP_OF_DONORS | 1 | -0.809524 | -1.082416 | 0.373984 | 1 |
| CATION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY | 5 | -0.401341 | -1.079096 | 0.353612 | 1 |
| TRANSMEMBRANE_RECEPTOR_PROTEIN_KINASE_ACTIVITY | 3 | -0.508593 | -1.073268 | 0.346642 | 1 |
| PROTEIN_BINDING_BRIDGING | 3 | -0.487895 | -1.050788 | 0.37911 | 1 |
| PROTEIN_TYROSINE_KINASE_ACTIVITY | 3 | -0.492906 | -1.048714 | 0.369888 | 1 |
| MOLECULAR_ADAPTOR_ACTIVITY | 3 | -0.487895 | -1.046019 | 0.396947 | 1 |

(b)

**Table 8. GSEA Prerank Mode Immunological Signatures Gene Set Database Results**. Upregulation of gene sets for phenotype 3F8 FULL when compared to 1C10 FULL when analyzed by Gene Set Enrichment Analysis (GSEA) prerank mode using Broad Institute molecular signatures database (MsigDB) immunological signatures gene sets (a). Downregulation of gene sets for phenotype 3F8 FULL compared to 1C10 FULL (b). Name column indicating gene set description. Size column indicating how many individual gene sets in particular database set. Results indicated by enrichment score (ES), normalized enrichment score (NES), nominal p value (NOM p-val) and false discovery rate for p value (FDR q-val). Significance measured by false discovery rate ≤ 0.05 (b).

| NAME | SIZE | ES | NES | NOM p-val | FDR q-val |
|---|---|---|---|---|---|
| GSE24142_DN2_VS_DN3_THYMOCYTE_FETAL_DN | 9 | 0.870132 | 3.038425 | 0 | 0 |
| GOLDRATH_EFF_VS_MEMORY_CD8_TCELL_DN | 13 | 0.663165 | 2.691305 | 0 | 0.00532 |
| GSE17721_LPS_VS_PAM3CSK4_6H_BMDM_UP | 11 | 0.636343 | 2.513937 | 0 | 0.022855 |
| GSE27786_LIN_NEG_VS_MONO_MAC_UP | 9 | 0.709001 | 2.503286 | 0 | 0.0203 |
| GSE2706_UNSTIM_VS_2H_R848_DC_DN | 10 | 0.66145 | 2.494914 | 0 | 0.019116 |
| GSE3982_EOSINOPHIL_VS_TH1_UP | 10 | 0.676243 | 2.482843 | 0 | 0.017427 |
| GSE360_L_MAJOR_VS_B_MALAYI_LOW_DOSE_MAC_UP | 11 | 0.629495 | 2.438442 | 0 | 0.02194 |
| GSE17974_CTRL_VS_ACT_IL4_AND_ANTI_IL12_72H_CD4_TCELL_DN | 8 | 0.742757 | 2.424921 | 0 | 0.022809 |
| GSE2706_UNSTIM_VS_2H_LPS_DC_DN | 10 | 0.662654 | 2.413104 | 0 | 0.022279 |
| GSE17721_CTRL_VS_LPS_8H_BMDM_DN | 16 | 0.520864 | 2.38934 | 0 | 0.024008 |
| GSE20715_WT_VS_TLR4_KO_24H_OZONE_LUNG_UP | 17 | 0.521199 | 2.388571 | 0 | 0.021825 |
| GSE27786_LSK_VS_BCELL_DN | 8 | 0.69787 | 2.381562 | 0.004357 | 0.021506 |
| GSE22886_NAIVE_CD4_TCELL_VS_12H_ACT_TH2_DN | 10 | 0.634213 | 2.3499 | 0 | 0.026126 |
| GSE22886_NAIVE_CD4_TCELL_VS_48H_ACT_TH2_DN | 10 | 0.645943 | 2.346459 | 0.00216 | 0.025302 |
| GSE17721_0.5H_VS_8H_GARDIQUIMOD_BMDM_DN | 10 | 0.634571 | 2.311752 | 0.002212 | 0.034913 |
| GSE29617_CTRL_VS_DAY3_TIV_FLU_VACCINE_PBMC_2008_UP | 7 | 0.733054 | 2.301823 | 0.002058 | 0.036577 |
| GSE11924_TH2_VS_TH17_CD4_TCELL_UP | 7 | 0.729191 | 2.299502 | 0 | 0.035053 |
| GSE17580_TREG_VS_TEFF_S_MANSONI_INF_DN | 6 | 0.775151 | 2.292438 | 0 | 0.0348 |
| GSE3982_NEUTROPHIL_VS_TH1_DN | 13 | 0.548318 | 2.279536 | 0.002283 | 0.036674 |
| GSE20715_0H_VS_48H_OZONE_TLR4_KO_LUNG_UP | 11 | 0.591004 | 2.278455 | 0 | 0.035113 |
| GSE37416_0H_VS_12H_F_TULARENSIS_LVS_NEUTROPHIL_DN | 12 | 0.560728 | 2.26078 | 0 | 0.038413 |
| GSE17721_POLYIC_VS_PAM3CSK4_8H_BMDM_DN | 16 | 0.4991 | 2.257453 | 0.002203 | 0.03782 |
| GSE20366_TREG_VS_NAIVE_CD4_TCELL_DEC205_CONVERSION_UP | 11 | 0.582715 | 2.255623 | 0 | 0.036571 |
| GSE14308_TH2_VS_NATURAL_TREG_DN | 10 | 0.604281 | 2.22772 | 0 | 0.045958 |

(a)

| NAME | SIZE | ES | NES | NOM p-val | FDR q-val |
|---|---|---|---|---|---|
| GSE14026_TH1_VS_TH17_UP | 2 | -0.869217 | -1.558639 | 0.027027 | 1 |
| GSE22886_NAIVE_CD8_TCELL_VS_MEMORY_TCELL_UP | 4 | -0.645734 | -1.536096 | 0.046693 | 1 |
| GSE17974_CTRL_VS_ACT_IL4_AND_ANTI_IL12_1H_CD4_TCELL_UP | 6 | -0.515921 | -1.472259 | 0.077213 | 1 |
| GSE14308_TH2_VS_TH17_UP | 7 | -0.463421 | -1.447395 | 0.109259 | 1 |
| GSE2706_UNSTIM_VS_8H_R848_DC_UP | 4 | -0.584456 | -1.431061 | 0.103846 | 1 |
| GSE360_L_MAJOR_VS_M_TUBERCULOSIS_MAC_DN | 6 | -0.508387 | -1.428143 | 0.11546 | 1 |
| GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_DN | 4 | -0.593362 | -1.399448 | 0.104508 | 1 |
| GSE17721_4_VS_24H_GARDIQUIMOD_BMDM_UP | 7 | -0.442564 | -1.355956 | 0.145129 | 1 |
| GSE10239_NAIVE_VS_KLRG1HIGH_EFF_CD8_TCELL_UP | 8 | -0.419777 | -1.355896 | 0.136194 | 1 |
| GSE3982_EFF_MEMORY_VS_CENT_MEMORY_CD4_TCELL_DN | 3 | -0.637642 | -1.338314 | 0.130841 | 1 |
| GSE10325_MYELOID_VS_LUPUS_MYELOID_UP | 6 | -0.472944 | -1.338306 | 0.153101 | 1 |
| GSE7460_TCONV_VS_TREG_THYMUS_DN | 8 | -0.404128 | -1.316337 | 0.13308 | 1 |
| GSE360_L_DONOVANI_VS_B_MALAYI_LOW_DOSE_MAC_DN | 6 | -0.450579 | -1.295702 | 0.177449 | 1 |
| GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP | 4 | -0.524312 | -1.274083 | 0.155769 | 1 |
| GSE17721_12H_VS_24H_PAM3CSK4_BMDM_UP | 11 | -0.333652 | -1.254922 | 0.20073 | 1 |
| GSE3982_EFF_MEMORY_VS_CENT_MEMORY_CD4_TCELL_UP | 4 | -0.525692 | -1.253477 | 0.209259 | 1 |
| GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_48H_ACT_CD4_TCELL_UP | 6 | -0.431581 | -1.249441 | 0.189922 | 1 |
| GSE15659_NAIVE_CD4_TCELL_VS_RESTING_TREG_UP | 4 | -0.524312 | -1.24784 | 0.213235 | 1 |
| GSE9006_TYPE_1_DIABETES_AT_DX_VS_4MONTH_POST_DX_PBMC_DN | 7 | -0.399672 | -1.242851 | 0.197782 | 1 |

(b)

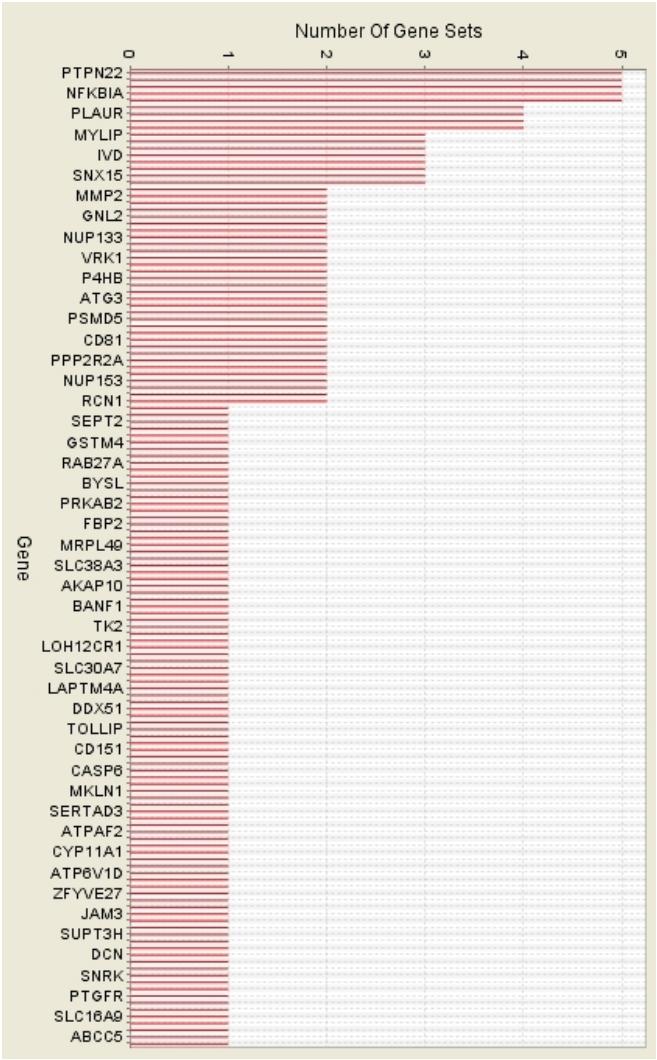


**Figure 18**. **GSEA Leading Edge Analysis for Prerank Immunological Signatures Database**. Leading edge analysis performed on Gene Set Enrichment Analysis (GSEA) significantly upregulated gene sets from Broad Institute molecular signatures database (MsigDB) immunological signatures gene sets. Shown are genes on Y axis contributing most to enrichment of leading edge gene sets. Number of gene sets comprising dataset indicated on X axis. False discovery rate ≤0.05.

### *4.5 Quantitative Polymerase Chain Reaction*

Ribonucleic acid (RNA) quality analysis was performed on phenotype groups 3F8 FULL (n=5), 3F8 SHAM (n=5), 1C10 FULL (n=5) and 1C10 SHAM (n=5). Total RNA ratio of the peak areas of the ribosomal bands 18s and 28s were measured. Complementary RNA

(cRNA) ratio of the peak areas of the ribosomal bands 18*s* and 28s were measured.

Electropherograms for cRNA were analyzed (Figure 19).

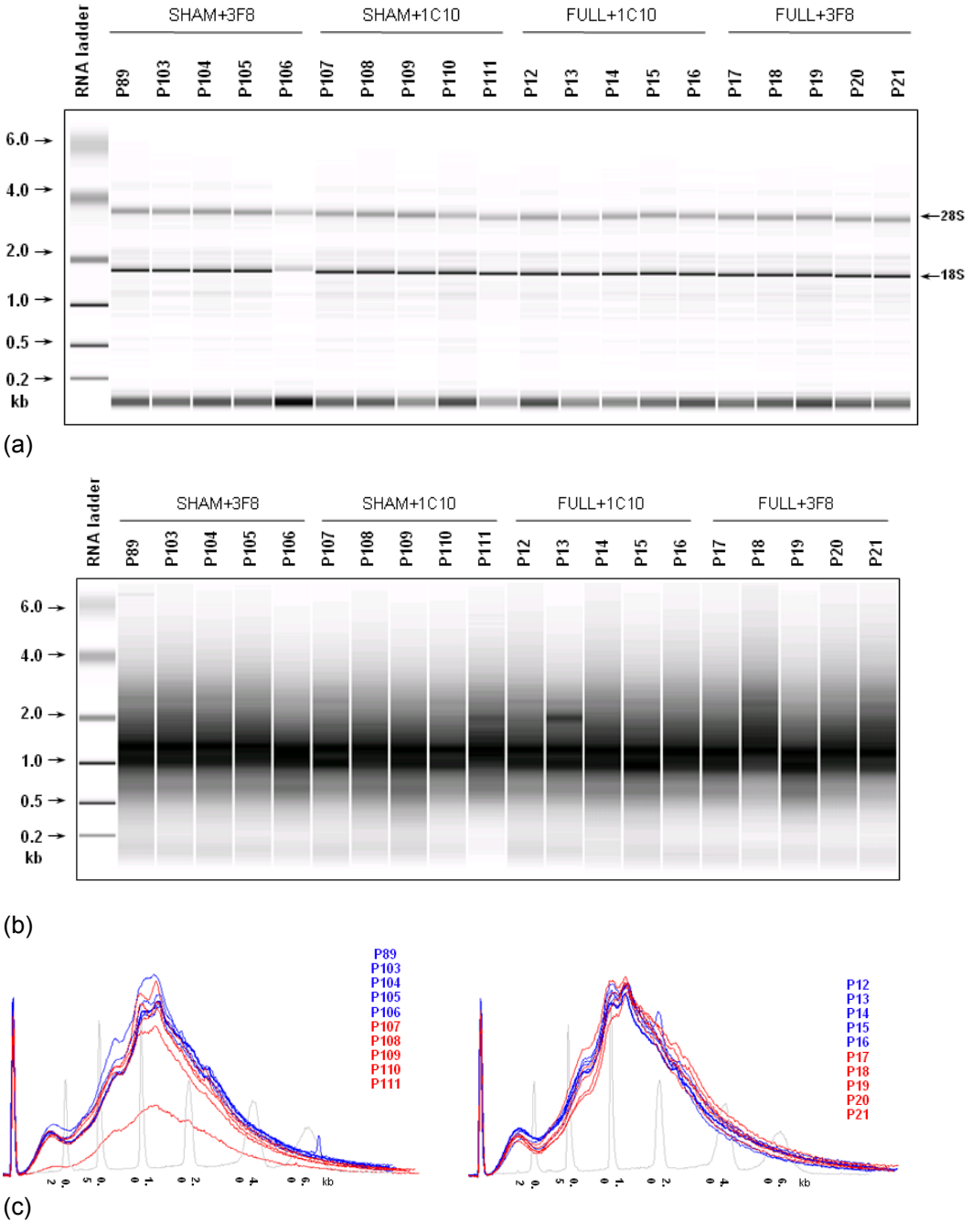


**Figure 19. RNA Quality Analysis**. Phenotype groups 3F8 SHAM (n=3), 1C10SHAM (n=3),1C10 FULL and 3F8 FULL (n=3). Total RNA ratio of peak areas of ribosomal bands 18s/28s. (a). Complementary RNA (cRNA) ratio of peak areas of ribosomal bands 18s/28s quality analysis (b). cRNA quality analysis electropherograms (c).

Quantitative polymerase chain reaction was performed between the two phenotype grouping 3F8 FULL vs. 3F8 SHAM and 3F8 FULL vs. 1C10 FULL. The 3F8 FULL vs. 3F8 SHAM examined expression change for genes *Hspa1a* and *Xrl4a*. There was significant difference in Ct value of *Hspa1a* in the 3F8 FULL (n=6) phenotype verses 3F8 SHAM (n=6) mice (Figure 20). Delta (D) Ct value showed a significant difference between 3F8 FULL (n=6) and 3F8 SHAM (n=6) when compared to reference gene *18s* (Figure 21). The Delta – Delta Ct change was – 3.04 indicating an upregulated 8.22 fold increase in phenotype 3F8 FULL compared to 3F8 SHAM. There was no significant difference in expression of *Xrl4a* as indicated by Ct value between 3F8 FULL (n=6) and 3F8 SHAM (n=6) (Figure 22). Delta (D) Ct value showed no significant difference between 3F8 FULL (n=6) and 3F8 SHAM (n=6) when compared to reference gene *18s* (Figure 23). The Delta – Delta Ct change was – 0.75 indicating an upregulated 1.68 fold increase in phenotype 3F8 FULL compared to 3F8 SHAM.

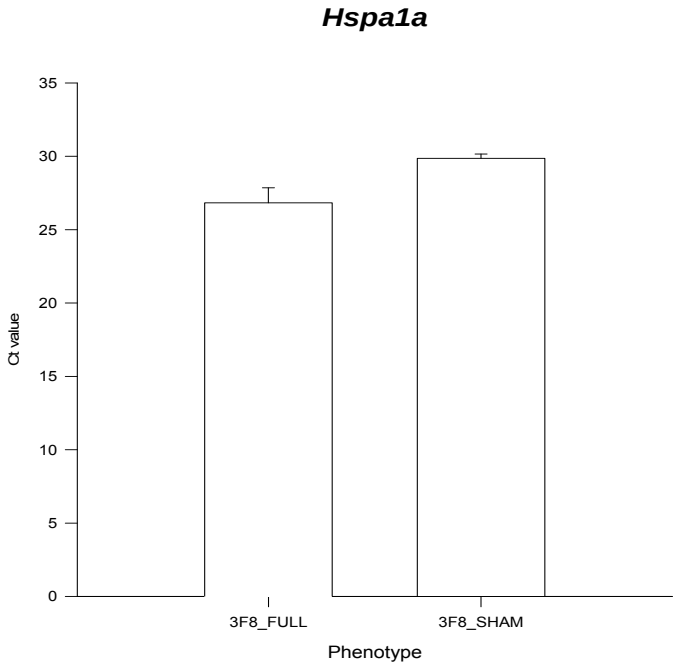


**Figure 20. qPCR Ct Values for Gene Hspa1a In Phenotype 3F8 FULL vs. 3F8 SHAM**. Absolute quantitative polymerase chain reaction (qPCR) Ct values for gene *Hspa1a* in phenotypes 3F8_FULL and 3F8_SHAM. Phenotype 3F8_FULL significantly different from 3F8_SHAM ($p \leq 0.001$).

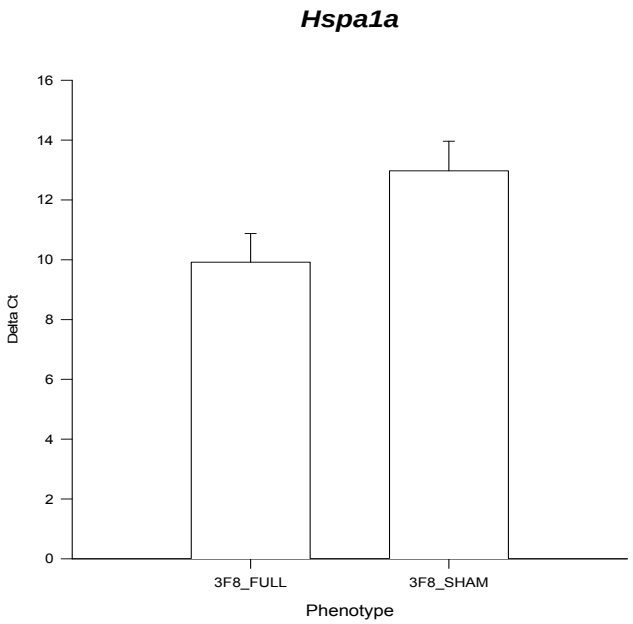


**Figure 21. qPCR Delta Ct Values Gene Hspa1a In Phenotype 3F8 FULL vs. 3F8 SHAM**. Quantitative polymerase chain reaction (per) Delta (D) Ct values for gene *Hspa1a* in phenotypes 3F8_FULL and 3F8_SHAM. Phenotype 3F8_FULL significantly different from 3F8_SHAM (p = 0.002).

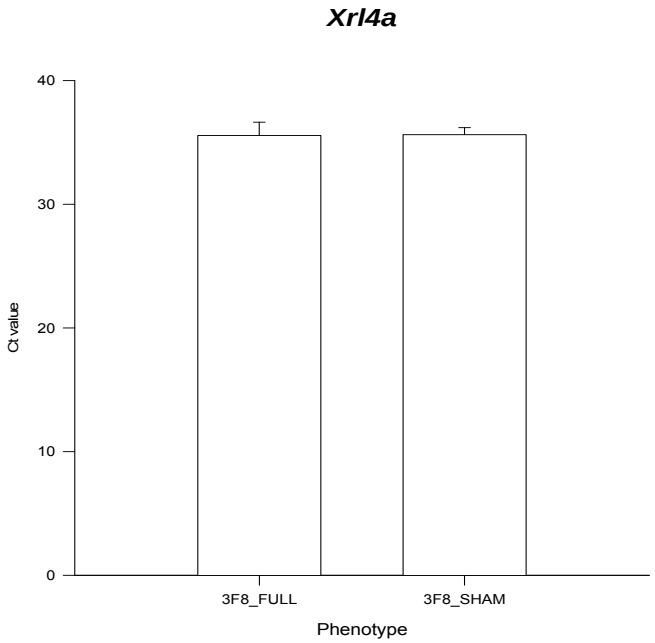


**Figure 22. qPCR Ct Values for Gene Xrl4a In Phenotype 3F8 FULL vs. 3F8 SHAM**. Absolute quantitative polymerase chain reaction (per) Ct values for gene *Xrl4a* in phenotypes 3F8_FULL and 3F8_SHAM. Phenotype 3F8_FULL no significance from 3F8_SHAM (p = 0.906).

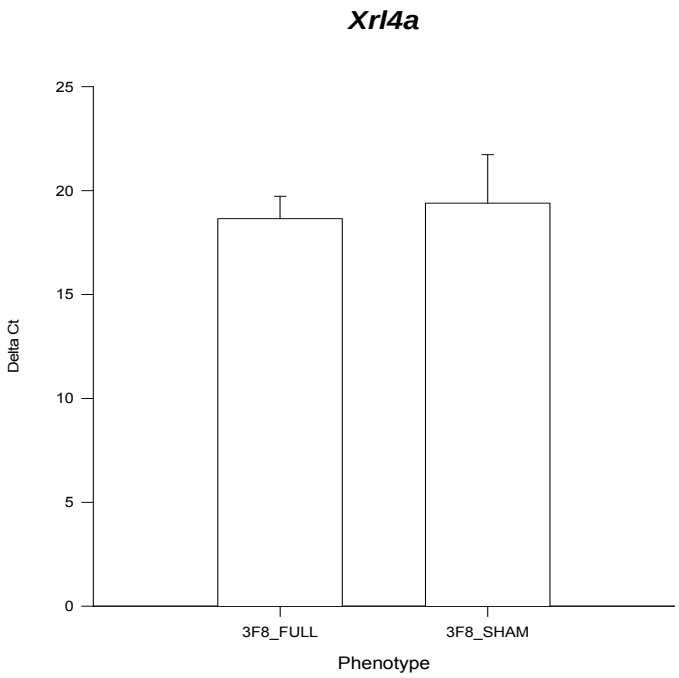


**Figure 23. qPCR Delta Ct for Gene Xrl4a In Phenotype 3F8 FULL vs. 3F8 SHAM**. Quantitative polymerase chain reaction (per) Delta (D) Ct values for gene *Xrl4a* in phenotypes 3F8_FULL and 3F8_SHAM. Phenotype 3F8_FULL had no significant difference from 3F8_SHAM ($p = 0.496$).

The second phenotype grouping, 3F8 FULL vs. 1C10 FULL underwent per. There was no significant difference in expression of gene *BC030870* (A_66_P126348) as indicated by Ct value between 3F8 FULL (n=6) and 1C10 FULL (n=6) (Figure 24). Delta (D) Ct value showed no significant difference between 3F8 FULL (n=6) and 1C10 FULL (n=6) when compared to reference gene *18s* (Figure 25). The change in delta – delta Ct value was – 0.33 indicating an upregulated 1.2 fold increase in phenotype 3F8 FULL compared to 1C10 FULL.

### *4.6 Molecular Modeling*

Thioredoxin (PDB 2TRX) imaged in UCSF Chimera version 1.9 identifying disulfide bond and protein sequence within active loop between Cysteine (Cys) 32 and Cys 35 (Figure 26). Computational insertion of flitrix peptide sequence FGQHFDGLPTSA into active loop site

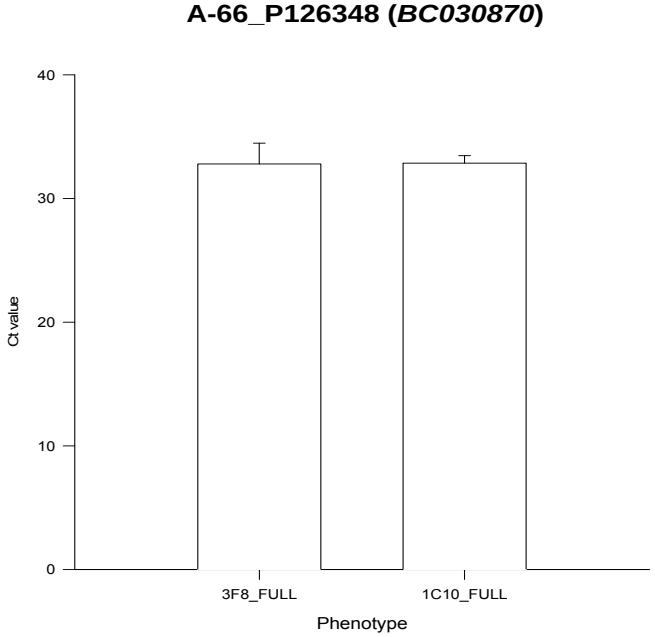


**Figure 24. qPCR Ct Values for Gene BC030870 In Phenotype 3F8 FULL vs. 1C10 FULL**. Absolute quantitative polymerase chain reaction (qPCR) Ct values for gene *BC030870* in phenotypes 3F8_FULL and 1C10_FULL. Phenotype 3F8_FULL no significance from 1C10_FULL (p = 0.496).

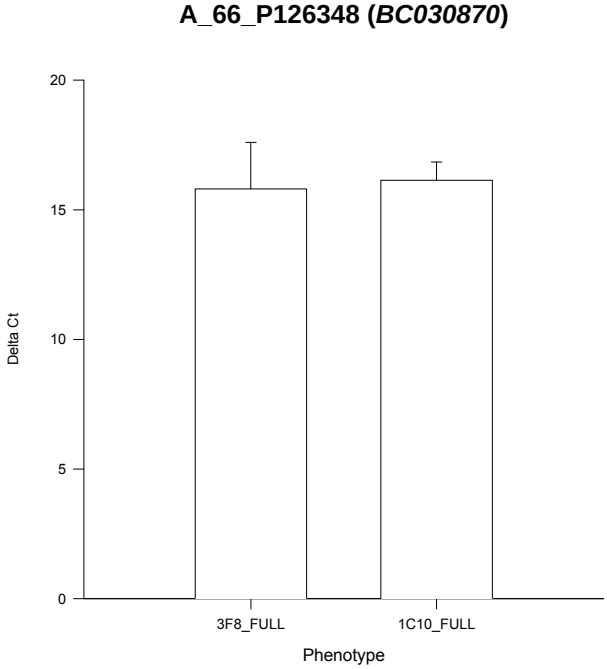


**Figure 25. qPCR Delta Ct Values for Gene BC030870 In Phenotype 3F8 FULL vs. 1C10 FULL**. Quantitative polymerase chain reaction (qPCR) Delta (D) Ct values for gene *BC030870* in phenotypes 3F8_FULL and 1C10_FULL. Phenotype 3F8_FULL had no significant difference from 1C10_FULL (p = 0.678).

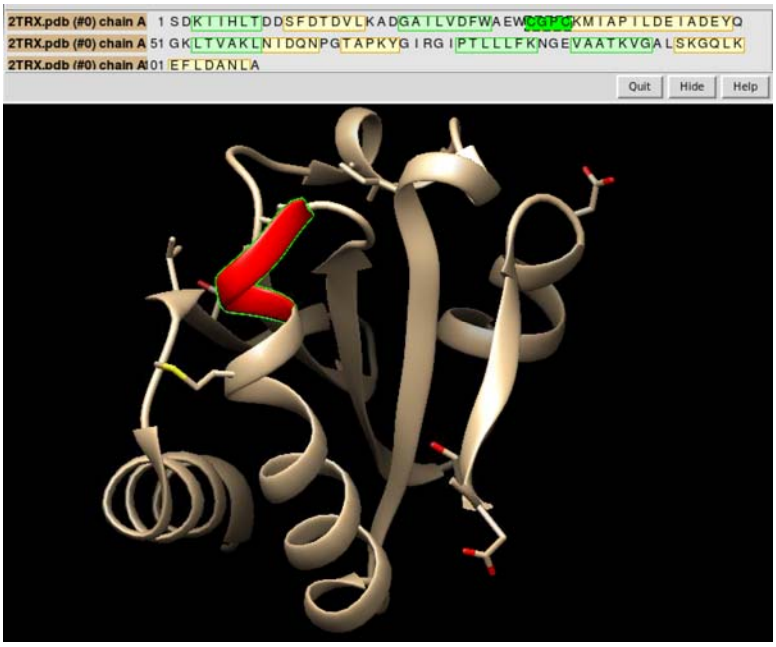


**Figure 26. Thioredoxin With Active Loop and Disulfide Bond**. Thioredoxin (PDB 2TRX) Chain A. Red indicates the disulfide bond and active loop between residues Cys 32 and Cys 35 in native structure.

of Thioredoxin between Cys 32 and Cys 35 (Figure 27). Thioredoxin (2TRX) with flitrix peptide having undergone loop modeling and refinement in ModLoop (Figure 28). Isolated flitrix peptide sequence in Sybyl-X which will be used for three dimensional protein structure comparison (Figure 29)(Table 9). RCSB PDB protein comparison tool FATCAT flexible comparison algorithm indicating LGKQVGNKFFL with a root mean square deviation (RMSD) of 1.76 as mAb 3F8 epitope for hMBL (Figure 30). RCSB PDB protein comparison tool FATCAT rigid comparison algorithm indicating LGKQVGNKFFL with a RMSD of 1.76 as mAb 3F8 epitope for hMBL (Figure 31). MArkovian TRAnsition of Structure evolution protein three dimensional structure comparison algorithm indicating LGKQVGNKFFL with a RMSD of 3.01 as mAb 3F8 epitope for hMBL (Figure 32). RCSB PDB protein comparison algorithm tool CE indicating protein sequence CVLLLK-QWNDV with a RMSD of 12.72 as mAb 3F8 epitope for hMBL (Figure 33). RCSB PDB protein comparison algorithm tool Smith Waterman indicating protein sequence EGEPNNA with a RMSD of 3.45 as mAb 3F8 epitope for hMBL (Figure 34).

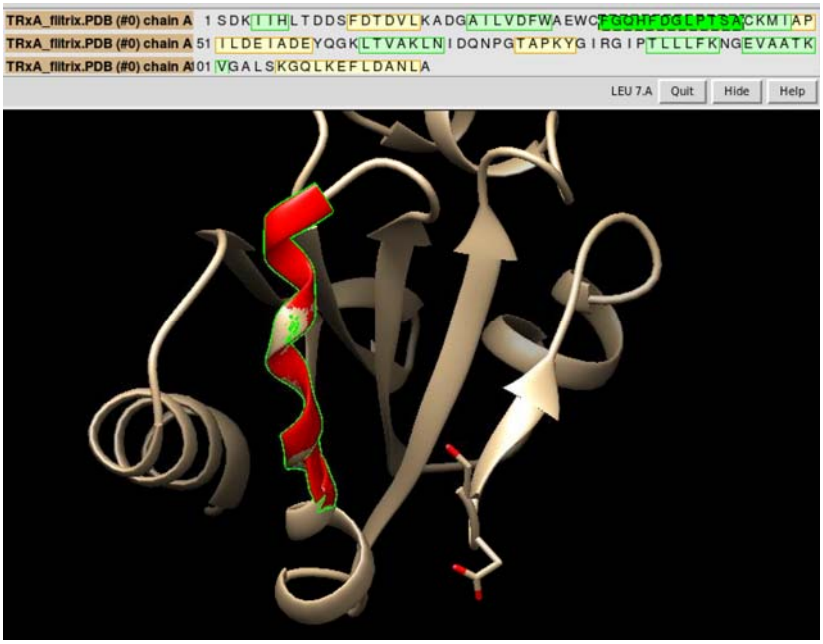


**Figure 27. Thioredoxin With Flitrix Peptide Inserted Into Active Loop From SWISS MODEL**. Red indicates inserted flitrix peptide residues recognized by mAb 3F8 into active loop between Cys 32 and Cys 35 (now Cys 45), FGQHFDGLPTSA. Dark green is protein FASTA residues inserted. Noted in this model is that the disulfide bond is not preserved.


**Figure 28. Thioredoxin With Flitrix Peptide Recognized by mAb 3F8 After Loop Modeling In ModLoop Program**. Red indicates successful computational insertion of flitrix peptide residues recognized by mAb 3F8 and preservation of disulfide bond between Cys 32 and Cys 45 indicated by white arrows.

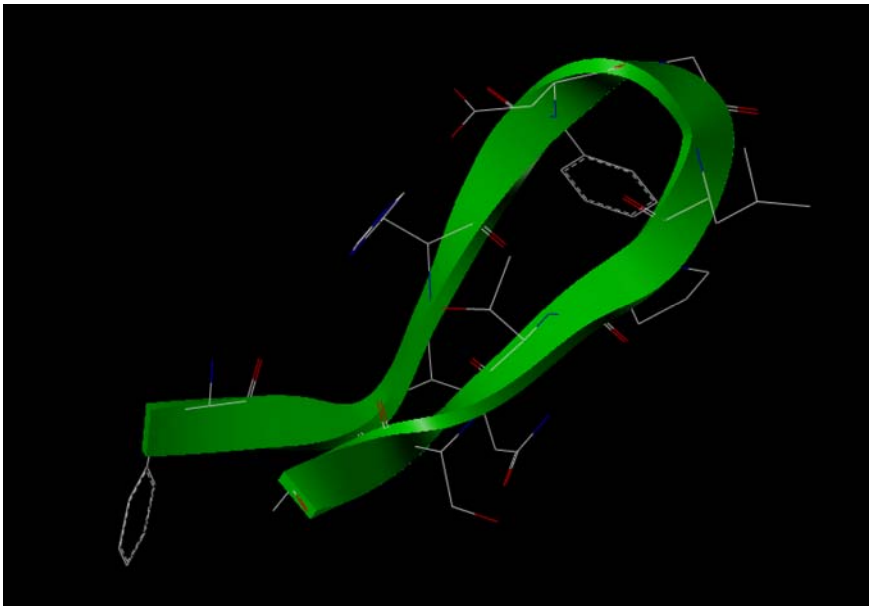

**Figure 29. Computationally Isolated Flitrix Peptide Recognized by mAb 3F8.** Protein Sequence FGQHFDGLPTSA three dimensional loop structure for use in molecular modeling of mAb 3F8 epitope for hMBL.

**Table 9. 3D Protein Structural Comparison Results**. Summary of three dimensional protein comparison results determining structural similarity for epitope identification between flitrix peptide recognized by mAb 3F8 and human mannose binding lectin (hMBL). Above are programs used, epitope sequence identified and calculated root mean square deviation (RMSD).

| Flitrix Peptide | 3D Structural Comparison Program | Epitope Identified | RMSD |
|---|---|---|---|
| FGQHFDGLPTSA | FATCAT_RIGID | LGKQVGNKFFL | 1.76 |
| FGQHFDGLPTSA | FATCAT_FLEX | LGKQVGNKFFL | 1.76 |
| FGQHFDGLPTSA | MATRAS | LGKQVGNKFFL | 3.01 |
| FGQHFDGLPTSA | Smith Waterman | EGEPNNA | 3.45 |
| FGQHFDGLPTSA | Combinational Extension (CE) | CVLLK-QWNDV | 12.72 |

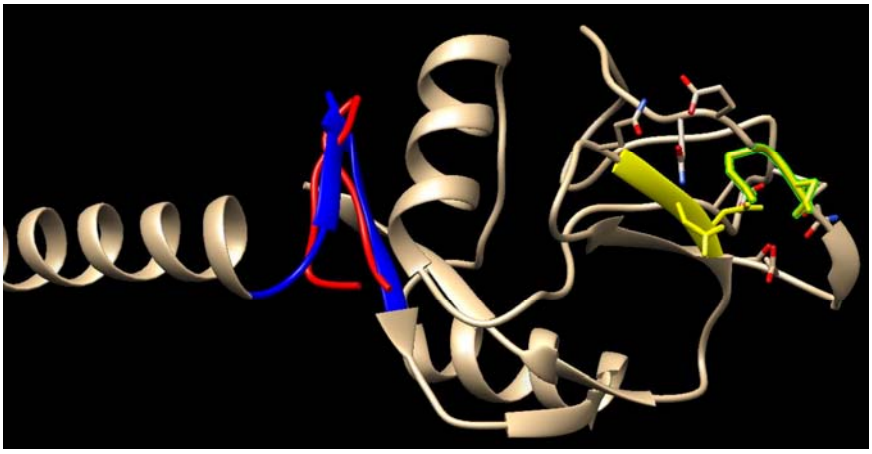

**Figure 30. 3D Structure Comparison Using FATCAT Flexible Algorithm**. Human mannose binding lectin (hMBL, PDB 1HUP) superimposed three dimensional structural alignment with flitrix peptide recognized by mAb 3F8 based on FATCAT flexible algorithm. Red indicates flitrix peptide residues FGQHDGLPTSA. Blue indicates mAb 3F8 epitope residues on 1HUP, LGKQVGNKFFL. Yellow indicates mannose binding residues EPN and WND.

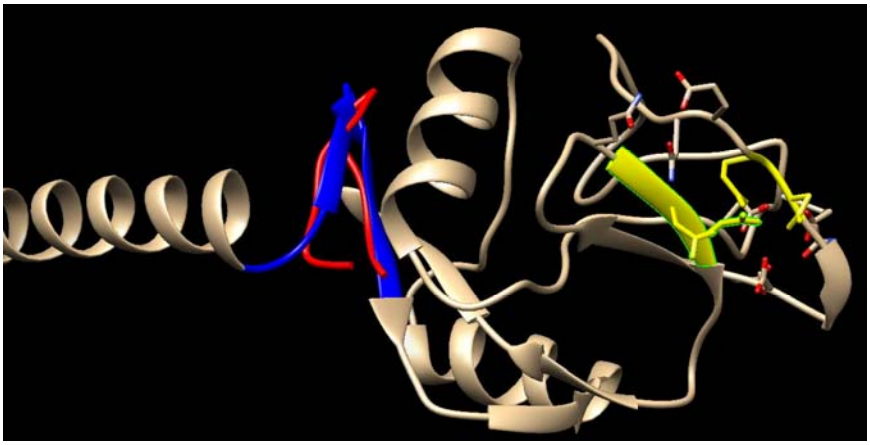

**Figure 31. 3D Structure Comparison Using FATCAT Rigid Algorithm**. Human mannose binding lectin (hMBL, PDB 1HUP) superimposed three dimensional structural alignment with flitrix peptide recognized by mAb 3F8 based on FATCAT rigid algorithm. Red indicates flitrix peptide residues FGQHDGLPTSA. Blue indicates mAb 3F8 epitope residues on 1HUP, LGKQVGNKFFL. Yellow indicates mannose binding residues EPN and WND.

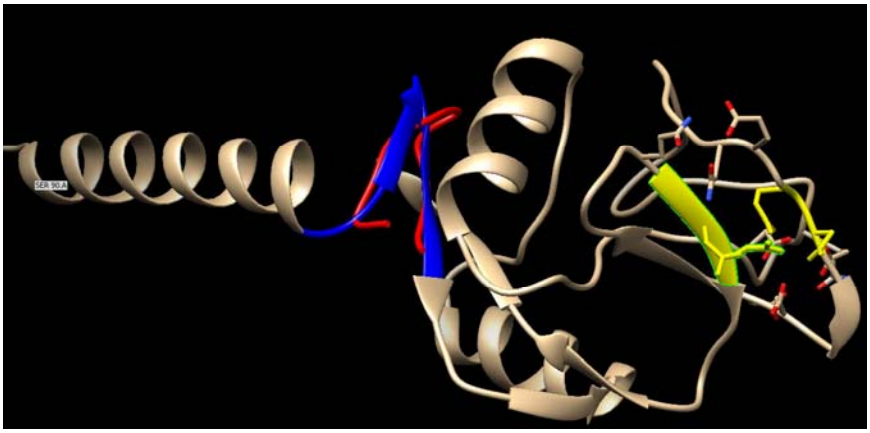

**Figure 32. 3D Structure Comparison Using MATRAS Algorithm**. Human mannose binding lectin (hMBL, PDB 1HUP) superimposed three dimensional structural alignment with flitrix peptide recognized by mAb 3F8 based on MATRAS algorithm. Red indicates flitrix peptide residues FGQHDGLPTSA. Blue indicates mAb 3F8 epitope residues on 1HUP, LGKQVGNKFFL. Yellow indicates mannose binding residues EPN and WND.

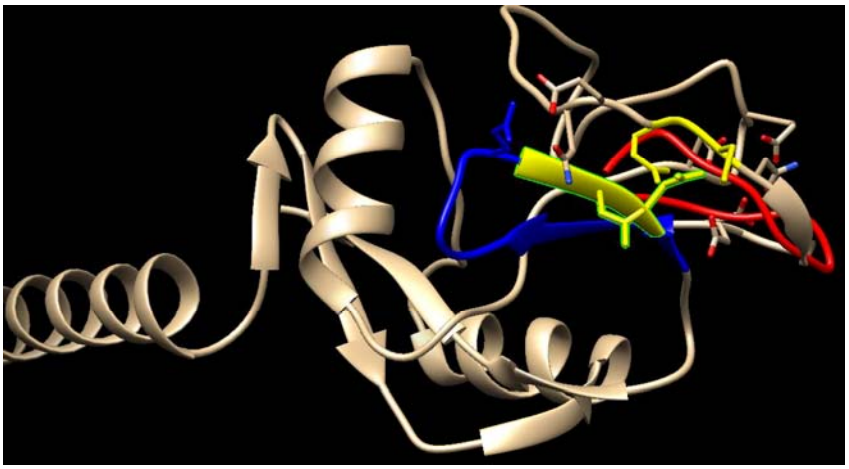

**Figure 33. 3D Structure Comparison Using Combination Extension (CE) Algorithm**. Human mannose binding lectin (hMBL, PDB 1HUP) superimposed three dimensional structural alignment with flitrix peptide recognized by mAb 3F8 based on Combinational Extension (CE) alignment. Red indicates flitrix peptide residues FGQHDGLPTSA. Blue indicates mAb 3F8 epitope residues on 1HUP, CVLLLK-QWNDV. Yellow indicates mannose binding residues EPN and WND.

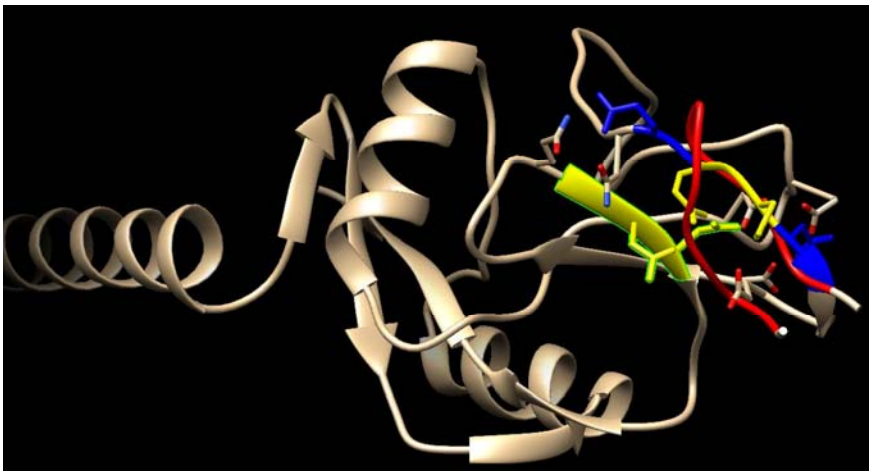

**Figure 34. 3D Structure Comparison Using Smith Waterman Algorithm**. Human mannose binding lectin (hMBL, PDB 1HUP) superimposed three dimensional structural alignment with flitrix peptide recognized by mAb 3F8 based on Smith Waterman superposition alignment. Red indicates flitrix peptide residues FGQHDGLPTSA. Blue indicates mAb 3F8 epitope residues on 1HUP, EGEPNNA. Yellow indicates mannose binding residues EPN and WND.

## CHAPTER V

## DISCUSSION

Drug target elucidation can be very difficult to assess in non-human models of disease. Recently there has been increased debate on the efficacy on such use. For many years the paradigm of disease models included the use of mice. With the significant advances in –omic driven technologies, including genomics, the apparent debate over models such as mice in studying human disease may show some warrant. Mice for example in regards to the complement system encode 2 genes for MBL, *Mbl-a* and *Mbl-c*. Humans however only encode one such gene *MBL2*. There is a second gene *MBL1* which is a pseudogene and has not been evolutionarily conserved in humans.[24,25,26]

In this study and through the Stahl lab we have been the first to accomplish and show successful expression of $hMBL^{+/+}$ mice. In MI/R injury models when assessing the AAR, variability must be avoided and show no statistical difference in the AAR/LV, as was the case in this research.[39] This uniform reliability allows for appropriate statistical assessment from one subject to the next and within similar groupings. This study was able to reproduce an MI/R mouse model with the use of a novel therapeutic mAb 3F8 in novel $hMBL^{+/+}$ mice with no variability in AAR. Results demonstrated that $hMBL^{+/+}$ mice with the same AAR treated with mAb 3F8 had less percent infarct than untreated mice. The success of such a model opens to door to understanding the role of hMBL in many human disease processes. Continued gene studies will need to be examined as to the extent that hMBL in the mouse regulates biological disease processes that mimic humans.

No study before this has examined the role of $hMBL^{+/+}$ mice in MI/R with use of whole genome microarray. With the influx of significant and ever changing genomic tools and methods assessing this data and cross referencing any such studies can become difficult. The sheer volume of data collected also becomes difficult to manage in a consistent cross platform manner. Variation includes how the microarray data is processed from its raw form,

background noise correction, normalization techniques, handling of replicate spots and which expression data value to use. There is also debate on the appropriate statistical markers to use for assessing expression importance such as p values, adjusted p values, false discovery rates (FDR), enrichment scores, ranking, etc.

This study was successful in its ability to process the raw microarray data in a reliable manner while preserving the integrity of such data. As with any microarray study the sheer volume of expression values from whole genome probes made interpreting the results challenging. In this study when examining the contrast matrix of most importance to us, hMBL undergoing full surgery treated with mAb 3F8 versus those treated with mAb 1C10, surprisingly out of the top 5,000 genes with p value ≤ 0.05 only five had a significant adjusted p value. The adjusted p value which is the p value adjusted for multiple testing and is thought to be an indication of the false discovery rate or chance that an expression value is misrepresented or overly confident. Thus, although the top five genes with significant adjusted p value may be of special interest, the remaining statistically significant genes based on p value may also help derive biological function knowledge and thus the importance of such programs like GSEA, as used in this study.

The importation, manipulation, R code writing and statistical analysis was duplicated and confirmed by an outside research source from the Computational Biology and Functional Genomics Laboratory at the Dana Farber Cancer Institute. Three out of the five significant limma topTable genes as indicated by adjusted p value are also uncharacterized genes belonging to the LncRNA family. Assessing their significance was found to be difficult. Identified in correspondence analysis was one model from each of the aforementioned group as outliers on the hierarchal clustering graphs and scatter plots. It was unable to be determined whether those mice were more accurate surgically and in treatment delivery or less accurate and poorer controlled than the others. All mice underwent surgery and treatment in an operator blinded manner with the same method of delivery and surgical technique being confirmed with electrocardiographic evidence prior, during and after MI/R.

The handling of all tissue samples were also carried out in the same strict RNA free driven protocols as indicated by the strength in the RNA quality results.

As this was the first mouse hMBL$^{+/+}$ model to be developed and used in MI/R and whole genome microarray further evaluation is necessitated. The microarray chip used was for whole mouse genome from Agilent Technologies. These gene probes were then mapped to other varying forms of identifiers. This mapping for assessment was unexpectedly difficult in that the top five genes were uncharacterized. This also brings about the future need to assess what downstream gene expression is hMBL in the mouse model regulating and what best chip platform to accurately identify those changes. If the mouse and human genome differ in not just *MBL* gene conservation then the affect this has on interpreting whole genome microarray use of human genes in animal models, such as used in this study will need to be explored.

Gene set enrichment analysis determines gene sets based on prior biological knowledge.[132] The program runs internally a modified t-test that produces a ranking of all genes. With that ranking it then calculates an enrichment score for each gene set based on a weighted Kolmorgorov-Smirnoff score. It then permutes the class labels of the samples and recalculates the enrichment score on those randomized data to derive an empirical p value. False positives are controlled by calculating the FDR which is the probability that a gene set with a given nominal enrichment score represents a false positive.[126]

Difficulties arise in using significant volumes of genes and probes such as in this study because correcting for multiple testing to derive a FDR will result in the loss of many genes. To account for such a problem this study devised two methods for GSEA use. The first was to use the top one hundred genes from limma topTable in raw data form. The second was to use the top five thousand genes from limma topTable using their respected p value for ranking selection. Furthermore, gene set databases for comparison were selected for their anticipated similar biological pathway functioning and included biological process, molecular functioning and immunological signatures databases. The top one hundred method when compared to

The Broad Institutes gene sets biological processes, molecular functions and immunological signatures showed no significant gene set enrichment either up or down. When the top five thousand gene ranking method was used already having been preranked by limma the biological process database showed significantly upregulation of gene sets consistent with metabolic pathways. The immunological signatures database also showed a significant upregulation. Significant down regulation based on nominal p values were reported but they were not significant when assessed for FDR values.

When examining this data it is apparent that a majority of the genes in the top one hundred raw data method are uncharacterized genes. These are genes that most likely because of the lack of whole genome expression models in mice have limited prior knowledge discovery in those gene set databases used. Although significant from a statistical standpoint and biological function standpoint this makes assessing them difficult. This does however offer novel areas for biological function discovery. The findings of this study and the significance of such uncharacterized genes can now be added to current and future databases for continued biological knowledge comparisons, especially in regards to the effects of mAb 3F8 attenuating MI/R injuries in $hMBL^{+/+}$ mice.

There are statistical concerns when using the GSEA preranking method. The use of Limma in advance may result in overconfident results because it has already corrected for multiple testing and enriched the genes that are differentially expressed. Preranked genes as the ones used in this study cannot undergo phenotype permutation but instead gene permutation which may result in overconfident results.[126] Using only those genes indicated as significant for input into GSEA also blunts the power of GSEA and caution is recommended according to the developers (rt. broad #164113). This study attempted to navigate the many possible pitfalls when using GSEA by using both raw intensity prove values as well as the preranked method and in selecting those gene set databases with most probable similar biological pathways. As a result, the top one hundred raw data method does mimic some of the same genes of significance importance as did the Limma program in terms of up or

downregulation ranking. The GSEA program may not have shown significant gene set enrichment also due to the uncharacterized nature of those genes.

Quantitative polymerase chain reaction often accompanies gene microarray results in attempts by the authoring researchers to deem those results as valid or to correlate such data. This can problematic at times when qPCR does not confirm those results or in fact reveal more conflicting results. There are also many pitfalls that can occur when acquiring and processing RNA. In this study, harvested left ventricles of mice undergoing sham or full MI/R injury followed a strict RNA procedure including not only the work environment but the handling and storage of tissue samples. The resultant total and complimentary RNA quality assessments as reported reveal high level quality RNA. Microarray results show that Hspa1a is significantly upregulated in mice undergoing MI/R treated with mAb 3F8 compared to sham mice. The qPCR in this study correlates with those results. However, when comparing the same phenotype model gene *Xlr4a* showed no statistical significance in fold change when using the change in ΔΔCt method. This is in contrast to previously reported results in this study in which microarray results showed significant downregulation in phenotype 3F8 FULL versus 3F8 SHAM for gene *Xrl4a*. The most significantly downregulated gene of interest when comparing phenotypes 3F8 FULL and 1C10 FULL and the one with the most significant fold change reported previously was gene probe A_66_P126348, referencing gene *BC030870*. During qPCR however and using the change in ΔΔCt to then calculate fold change, no significant difference was noted. This was an unexpected result in that the microarray data analysis revealed this gene to be the most significant and most significantly downregulated gene. This was mirrored in the GSEA ranking with this gene having the highest downregulated score.

The aforementioned differing results can be attributable to many possibilities. First, both genes although assessed in different phenotype models are highly uncharacterized genes of interest. Gene *BC030870* also belongs to the lncRNA family of genes. lncRNA expression is developmentally regulated and can be tissue and cell specific.[162] lncRNA

consist of many varying transcripts and can be transcribed as partial or whole antisense transcripts to coding genes.[163] Exertion of LncRNA function is by binding to DNA or RNA in a specific sequence manner and at low levels of expression.[163] There are four archetypes of lncRNA; signal, decoy, guide and scaffold.[165] Various disease types including cancer, cardiovascular disease, neurological disease and immune-mediated diseases have been associated with lncRNA's.[166] Protein structure function prediction is difficult in assessing lncRNA's because unlike protein-coding genes whose sequence motifs indicate function, the lncRNA sequences and their secondary structures are not well conserved as is their motifs.[167] These aforementioned characteristics of lncRNA may have contributed to the difficulty in correlating the gene microarray results and per results.

There are also technical issues which could explain the different results between microarray and qPCR. The gene of interest *BC030870* could also have been a false positive. However, using the FDR value of 8.39E-05 the probability of it being a false value is statistically highly unlikely. More plausible would be that the primers that were developed are targeting other splice variants than the probes on the microarray. Also, this study only used one reference gene, *18s*, and for qPCR normalization may not have been adequate. However, all the Ct values for both phenotype groups, 3F8 FULL and 1C10 FULL were not of significant difference. The probes on the microarray and the qPCR primers may also be designed to different isoforms or are hitting different genes. Further work will need to be done to resolve these discrepancies.

When discussing the three dimensional structure characteristics between mAb 3F8 and hMBL included should be an explanation of previous N and C terminal deletion construct studies, the hinge region, epitope region and C terminal residue region. The mAb 3F8 epitope for hMBL has previously been reported to be within the hinge region. Traditionally in structural biology a 'hinge' region refers to flexible residues in a protein structure connecting two distinct domains. This allows for variation in relative orientation due to its flexibility. Structurally, hMBL sequence LGKQVGNKFFL forms one unique domain whereas sequence

MARIKKWLTFLS forms an alpha helix. Helix structures however are not considered flexible. Thus, part of MARIKKWLTFLS and C-type lectin region could possibly be deemed a 'hinge'.

Disulfide bonds are one of the most critical structural determinants and are preserved throughout evolution. Protein structure will be different for those with and without disulfide bonding. Three dimensional structure is the most important characteristic for mAb recognition and modifications such as deletions which do not preserve the three dimensional structure destroy the native structure of the protein. Human mannose binding lectin contains several disulfide bonds. It has been previously reported that during N and C terminal deletion constructs mAb 3F8 did not recognize any of the C terminal deletion constructs. Those constructs did not preserve disulfide bonding and thus the three dimensional structure in that region most likely changed resulting in a poor 3F8 epitope for hMBL. This could also indicate that C terminal residues are in fact not part of the epitope.

None of the N terminal deletion constructs affected disulfide bonding, but only the N1 deletion was recognized by mAb 3F8.[97] Because no three dimensional local region structure changes would have occurred, the helix residues MARIKKWLTFSL, which, is removed during N2 and N3 deletions may play an important factor in mAb 3F8 recognition of hMBL. A helix cannot be a hinge, but a part of this helix, such as the end, may be able to function as a hinge residue. N2 truncation studies preserve all but L in the hypothesized epitope region LGKQVGNKFFL in this study. This in combination with the removal of previously reported epitope MARIKKWLTFSL places what we believe to be the true epitope in the position of being the N terminal residues. The N2 truncation may result in a gained flexibility of the epitope GKQVGNKFFL which in 3D crystal form alters its structure and biological function, thus its ability to recognize mAb 3F8.

In this study and after thorough review of the Stahl lab's previous work the three dimensional structural characteristics of mAb 3F8 recognition of hMBL was performed utilizing molecular modeling techniques.[97] Small free peptide modeling could not be performed because it would be too flexible and would not have a fixed structure in solution. Random

short peptides also do not have fixed conformations for mAb recognition. Thus, whether performed experimentally or computational as was performed in this study, a model peptide was needed to be constructed within the frame of a template protein. This component was not evident in review of the previous work. The purpose of this study performing local structure alignment is to identify the three dimensional elements of native hMBL (PDB # 1HUP) with the flitrix peptide identified from the Stahl lab work that is recognized by mAb 3F8 and inserted in active loop of Thioredoxin . The difference among mAb binding to flitrix random peptides is not due to Thioredoxin, but the peptide portion which shares the recognition of hMBL in mAb. As a result, the flitrix peptide within Thioredoxin should have the three dimensional structure that is similar to a portion of hMBL that is recognized by mAb 3F8. This is not true for the whole hMBL protein which, does not have similar structure to Thioredoxin and thus cannot be compared in its entirety.

This study using the RCSB PDB protein comparison tool located three possible mAb 3F8 epitopes for hMBL. Four different algorithm based comparison tools were utilized. The RCSB PDB protein comparison tool web server itself is based on the FATCAT and CE algorithms. RCSB PDB tool also allows for sequence based structure alignment using the Smith Waterman superposition. Also used was the MATRAS three dimensional structure alignment. The MATRAS and FATCAT programs isolated the same epitope region LGKQVGNKFFL. This was regardless if FATCAT utilized flexible or rigid parameters. The identified sequence is just proceeding the closest N terminal hinge region MARIKKWLTFSL. These results support the Stahl lab's previous findings that mAb 3F8 only recognized N1 terminus truncation and that by removing a peptide sequence on either side of hinge region, such as LGKQVGNKFFL, abolished the binding ability of mAb 3F8 to hMBL. This may indicate and support the importance of MARIKKWLTFSL is stabilizing the LGKQVGNKFFL epitope three dimensional structure.

Also within the RCSB PDB protein comparison tool used in this study was the CE algorithm method. The resultant epitope sequence identified was CVLLLK-QWNDV. The

later part of the sequence lies within one of the two mannose binding recognition sequences *WND*. Lastly, the Smith Waterman superposition algorithm on the RCSB PDB protein comparison tool was used. The epitope sequence identified was EGEPNNA. This sequence segment also is located within the hinge region of hMBL and is also comprised partly of one of the mannose binding recognition sequences.

The algorithms and techniques utilized by these aforementioned tools used in this study themselves offer some explanation as to why the epitope identification results varied in all but the FATCAT and MATRAS algorithms. The Smith Waterman superposition algorithm is a local sequence based comparison for three dimensional structure alignments. Algorithms that use a template based prediction with sequence information purely as a starting point are less reliable due to the presumption that the available database pool of protein-protein structures to serve as putative templates for three dimensional structure building may not be sufficiently large enough.[168] The Smith Waterman superposition performs a local alignment instead of a total sequence examination comparing all possible lengths. Following this protein sequence alignment the three dimensional proteins are superimposed based solely on the initial sequence alignment step. The weakness lies in comparing three dimensional structure characteristics based on an alignment that is optimized for maximum sequence similarity and not three dimensional similarity as used by FATACAT, CE and MATRAS.[169] Furthermore, there is no rational that the sequence of flitrix peptide has any match to hMBL because it is a random sequence peptide, thus making algorithms based solely on sequence alignment inadequately reliable.

The MATRAS model for three dimensional protein structure comparison and alignment is based on a novel scoring scheme. A structure similarity score is defined as the log-odds of two probabilities. This score detects evolutionary homologous structure similarities. The MATRAS scheme is similar to Dayhoff's amino acid substitution model. Unlike the Smith Waterman superposition algorithm FATCAT and CE compare aligned fragment pairs (AFP) during alignment process. These algorithms offer strength in three

dimensional protein structure comparisons because those AFP's are based on similarity in local geometry. The FATCAT algorithm allows for twists termed flexible comparison in which different regions of the protein structure can undergo different geometric transformations. The FATCAT algorithm can also be run in rigid mode from the start or will convert to rigid comparison if twists are forced to "0". The CE algorithm explores continuous alignment paths by extending its structural comparison beyond the AFP's leading to an optimal alignment which, is only performed based on rigid alignment.[169]

Included in this research was molecular modeling of the newly identified mAb 3F8 epitope for hMBL. Examination of the 3D structure characteristics between the newly identified epitope from this research and the terminal N and C residues from previously proposed discontinuous epitope show that MARIKKWLTFSL and C terminal residues are not as close in 3D structural space. This research identified that newly hypothesized epitope LGKQVGNKFFL is almost between those regions in structural comparison (Figure 35). Measurement between these three fragments quantitatively reveals more 3D contacts between the newly identified epitope LGKQVGNKFFL and terminal sequences forming discontinuous epitope than that of those N and C terminal sequences to each other. This indicates that the newly identified epitope in 3D space in closer to those terminal sequences than they are to each other and located in between them which, may further support it's role as the true epitope and not as previously described solely a discontinuous epitope (Figure 36)(Table 10).

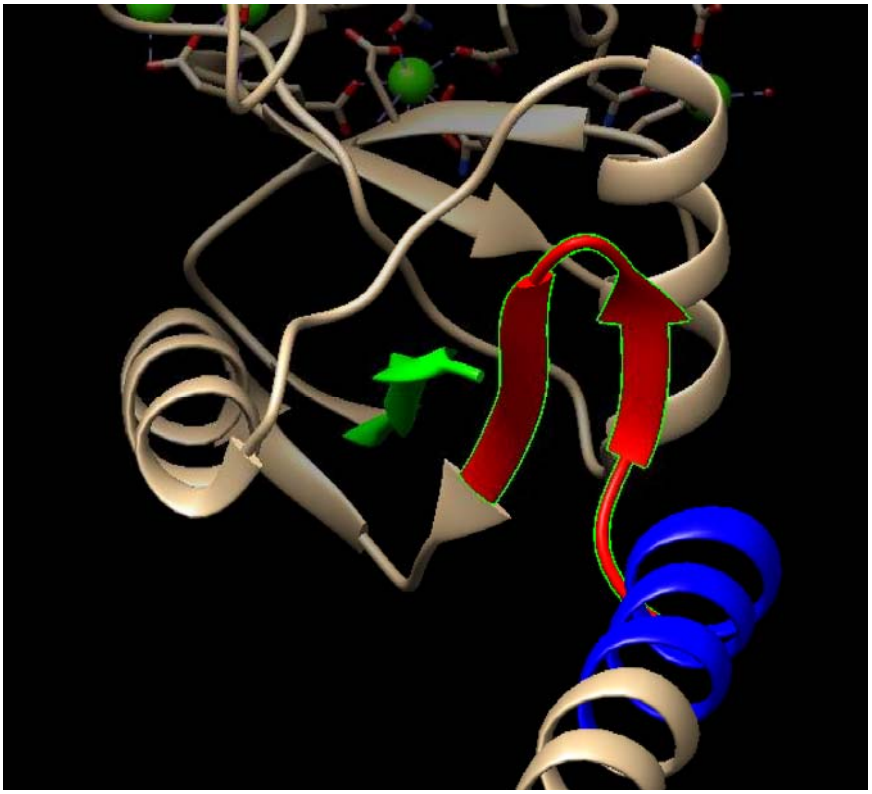

**Figure 35. Native Human Mannose Binding Lectin (hMBL) Epitopes**. Previously described discontinuous epitope indicating N terminal residues MARIKKWLTFSL in blue, C terminal residues LAVCEFPI in green, and newly identified epitope LGKQVGNKFFL in red (Gorsuch, W. B., unpublished data).

**Table 10. Quantitative Distance Measurement Between hMBL Epitope Atomic Contacts**. Red indicates N terminal fragment MARIKKWLTFSL, green indicates C terminal fragment and blue indicates newly identified epitope LGKQVGNKFFL (Gorsuch, W. B., unpublished data).

| Atom1 | Atom2 | Å |
|---|---|---|
| Lys 103:NZ | Ile 228: CG1 | 4.4 |
| Thr 107: CG2 | Ile 228: CD1 | 3.6 |
| Thr 107: CG2 | Gln 113: OE1 | 3.6 |
| Ile 228: CG2 | Gln 113: NE2 | 3.9 |
| Glu 225: OE1 | Lys 117: CE | 3.4 |
| Glu 225: OE1 | Lys 117: NZ | 2.8 |

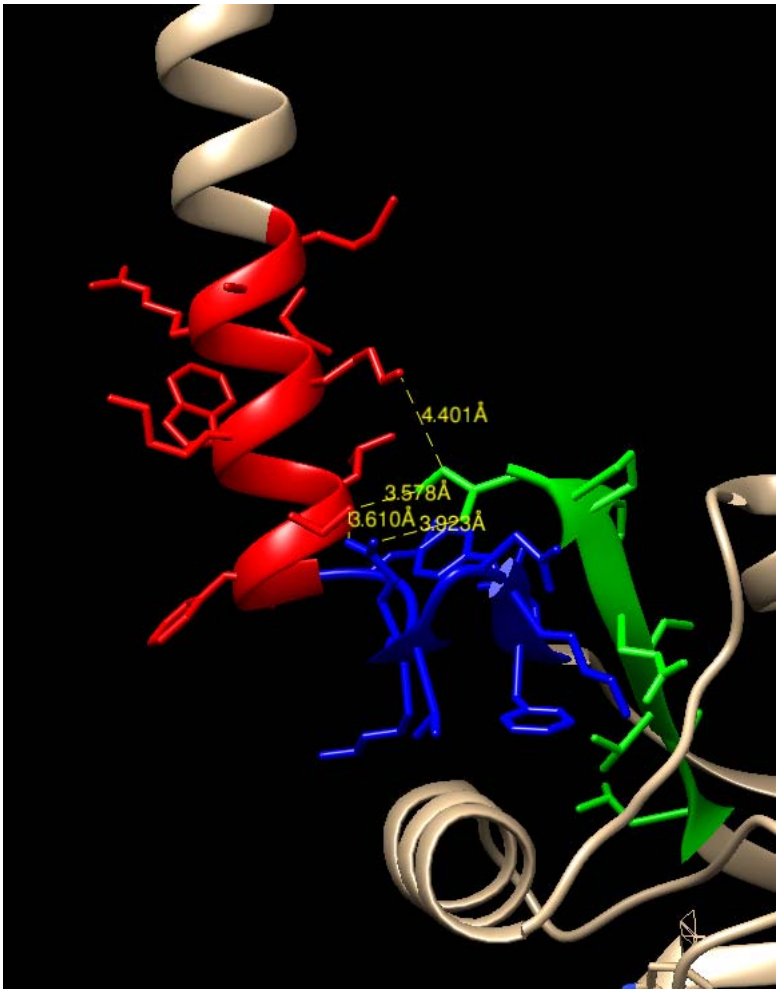


**Figure 36. 3D Image of Quantitative Distance Measurement Between Epitope Atoms**. Red indicates N terminal fragment MARIKKWLTFSL, green indicates C terminal fragment and blue indicates newly identified epitope LGKQVGNKFFL (Gorsuch, W. B., unpublished data).

Epitope residues must be exposed in 3D crystal structure in order to recognize mAb. Too many inward positioned residues will not allow for mAb recognition. When computationally evaluating the surface of all aforementioned fragments it is apparent that some of each may be involved with recognition. The C-terminal residues are less exposed to this study's hypothesized epitope (Figure 37). From molecular modeling in this research and

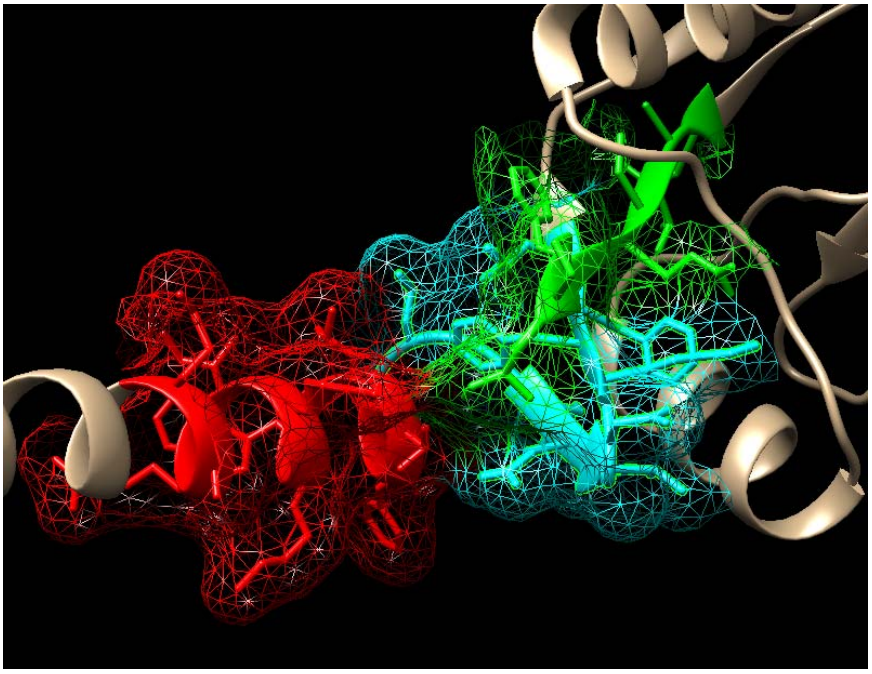

**Figure 37. Orientation of hMBL Epitope Surface Residues**. Surface residues facing outward and available to participate in mAb recognition. Red indicates N terminal fragment MARIKKWLTFSL, green indicates C terminal fragment and blue indicates newly identified epitope LGKQVGNKFFL (Gorsuch, W. B., unpublished data).

from knowledge gained in previous work done by the Stahl lab it is hypothesized that regardless of MARIKKWLTFLS, LGKQVGNKFFL or C terminal residue involvement, that mAb 3F8 may not actually inhibit hMBL through competitive binding to mannose binding site. Rather, mAb 3F8 attenuates hMBL expression by either altering monomer interaction to form a biological trimer or once mAb binds to newly identified epitope the rest of it may block carboydrate binding ability to hMBL mannose binding sites.

## CONCLUSION

This study is the first to successfully document the use of hMBL$^{+/+}$ mice and use them to not only perform an MI/R injury but to assess those affects through the use of whole genome microarrays. By comparing the infarct area and AAR, correlations may be made between the molecular and functional benefits of a novel mAb 3F8 as a therapeutic agent. The observation that mAb 3F8 in hMBL$^{+/+}$ mice undergoing MI/R limits not only the size of infarct when comparing left ventricle size and AAR, but also identified the significant down regulation of five uncharacterized genes belonging to the lncRNA family. Also identified was the difficulty in performing qPCR with such lncRNA in attempts to understand and correlate their expression characteristics. Further work will need to be done to distinguish the possible probe difficulties targeting such lncRNA and how to detect such a gene with native low levels of expression variance. The lncRNA may hold vital importance in assessing cardiac disease as this molecular class is necessary to regulate differentiation of embryonic stem cells into cardiac cells (Scheuermann et al., 2013).

This study added to the research previously performed by the Stahl lab in identifying possible mAb 3F8 epitopes for hMBL. This was achieved through new techniques utilizing homology modeling and three dimensional structure comparison. Molecular modeling assisted in better characterizing the possible effects of the hinge region in hMBL and its affect on the 3F8 epitope as well the effects of disulfide bond destruction. The goal met by this study was to better characterize the local three dimensional characteristics between mAb 3F8 and hMBL.

## References


1. Neyer JR, Greenlund KJ, Denny CH. Prevalence of heart disease—United States, 2005. *MMWR Morb Mortal Wkly Rep* 2005; 56: 113-118.

2. Heron, M. Deaths: leading causes for death. *Natl Vital Stat Rep* 2004; 56: 1-95.

3. American Heart Association. Know the facts, get the stats: our guide to heart disease, stroke and risks 2007; 55-0576.

4. Thom T, Haase N, Rosamond W, Howard VJ, Rumsfeld J, Manoliot T, Zheng, ZJ, Flegal K, O'Donnell C, Kittners S, Lloyd-Jones D, Goff DC, Hong Y, Adams R, Friday G, Furie K, Gorellick S, Wilson M, & Wolf P. Heart disease and stroke statistics update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 2006; 113: 69-171.

5. Riedemann NC, Ward PA. Complement in ischemia reperfusion injury. *Am J Pathol* 2003; 162: 363-367.

6. Asakura M, Kitakaze M. Global gene expression profiling in the failing myocardium. *Circ J* 2009; 73: 1568-1576.

7. Pavlov V, La Bonte LR, Baldwin WM, Markiewski M, Lambris J, Stahl GL. Absence of mannose-binding lectin prevents hyperglycemic cardiovascular complications. *Am J Pathol* 2012; 180: 104-112.

8. Hart ML, Ceonzo KA, Shaffer LA, Takahashi K, Rother RP, Reenstra WR, Buras JA, Stahl GL. Gastrointestinal ischemia-reperfusion injury is lectin complement pathway dependent without involving C1q. *J Immunol* 2005; 174: 6373-6380.

9. Walsh MC, Bourcier T, Takahashi K, Shi L, Busche MN, Rother RP, Solomon SD, Ezekowitz RA, Stahl GL. Mannose-binding lectin is a regulator of inflammation that accompanies myocardial ischemia and reperfusion injury. *J Immunol* 2005; 175: 541-546.

10. Murata K, Fox-Talbot K, Qian Z, Takahashi K, Stahl GL, Baldwin WM, Wasowska BA. Synergistic deposition of C4d by complement-activating and nonactivating antibodies in cardiac transplants. *Am J Transplant* 2007; 7: 2605-2614.

11. La Bonte .R, Davis-Gorman G, Stahl GL, McDonagh PF. Complement inhibition reduces injury in the type 2 diabetic heart following ischemia and reperfusion. *Am J Physiol Heart Circ Physiol* 2008; 294: H1282-H1290.

12. Busche MN, Walsh MC, McMullen ME, Guikema BJ, Stahl GL. Mannose binding lectin plays a critical role in myocardial ischemia and reperfusion injury in a mouse model of diabetes. *Diabetologia* 2008; 51: 1544-1551.

13. La Bonte LR, Dokken B, Davis-Gorman G, Stahl GL, McDonagh PF. The mannose-binding lectin pathway is a significant contributor to reperfusion injury in the type 2 diabetic heart. *Diab Vasc Dis Res* 2009; 6: 172-180.

14. Busche MN, Pavlov V, Takahashi K, Stahl GL. Myocardial ischemia and reperfusion injury is dependent on both IgM and mannose-binding lectin. *Am J Physiol Heart Circ Physiol* 2009; 297: H1853-H1859.

15. Schwaeble WJ, Lynch NJ, Clark JE, Marber M, Samani NJ, Ali YM, Dudler T, Parent B, Lhotta K, Wallis R, Farrar CA, Sacks S, Lee H, Zhang M, Iwaki D, Takahashi M, Fujita T, Tedford CE, Stover CM. Targeting of mannan-binding lectinassociated serine protease-2 confers protection from myocardial and gastrointestinal ischemia/reperfusion injury. *Proc Natl Acad Sci USA* 2011; 108: 7523-7528.

16. Busche MN, Stahl GL. Role of the complement components C5 and C3a in a mouse model of myocardial ischemia and reperfusion injury. *Ger Med Sci* 2010; 8.

17. Yang J, Moravec CS, Sussman MA, DiPaola NR, Fu D, Hawthorn L, Mitchell CA, Young JB, Francis GS, McCarthy PM, Bond M. Decreased SLIM1 expression and increased gelsolin expression in failing human hearts measured by high-density oligonucleotide arrays. *Circulation* 2000; 102(25): 3046-3052.

18. Barrans J D, Allen PD, Stanatiou D, Dzau VJ, Liew CC. Global gene expression profiling of end stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. *Am J Pathol* 2002; 160(6): 2035-2043.

19. Yung CK, Halperin VL, Tomaselli GF, Winslow RL. Gene expression profiles in end stage human idiopathic dilated cardiomyopathy: altered expression of apoptotic and cytoskeletal genes. *Genomics* 2004; 83(2): 281-297.

20. Grzeskowiak R, Witt H, Drungowski M, Therman R, Hennig S, Perrot A, Osterziel KJ, Klingbiel D, Scheid S, Spang R, Lehrach H, Ruiz P. Expression profiling of human idiopathic dilated cardiomyopathy. *Cardiovas. Res* 2003; 59(2): 400-411.

21. Colak D, Kaya N, Al-Zahrani J, Bakheet A, Muiya P, Andres E. Left ventricle global transcriptional profiling in human end stage dilated cardiomyopathy. *Genomics* 2009; 94: 20-31.

22. Buermans HP, Redout EM, Schiel AE, Musters RJP, Zuidwijk M, Eijk PP, van Hardeveld C, Kasanmoentalib S, Visser FC, Yistra B, Simonides WS. Microarray analysis reveals pivotal divergent mRNA expression profiles early in the development of either compensated ventricular hypertrophy or heart failure. *Phsiol Genomics* 2005; 21: 314-323.

23. Dewey FE, Perez MV, Wheeler MT, Watt C, Spin J, Langfelder P, Horvath S, Hannenhalli S, Cappola TP, Ashley EA. Gene expression network topology of cardiac development, hypertrophy, and failure. *Circ Cardiovasc Genet* 2011; 4: 26-35.

24. Orsini F, Villa P, Parrella S, Zangari R, Zanier ER, Gesuete R, Stravalaci M, Fumagalli S, Ottria R, Reina J, Paladini A, Micotti E, Ribeiro-Viana R, Rojo J, Pavlov VI, Stahl GL, Bernardi A, Gobbi M, De Simoni MG. Targeting mannose-binding lectin confers long lasting protection with a surprisingly wide therapeutic window in cerebral ischemia. *Circulation 2*012; 126: 1484-1494.

25. Chang WC, White MR, Moyo P, McClear S, Thiel S, Hartshorn KL. Lack of pattern recognition molecule mannose-binding lectin increases susceptibility to influenza A virus infection. *BMC Immunol* 2010; 11: 64.

26. Liu H, Jensen L, Hansen S, Petersen SV, Takahashi K, Ezekowitz AB, Hansen FD, Jensenius JC, Thiel S. Characterization and quantification of mouse mannon binding lectins (MBL-A and MBL-C) and the study of acute phase responseses. *Scandinavian J Immunol* 2001; 53: 489-497.

27. Collard CD, Vakeva A, Morrissey MA, Agah A, Rollins SA, Reenstra WR, Buras JA, Meri S, Stahl GL. Complement activation following oxidative stress: Role of the lectin complement pathway. *Am J Pathol* 2000; 156: 1549-1556.

28. Montalto MC, Collard CD, Buras JA, Reenstra WR, McClaine R, Gies DR, Rother RP, Stahl GL. A keratin peptide inhibits mannose-binding lectin. *J Immunol* 2001; 166: 4148-4153.

29. Wada K, Montalto MC, Stahl GL. Inhibition of complement C5 reduces local and remote organ injury after intestinal ischemia/reperfusion in the rat. *Gastroenterology* 2001; 120: 126-133.

30. Liangos O, Domhan S, Schwager C, Zeier M, Huber PE, Addabbo F, Goligorsky MS, Hlatky L, Jaber BL, Abdollahi A. Whole blood transcriptomics in cardiac surgery identifies a gene regulatory network connecting ischemia reperfusion with systemic inflammation. *PLoS ONE* 2010; 5: e13658.

31. Khalil AA, Aziz FA, Hall JC. Reperfusion injury. *Plast Reconstr Surg* 2006; 117: 1024-1033.

32. Arumugam TV, Shiels IA, Woodruff TM, Granger DN, Taylor SM. The role of the complement system in ischemia-reperfusion injury. *Shock* 2004; 21: 401-409.

33. Nolte D, Lehr HA, Messmer K. Adenosine inhibits postischemic leukocyteendothelium interaction in postcapillary venules of the hamster. *Am J Physiol Heart Circ Physiol* 1991; 261: H651-H655.

34. Lefer AM, Weyrich AS, Buerke M. Role of selectins, a new family of adhesion molecules, in ischaemia-reperfusion injury. *Cardiovasc Res* 1994; 28: 289-294.

35. Stotland MA, Kerrigan CL. E- and L-selectin adhesion molecules in musculocutaneous flap reperfusion injury. *Plast Reconstr Surg* 1997; 99: 2010-2020.

36. Lehr, HA, Olofsson AM, Carew TE, Vajkoczy P, Von Andrian UH, Hübne C, Berndt MC, Steinberg D, Messmer K, Arfors KE. P-selectin mediates the interaction of circulating leukocytes with platelets and microvascular endothelium in response to oxidized lipoprotein in vivo. *Lab Invest* 1994; 71: 380–386.

37. Piper HM, Meuter K, Schafer C. Cellular mechanisms of ischemia Reperfusion injuries. *AnnThorac Surg* 2003; 75: 644-648.

38. Atsuumi T, Yakima Y, Shichishima T, Maehara K, Fujita T, Maruyama Y. Complement and polymorph nuclear leukocyte activation each play a role in determining myocardial ischemia-reperfusion injury. *Jpn Circ J* 2001; 65: 659-666.

39. Monsinjon T, Richard V, Fontaine M. Complement and its implications in cardiac ischemia/reperfusion: strategies to inhibit complement. *Fundam Clin Pharmacol* 2001; 15: 293-306.

40. Vakeva, A, Meri A. Complement activation and regulator expression after anoxic injury of human endothelial cells. *APMIS* 1998; 106: 1149-1156.

41. Verma S, Fedak PW, Weisel RD, Butany J, Rao AM, Li RK, Dhillon B, Yau TM. Fundamentals of reperfusion injury for the clinical cardiologist. *Circulation* 2002; 105: 2332-2336.

42. Fujita T, Ohi H, Endo M, Ohsawa I, Kanmatsuse K. Complement activation accelerates glomerular injury in diabetic rats. *Nephron* 1999; 81: 208-214.

43. Fujita T. Evolution of the lectin-complement pathway and its role in innate immunity. Nat. Rev. Immunol. 2002; 2: 346-353.

44. Fujita T, Matsushita M, Endo Y. The lectin-complement pathway--its role in innate immunity and evolution. *Immunol Rev* 2004; 198: 185-202.

45. Walport MJ. Complement. First of two parts. *N Engl J Med* 2001a; 344: 1058-1066.

46. Walport MJ. Complement. Second of two parts. *N Engl J Med* 2001b; 344: 1140-1144.

47. Lekowski R, Collard CD, Reenstra WR, Stahl GL. Ulex europaeus agglutinin II (UEA-II) is a novel, potent inhibitor of complement activation. *Protein Sci* 2001; 10: 277-284.

48. Collard CD, Lekowski R, Jordan JE, Agah A, Stahl GL. Complement activation following oxidative stress. *Mol Immunol* 1999; 36: 941-948.

49. Collard CD, Montalto MC, Reenstra WR, Buras JA, Stahl GL. Endothelial oxidative stress activates the lectin complement pathway: role of cytokeratin 1. *Am J Pathol* 2001; 159: 1045-1054.

50. Collard CD, Vakeva A, Bukusoglu C, Zund G, Sperati CJ, Colgan SP, Stahl GL. Re-oxygenation of hypoxic human umbilical vein endothelial cells (HUVECs) activates the classic complement pathway. *Circulation* 1997; 96: 326-333.

51. Jordan JE, Montalto MC, Stahl GL. Inhibition of mannose-binding lectin reduces postischemic myocardial reperfusion injury. *Circulation* 2001; 104: 1413-1418.

52. Vakeva AP, Agah A, Rollins SA, Matis LA, Li L, Stahl GL. Myocardial infarction and apoptosis after myocardial ischemia and reperfusion. *Circulation* 1998; 97: 2259-22672.

53. Pavlov VI, Skjoedt MO, Tan YS, Rosbjerg A, Garred P, Stahl GL. Endogenous and natural complement inhibitor attenuates myocardial injury and arterial thrombogenesis. Immunobiology. *Circulation* 2012; 126: 2227-2235.

54. Zhang M, Hou YJ, Cavusoglu E, Lee DC, Steffensen R, Yang L, Bashari D, Villamil J, Moussa M, Fernaine G, Jensenius JC, Marmur JD, Ko W, Shevde K. MASP-2 activation is involved in ischemia-related necrotic myocardial injury in humans. *In. J Cardiol* 2011; 166(2): 499-504.

55. Krijnen PA, Ciurana C, Cramer T, Hazes T, Meijer CJ, Visser CA, Niessen HW, Hack CE. IgM colocalises with complement and C reactive protein in infarcted human myocardium. *J Clin Pathol* 2005; 58: 382–388.

56. Nevens JR, Mallia AK, Wendt MW, Smith PK. Affinity chromatographic purification of immunoglobulin M antibodies utilizing immobilized mannan binding protein. *J Chromatogr* 1992; 597: 247-256.

57. Arnold JN, Wormald MR, Suter DM, Radcliffe CM, Harvey DJ, Dwek RA, Rudd PM, Sim RB. Human serum IgM glycosylation: identification of glycoforms that can bind to mannan-binding lectin. *J Biol Chem* 2005; 280: 29080-29087.

58. Collins C, Tsui FW, Shulman MJ. Differential activation of human and guinea pig complement by pentameric and hexameric IgM *Eur J Immunol* 2002; 32: 1802-1810.

59. Hughey CT, Brewer JW, Colosia AD, Rosse WF, Corley RB. Production of IgM hexamers by normal and autoimmune B cells: implications for the physiologic role of hexameric IgM. *J Immunol* 1998; 161: 4091-4097.

60. McMullen ME, Hart ML, Walsh MC, Buras J, Takahashi K, Stahl GL. Mannosebinding lectin binds IgM to activate the lectin complement pathway in vitro and in vivo. *Immunobiology* 2006; 211: 759-766.

61. Zhang M, Takahashi K, Alicot EM, Vorup-Jensen T, Kessler B, Thiel S, Jensenius JC, Ezekowitz RA, Moore FD, Carroll MC. Activation of the lectin pathway by natural IgM in a model of ischemia/reperfusion injury. *J Immunol* 2006; 177: 4727-4734.

62. Hill JH, Ward PA. The phlogistic role of C3 leukotactic fragments in myocardial infarcts in rats. *J Ex Med* 1971; 133: 885-900.

63. MacLean D, Fishbein MC, Braunwald E, Maroko PR. Long-term preservation of ischemic myocardium after experimental coronary artery occlusion. *J Clin Invest* 1978; 61: 541-551.

64. Maroko PR, Carpenter CB, Chiariello M, Fishbein MC, Radvany P, Knostman JD, Hale SL. Reduction by cobra venom factor of myocardial necrosis after coronary artery occlusion. *J Clin Invest* 1978; 61: 661-670.

65. Pinckard RN, O'Rourke RA, Crawford MH, Grover FS, McManus LM, Ghidoni JJ, Storrs SB, Olson MS. Complement localization and mediation of ischemic injury in baboon myocardium. *J Clin Invest* 1980; 66: 1050-1056.

66. Vogel CW, Fritzinger DC. Humanized cobra venom factor: experimental therapeutics for targeted complement activation and complement depletion. *Curr Pharm Des* 2007; 13: 2916-2926.

67. Gorsuch WB, Guikema BJ, Fritzinger DC, Vogel CW, Stahl GL. Humanized cobra venom factor decreases myocardial ischemia-reperfusion injury. *Mol Immunol* 2009; 47: 506-510.

68. Xiao F, Eppihimer MJ, Willis BH, Carden DL. Complement-mediated lung injury and neutrophil retention after intestinal ischemia–reperfusion. *J Appl Physiol* 1997; 82: 1459–1465.

69. Weisman HF, Bartow T, Leppo MK, Marsh HCJ, Carson GR, Concino MF, Boyle MP, Roux KH, Weisfeldt ML, Fearon DT. Soluble human complement receptor type 1: in vivo inhibitor of complement suppressing post-ischemic myocardial inflammation and necrosis. *Science* 1990; 249: 146-151.

70. Rioux P. TP-10 (AVANT Immunotherapeutics). *Current Opinions on Investigation Drugs* 2001; 2: 364-371.

71. Pratt JR, Hibbs MJ, Laver AJ, Smith RA, Sacks SH. Effects of complement inhibition with soluble complement receptor-1 on vascular injury and inflammation during renal allograft rejection in the rat. *Am J Pathol* 1996; 149: 2055-2066.

72. Mulligan MS, Yeh CG, Rudolph AR, Ward PA. Protective effects of soluble CR1 in complement- and neutrophil-mediated tissue injury. *J Immunol* 1992; 148: 1479-1485.

73. Kyriakides C, Wang Y, Austen WG, Favuzza J, Kobzik L, Moore FD, Hechtman HB. Moderation of skeletal muscle reperfusion injury by a sLe(x)- glycosylated complement inhibitory protein. *Am J Physiol Cell Physiol* 2001; 281: C224-C230.

74. Eror AT, Stojadinovic A, Starnes BW, Makrides SC, Tsokos GC, Shea-Donohue T. Anti-inflammatory effects of soluble complement receptor type 1 promote rapid recovery of ischemia/reperfusion injury in rat small intestine. *Clin Immunol* 1999; 90: 266–275.

75. Lazar HL, Bokesch PM, van LF, Fitzgerald C, Emmett C, Marsh HC, Ryan U. Soluble human complement receptor 1 limits ischemic damage in cardiac surgery patients at high risk requiring cardiopulmonary bypass. *Circulation* 2004; 110: II274-II279.

76. Li JS, Sanders SP, Perry AE, Stinnett SS, Jaggers J, Bokesch P, Reynolds L, Nassar R, Anderson PA. Pharmacokinetics and safety of TP10, soluble complement receptor 1, in infants undergoing cardiopulmonary bypass. *Am Heart J* 2004; 147: 173-180.

77. Cedzynski M, Madalinski K, Gregorek H, Swierzko AS, Nowicka E, Obtulowicz K, Dzierzanowska-Fangrat K, Wojda U, Rabczenko D, Kawakami M. Possible disease modifying factors: the mannan-binding lectin pathway and infections in hereditary angioedema of children and adults. *Arch Immunol Ther Exp* 2008; 56: 69-75.

78. Rossi V, Cseh S, Bally I, Thielens NM, Jensenius JC, Arlaud GJ. Substrate specificities of recombinant mannan-binding lectin-associated serine proteases-1 and-2. *J Biol Chem* 2001; 276: 40880-40887.

79. Rossi V, Teillet F, Thielens NM, Bally I, Arlaud GJ. Functional characterization of complement proteases C1s/mannan-binding lectin-associated serine protease-2 (MASP-2) chimeras reveals the higher C4 recognition efficacy of the MASP-2 complement control protein modules. *J Biol Chem* 2005; 280: 41811-41818.

80. Davis AE. Biological effects of C1 inhibitor. Drug News Perspect. 2004; 17: 439-446. Degn SE, Hansen AG, Steffensen R, Jacobsen C, Jensenius JC, Thiel S. MAp44, a human protein associated with pattern recognition molecules of the complement system and regulating the lectin pathway of complement activation. *J Immunol* 2009; 183: 7371-7378.

81. Buerke M, Murohara T, Lefer AM. Cardioprotective effects of a C1 esterase inhibitor in myocardial ischemia and reperfusion. *Circulation* 1995; 393-402.

82. Horstick G, Berg O, Heimann A, Gotze O, Loos M, Hafner G, Bierbach B, Petersen S, Bhakdi S, Darius H, Horstick M, Meyer J, Kempski O. Application of C1-esterase inhibitor during reperfusion of ischemic myocardium: dose-related beneficial versus detrimental effects. *Circulation* 2001a; 104: 3125-3131.

83. Horstick G, Heimann A, Gotze O, Hafner G, Berg O, Boehmer P, Becker P, Darius H, Rupprecht HJ, Loos M, Bhakdi S, Meyer J, Kempski O. Intracoronary application of C1 esterase inhibitor improves cardiac function and reduces myocardial necrosis in an experimental model of ischemia and reperfusion. *Circulation* 1997b; 95: 701-708.

84. Horstick G, Kempf T, Lauterbach M, Bhakdi S, Kopacz L, Heimann A, Malzahn M, Horstick M, Meyer J, Kempski O. C1-esterase-inhibitor treatment at early reperfusion of hemorrhagic shock reduces mesentry leukocyte adhesion and rolling. *Microcirculation* 2001; 8: 427-433.

85. Fattouch K, Bianco G, Speziale G, Sampognaro R, Lavalle C, Guccione F, Dioguardi P, Ruvolo G. Beneficial effects of C1 esterase inhibitor in ST-elevation myocardial infarction in patients who underwent surgical reperfusion: a randomized double-blind study. *Eur J Cardiothorac Surg* 2007; 32: 326-332.

86. Bauernschmitt R, Bohrer H, Hagl S. Rescue therapy with C1-esterase inhibitor concentrate after emergency coronary surgery for failed PTCA*. Intensive Care Med* 1998; 24: 635-638.

87. Gesuete R, Storini C, Fantin A, Stravalaci M, Zanier ER, Orsini F, Vietsch H., Mannesse, M.L., Ziere, B., Gobbi, M., De Simoni, M.G. Recombinant C1 inhibitor in brain ischemic injury. *Ann. Neurol* 2009; 66: 332-342.

88. Ricklin D, Lambris JD. Complement-targeted therapeutics. *Nat Biotechnol* 2007; 25: 1265-1275.

89. Woodruff TM, Nandakumar KS, Tedesco F. Inhibiting the C5-C5a receptor axis. Mol. Immunol. 48, 1631-1642. Xiao, F., Eppihimer, M.J., Willis, B.H., Carden, D.L., 1997. Complement-mediated lung injury and neutrophil retention after intestinal ischemia-reperfusion. *J Appl Physiol* 2011; 82: 1459-1465.

90. Armstrong PW, Mahaffey KW, Chang WC, Weaver WD, Hochman JS, Theroux P, Rollins S, Todaro TG, Granger CB, COMMA Investigators. Concerning the mechanism of pexelizumab's benefit in acute myocardial infarction. *Am Heart J* 2006; 151: 787-790.

91. Granger CB, Mahaffey KW, Weaver WD, Theroux P, Hochman JS, Filloon TG, Rollins S, Todaro TG, Nicolau JC, Ruzyllo W, Armstrong PW. Pexelizumab, an anti-C5 complement antibody, as adjunctive therapy to primary percutaneous coronary intervention in acute myocardial infarction: the Complement inhibition in Myocardial infarction treated with Angioplasty (COMMA) trial. *Circulation* 2003; 108: 1184-1190.

92. Theroux P, Armstrong PW, Mahaffey KW, Hochman JS, Rollins SA, Malloy KJ, Parish T, Nicolau NC, Lavoie J, Granger CB. Markers of inflammation predict mortality and are reduced by pexelizumab in patients with acute myocardial infarction: insights from the Complement Inhibition in Myocardial Infarction Treated with Angioplasty (COMMA) trial. *JACC* 2004; 43: 287A.

93. Shernan SK, Fitch JC, Nussmeier NA, Chen JC, Rollins SA, Mojcik CF, Malloy KJ, Todaro TG, Filloon T, Boyce SW, Gangahar DM, Goldberg M, Saidman LJ, Mangano DT. Impact of pexelizumab, an anti-C5 complement antibody, on total mortality and adverse cardiovascular outcomes in cardiac surgical patients undergoing cardiopulmonary bypass. *Ann Thorac Surg* 2004; 77: 942-949.

94. Smith PK, Shernan SK, Chen JC, Carrier M, Verrier ED, Adams PX, Todaro TG, Muhlbaier LH, Levy JH. Effects of C5 complement inhibitor pexelizumab on outcome in high-risk coronary artery bypass grafting: combined results from the PRIMO-CABG I and II trials. *J Thorac Cardiovasc Surg* 2011; 142: 89-98.

95. Skjoedt MO, Hummelshoj T, Palarasah Y, Honore C, Koch C, Skjodt K, Garred P. A novel MBL/ficolin associated protein is highly expressed in heart and skeletal muscle tissues and inhibits complement activation. *J Biol Chem* 2010; 285: 8234-8243.

96. Degn SE, Jensenius JC, Thiel S. Disease-causing mutations in genes of the complement system. *Am J Hum Genet* 2011; 88: 689-705.

97. Zhao H, Montalto MC, Pfeiffer KJ, Hao L, Stahl GL. Murine model of gastrointestinal ischemia associated with complement-dependent injury. *J Appl Physiol* 2002; 93: 338-345.

98. Harboe M, Garred P, Karlstrom E, Lindstad JK, Stahl GL, Mollnes TE. The down stream effects of mannan-induced lectin complement pathway activation depend quantitatively on alternative pathway amplification. *Mol Immunol* 2009; 47(2-3): 373-380.

99. Zhao H, Wakamiya N, Suzuki Y, Hamonko MT, Stahl GL. Identification of human mannose binding lectin recognition sites for novel inhibitory antibodies. *Hybridoma and Hybridomics* 2002; 21(1): 25-36.

100. Heitzeneder S, Seidel M, Forster-Waldl Heitger A. Mannon-binding lectin deficiency-Good news, bad news, doesn't matter? *Clin. Immunol* 2012; 143: 22-38.

101. Garred P. Mannose binding lectin genetics: from A to Z. *Biochem Soc Trans* 2008; 36: 1461.

102. Sheriff S. Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple helical coiled-coil. *Nat Struct Bio* 1994; 1(11): 789.

103. Hans-Joachim G. The sugar code: Fundamentals of glycosciences. Wiley-Blackwell 2009; Hoboken, NJ.

104. Nauta AJ, Raaschou-jensen N, Roos A, Daha MR, Madsen HQ, Borrias-Essers MC, Ryder LP, Koch C, Garred P. Mannose-binding lectin engagement with late apoptotic and necrotic cells. *Eur J Immunol* 2003; 33: 2853-2863.

105. Palaniyar N, Nadesalingam J, Clark H, Shis MJ, Dodds AW, Reid KB. Nucleic acid is a novel ligand for innate immune pattern recognition collectins surfactant proteins A and D and mannose-binding lectin. *J Biol Chem* 2004; 279: 32728-32736.

106. Hirano M, Ma BY, Kawasaki N, Okimura K, Baba M, Nakagawa T, Miwa K, Kawasaki N, Oka S, Kawasai T. Mannon-binding protein blocks the activation of metalloproteases meprin alpha and beta. *J Immunol* 2005; 175: 3177-3185.

107. Garred P, Honore C, Ma YJ. MBL2, FCN1, FCN2 and FCN3-The genes behind the initiation of the lectin pathway of complement. *Mol Immunol* 2009; 46: 2737.

108. Seyfarth J, Garred P, Madsen HO. Extra-hepatic transcription of the human mannose-bindinglectin (MBL2) and the MBL-associated serine protease 1-3 genes. *Mol Immunol* 2006; 43: 963-971.

109. Genecards. Weizmann Institute of science. 1996-2012. Accessed on 9/21/12 at www.genecards.org.

110. Steinberger P, Szekeres A, Wille S, Selenko N, Prager E, Staffer G, Madic O, Stockinger H, Knapp W. Identification of human CD93 as a phagocytic C1q receptor (C1qRp) by expression cloning. *J Leukoc Biol* 2002; 71: 133-140.

111. Whittock NV. Genomic organization and amplification of the human keratin 15 and keratin 19 genes. *Biochem Biophys Res Commun* 2000; 267: 462-465.

112. Ambrus G, Gal P, Kojima M, Szilagyi K, Balczer J, Antal J, Graf L, Laich A, Moffatt BE, Schwaeble W, Sim RB, Zavodszky P. Natural substrates and inhibitors of mannan binding lectin associated serine protease-1 and -2: a study on recombinant catalytic fragments. *J Immunol* 2003; 170: 1374-1382.

113. Gemma A, Hosoya Y, Seike M, Uematsu K, Kurimoto F, Hibino S, Yoshimura A, Shibuya M, Kudoh S, Emi M. Genomic structure of the human MAD2 gene and mutation analysis in human lung and breast cancers. *Lung Cancer* 2001; 23: 289-295.

114. Holaska JM, Black BE, Love DC, Hanover JA, Leszyk J, Paschal BM. Calreticulin is a receptor for nuclear export. *J Cell Bio* 2001; 152: 127-140.

115. Nam HJ, Poy F, Salto H, Frederick CA. Structural basis for the function and regulation of the receptor protein tyrosine phosphatase CD45. *J Exp Med* 2005; 201: 441-452.

116. Duncan RC, Bergstrom F, Coetzer TH, Blom AM, Wijeyewickrema LC, Pike RN. Multiple domains of MASP-2, an initiating complement protease, area required for interaction with its substrate C4. *Mol Immunol* 2012; 49(4): 593-600.

117. Verga MV, Segat L, Puppini B, Amoroso A, Crovella S. Evolution of the mannose-binding lectin gene in primates. *Genes Immun* 2004; 5(8): 653-661.

118. Biezeveld MH, Geissler J, Weverling GJ, Kuipers IM, Lam J, Ottenkamp J, Kuijpers TW. Polymorphisms in the mannose binding lectin gene as determinants of age-defined risk of coronary artery lesions in Kawasaki disease. *Arthritis Rheum* 2006; 54(1): 369-376.

119. Zimin AV, Deicher AL, Florea L, Kelley DR, Schatz MC, Pulu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marcias G, Roberts M, Subramanian P, Yorke JA, Saizberg SL. A whole genome assembly of the domestic cow, Bos taurus. *Genome Bio* 2009: 10(4): R42.

120. Vitved L, Holmskov U, Teisner B, Hansen S, Salomonsen J, Skjodt K. The homologue of mannose binding lectin the the carp family Cyprinidae is expressed at high level in spleen, and the deduced primary structure predicts affinity for galactose. *Immunogenetics* 2000; 51(11): 955-964.

121. La Bonte LR, Pavlov VI, Tan YS, Takahashi K, Takahashi M, Banda NK, Fujita T, Stahl GL. MBL-associated serine protease -1 (MASP-1) is a significant contributor to coagulation in a murine model of occlusive thrombosis. *Journal of Immunology* 2012; 188: 885-91.

122. Flicek P. Ensemble. *Nucleic Acids Res* 2012; 40: 84-90.

123. Holmskov U, Thiel S, Jensenius JC. Collectins and Ficolins: Humoral lectins of the innate immune defense. *Annu Rev Immunol* 2003; 21: 547-578.

124. Uemura K, Saka M, Nakagawa T, Kawasaki N, Thiel S, Jensenius JC, Kawasaki T. L-MBP is expressed in epithelial cells of mouse small intestine. *J Immunol* 2002; 169(12): 6945-6950.

125. Botto M, Kirschfink M, Macor P, Pickering MC, Wurzner R, Tedesco F. Complement in human diseases: Lessons from complement deficiencies. *Mol Immunol* 2009; 46: 2774-2783.

126. Allison DB, Cui P, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. *Nature Reviews Genetics* 2006; *7*(1): 55-65.

127. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* 2007; doi:10.1093.

128. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* 2005; 102 (43): 15545-15550.

129. Mootha VK, Lindgren CM, Erickson KF, Subramanian A, Sihag S, Lehar J, et al. PDC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet* 2003; 34 (3): 267-273.

130. Smyth GK, Michaud J, Scott H. The use of within-array replicate spots for assessing di_erential expression in microarray experiments. *Bioinformatics* 2005; 21(9): 2067{2075.

131. Mar JC, Matigan NA, Quackenbush J. Defining an informativeness metric for clustering gene expression data. *Bioinformatics* 2011; 27(8): 1094-1100.

132. Kaminiski N, Friedman N. Practical approaches to analyzing results of microarray experiments. Am. J. Respir. *Cell Mol Biol* 2002; 227: 125-132.

133. Evangelou M, Augusto R, Ouwehand WH, Wernisch L, Dudbridge F. Comparison of methods for competitive tests of pathway analysis. *PloS ONE* 2012; 7(7): e41018.

134. Tamayo P, Steinhardt G, Liberzon A, Mesirov JP. Gene set enrichment analysis made right. *arXiv preprint arXiv*:1110.4128 2011; 846.

135. Maruschke M, Reuter D, Koczan D, Hakenberg OW, Thiesen HJ. Gene expression analysis in clear cell renal cell carcinoma using gene set enrichment analysis for biostatistical management. *BJUI* 2011; 108: E29-E35.

136. Smyth GK. Limma: linear models for microarray data. In:*Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Springer, New York 2005; 397--420.

137. Smyth GK. Linear models and empirical Bayes methods for assessing dierential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* 2004; 3(1): Article 3.

138. Smyth GK, Speed TP. Normalization of cDNA microarray data. *Methods* 2003; 31: 265.

139. Ritchie ME, Silver J, Oshlack A, Silver J, Holmes M, Diyagama D, Holloway A, Smyth GK. A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 2007; 23: 2700-2707.

140. Ritchie ME, Diyagama D, Neilson van Laar RJ, Dobrovic A, Holloway A, Smyth GK. Empirical array quality weights in the analysis of microarray data. *BMC Bioinformatics* 2006; 7: 261.

141. Mar JC, Matigian NA, Mackay-Sim A, Mellick GD, Sue CM, Silburn PA, McGrath JJ. Variance of gene expression identifies altered network constraints in neurological disease. *PloS Genetics* 2011; 7(8): e1002207.

142. Babu M. M. Introduction to microarray data analysis**.** *Computational Genomics: Theory and Application* 2004; 225-249.

143. Haibe-Kains B, Olsen C, Djebbari A, Bontempi G, Correll M, Bouton C, Quackenbush J. Predictive networks: a flexible, open source, web application for integration and analysis of human gene networks. *Nucleic Acids Res* 2012; 40: 866-875.

144. Derry JMJ, Mangravite LM, Suver C, Furia MD, Henderson D, Schildwachter X, Bot B, Izant J, Sieberts S. K., Kellen, M. R., Friend, S. H. Developing predictive molecular maps of human disease through community based modeling. *Nat Genet* 2012; 44(2): 127-131.

145. Culhane AC, Schroder MS, Sultana R, Picard SC, Martinelli EN, Kelly C, Haibe-Kains B, Kapushesky M, St Pierre AA, Flahive W, Picard KC, Gusenleitner D, Papenhausen G, O'Connor N, Correll M, Quackenbush J. GeneSigDB: a manually curated database and resource for analysis of gene expression signatures. *Nucleic Acids Res* 2012; 40: 1060-1066.

146. Chittenden TW, Howe EA, Taylor JM, Mar JC, Aryee MJ, Gomez H, Sultana R, Braisted J, Nair SJ, Quackenbush J, Holmes C. nEASE: a method for gene ontology sub-classification of high-throughput gene expression data. *Bioinformatics* 2012; 28(5): 726-728.

147. Goh KI, Cusik ME, Valle D, Childs B, Vidal M, Barabas AL. The human disease network. *PNAS* 2007; 104(21): 8685-8690.

148. Perlman L, Gottlieb A, Atias N, Ruppin E, Sharan R. Combining drug and gene similarity measures for drug target elucidation. *J Comput Bio* 2011; 18(2): 1330145.

149. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ. Network-based elucidation of human disease similarities reveals common functional methods enriched for pluripotent drug targets. *PloS Comput Bio* 2010; 6(2): e1000662.

150. Morey JS, Ryan JC, Van Dolah FM. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. *Biological procedures online* 2006; *8*(1): 175-193.

151. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate. Normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology* 2002; *3*(7): 34.

152. Zhong X, Wang L, Wnag Y, Dong S, Leng X, Jia J, Zhao Y, Li H, Zhang X, Xu C, Wu L, Wang R, Lu F, Zhang W. Exogenouse hydrogen sulfide attenuates diabetic myocardial injury through cardiac mitochondrial protection. *Mol Cell Biochem* 2012; 371(1-2): 187-98.

153. Sergeev P, da Silva R, Lucchinetti E, Zaugg K, Pasch T, Schaub MC, Zaugg M. Trigger-dependent gene expression profiles in cardiac preconditioning: evidence for distinct genetic programs in ischemic and anesthetic preconditioning. *Anesthesiology* 2004; 100(3): 474-488.

154. Zeller T, Blankenberg S, Diemert P. Genomewide association studies in cardiovascular disease-An update 2011. *Clin Chem* 2012; 58(1): 92-103.

155. Gianfagna F, Cugino D, Santimone I, Jacoviello L. From candidate gene to genome-wide association studies in cardiovascular disease. *Thromb Res* 2012; 129: 320-324.

156. Zou C, La Bonte LR, Pavlov VI, Stahl GL. Murine hyperglycemic vasculopathy and cardiomyopathy: whole-genome gene expression analysis predicts cellular targets and regulatory networkds influenced by mannose binding lectin. Front. *Immunol* 2012; 3: doi: 10.3389/fimmu.2012.00015 .

157. Yu X, Murao K, Imachi H, Li J, Hosomi N, Masagata H, Zhang GX, Iwama H, Ishida T. Hyperglycemia suppresses ABCA1 expression in vascular smooth muscle cells. *Horm Metab Res* 2010; 42(4): 241-246.

158. Murao K, Imachi H, Cao WM, Chen K, Matsumoto K, Nishiuchi T, Wong NC, Ishida T. Hyperglycemia suppresses hepatic scavenger receptor clas B type I expression. *A J Physiol Emdocrinol Metab* 2008; 294(1): E78-87.

159. Mayrose I, Penn O, Erez E, Rubinstein ND, Shlomi T, Freund NT, Bublil EM, Ruppin E, Sharan R, Gershoni JM, Martz E. Pupko T. *Bioinformatics* 2007; 23(23): 3244-3246.

160. Irving MB, Pan O. Scott JK. Random peptide libraries and antigen fragment libraries for epitope mapping and the development of vaccines and diagnostics. *Curr Opin Chem Biol* 2001; 5: 314-324.

161. Barbas CF, Burton DR, Scott JK. Silverman GJ. *Phage display: a laboratory manual*. Cold Spring Laboratory Press 2001; Planview, NJ.

162. Denisova GF, Denisov DA, Ramson JL. Applying bioinformatics for antibody epitope prediction using affinity selected mimotopes – relevance for vaccine design. *Immunomol Res* 2010; 6.

163. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 2009; 23: 1494-1504.

164. Bhartiya D, Kapoor S, Jalali S, Sati S, Kaushik K, Sachindanandan S, Sivasubbu S, Scaria V. Conceptual approaches for lncRNA drug discovery and future strategies. *Expert Opin Drug Discov* 2012; 7: 503-513.

165. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell* 2011; 43(6): 904-914.

166. Gutschner T, Hammerle M, Eissman M, Hsu J, Kim Y, Hung G, Revenko A, Arun G, Stentrup M, Gross M, MacLeod, AR, Spector DL, Diederichs S. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* 2013; 73: 1180-1189.

167. Bernard D, Presanthi K, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourden L, Coulpier F, Triller A, Spector DL, Bessis A. A long nuclear-retained RNA regulates synaptogenesis by modulating gene expression. *EMBO J* 2010; 29: 3082-3093.

168. Ma H, Hao Y, Tian W. Molecular mechanisms and function prediction of long noncoding RNA. *Scientific World Journal* 2012; 541-786.

169. Kundrotas PL, Lensink MF, Alexov E. Protein-protein complexes using adjustments of modified sequences. *Int J Biol Macromol* 2008; 34: 2198-2208.

170. Prlic A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, Bourne PE. Pre-calculated protein structure alignments at the RCSB PDB website. *Structural Bioinformatics* 2010; 26: 2983-2985.