

Regulatory characterisation of the novel gene, myocyte stress 1.

**Thesis submitted for the degree of Doctor of Philosophy at the
University of Leicester**

by

**Samir Ounzain
Department of Cardiovascular Sciences
University of Leicester**

July 2008

Acknowledgements

I would like to thank my supervisor Dr Nelson Chong for giving me the opportunity to undertake this PhD in cardiovascular biology. His advice, guidance and encouragement throughout the duration of this project has been greatly appreciated. I am particularly grateful for the scientific freedom I was given which allowed me to develop my own idea's and discover the 'scientific bug'.

I would like to thank all the members of staff and students of the Division of Cardiology, specifically members of the Nilesh Samani, Alison Goodall and Nelson Chong laboratories. In particular Jo, Danny and Andrea, my fellow doctoral companions who ultimately became good friends. I will always have fond memories of our lunchtime chat's and other "scientific" discussions!!! I express sincere gratitude to Heart Research UK for providing the fund's to allow me to execute this project.

Finally, and most importantly, I offer sincere thanks to Mum, Dad and Nadia for all their love, support and encouragement throughout my life and academic career. They have always instilled me with the belief that I was capable of anything and without them I would not be where I am now and for that I am eternally grateful. I would also like to mention all of my auntie's, uncles and cousins both here in the UK and in Algeria. Special thanks go to my grandparents, to Nana who is still with us and my late Grandad, Bebasidi and Mamani, god bless you all.

Sincere thanks also go to Anna for her love, patience and support throughout the course of the last year. I know I have been a 'terror' !!! Thank you and love you lots!

I would like to dedicate the thesis to the memory of my late cousin, Nedjib. My thoughts are eternally with you and your family, god bless you.

Experimental Acknowledgements

In Chapter 4;

Whole tissue RNA extraction's were carried out by Dr Harin Mahadeva.

In Chapter 5;

GATA4 ChIP on adult feline myocytes was carried out by Dr Robert Paterson.

Adenoviral knockdown and over-expression of GATA4 in neonatal rat ventricular myocytes was carried out by Dr Saturo Kobayashi.

Type 1 and -2 diabetic model heart RNA and GATA4 TG heart RNA was provided by Dr Rong Liang and Dr Jeffrey Molkentin.

GATA4 gene-targeted murine allele's whole heart RNA was provided by Dr William Pu.

In Chapter 6;

MyoD and HDAC1/2 ChIP's were executed by Dr Caroline Dacwag.

TABLE OF CONTENTS

Acknowledgements	i
Experimental Acknowledgements	ii
Table of contents	iii
Abbreviations	X
List of Publications and Abstracts	XII
Abstract	XIV

CHAPTER 1

Introduction

1.1	General Introduction	1
1.2	Cardiovascular Disease	1
1.2.1	Epidemiology	1
1.3	Heart failure and left ventricular hypertrophy	2
1.3.1	Heart Failure	2
1.3.2	Left ventricular hypertrophy	2
1.3.3	Physiological hypertrophy	2
1.3.4	Classification of pathological LVH	3
1.3.5	Molecular basis of LVH	5
1.4	Signalling pathways in cardiac myocytes	6
1.4.1	G protein-coupled receptors	10
1.4.2	Small G proteins	11
1.4.3	Mitogen-activated protein kinases	12
1.4.3.1	Extracellular regulated kinases	14
1.4.3.2	<i>c-Jun</i> N-terminal kinases	14
1.4.3.3	p38 mitogen-activated protein kinases	15
1.4.4	Phosphoinositide 3-Kinase/Protein kinase B/Glycogen	

	synthase kinase-3	17
	1.4.5 Calcineurin	20
1.5	Transcription factors in cardiac hypertrophy	22
	1.5.1 GATA binding protein 4	22
	1.5.2 Myocyte enhancement factor 2	25
	1.5.3 Nuclear factor of activated T cells	28
	1.5.5 Smads	29
	1.5.6 Nkx2-5	31
	1.5.7 Hand Family	32
	1.5.10 Kruppel like factor family	32
	1.5.11 Serum response factor	33
	1.5.12 Myocardin family of SRF co-factors	35
	1.5.13 Role of the myocardin family in Rho signalling	38
1.6	Myocyte stress 1	39
	1.6.1 Initial identification and functional characterisation	39
	1.6.2 Role's in striated muscle development and disease	43
1.7	Regulation of <i>msl</i>	44
	1.7.2 Transcription	44
	1.7.2.1 Chromatin regulation	45
1.8	Aims	49

CHAPTER 2

Materials and methods

2.1	Materials	50
	2.1.1 Plasmids and reagents for molecular biology	50
	2.1.2 Reagents for Cell Culture	50
2.2	DNA Manipulation	51
	2.2.1 High Fidelity Polymerase Chain Reaction (PCR)	51
	2.2.2 Standard PCR	54
	2.2.3 Agarose gel electrophoresis	54

2.2.4	DNA extraction and purification from agarose gel	54
2.2.4.1	DNA determination	55
2.2.5	Ligation of isolated DNA into vector	55
2.2.6	Transformation of bacteria with plasmid vector	55
2.2.7	Restriction Enzyme Digest	56
2.2.8	Site Directed Mutagenesis	56
2.2.9	DNA Sequencing	57
2.3	Cell Culture	59
2.3.1	Maintenance of cell lines	59
2.3.2	Storage of cells	59
2.3.3	Cardiogenic and myogenic differentiation of cell lines	59
2.3.3.1	H9c2 cardiogenic differentiation	59
2.3.3.2	C2C12 myogenic differentiation	60
2.3.4	Trichostatin A treatment of H9c2 cells	60
2.3.5	Simulated sub-lethal ischemia/reperfusion	61
2.4	RNA manipulation	61
2.4.1	RNA isolation from cells and tissues	61
2.4.1.1	RNA determination	61
2.4.2	Deoxyribonuclease (DNaseI) treatment of RNA	62
2.4.3	cDNA synthesis	62
2.4.4	Semi-quantitative RT-PCR	62
2.4.4.1	Statistical analysis	65
2.4.5	Real-time quantitative RT-PCR using TaqMan® gene expression assays	65
2.4.6	Real-time quantitative RT-PCR using SYBR® Green and analysed using the Pfaffl method	66
2.4.7	Statistical analysis	67
2.5	Transient transfection analysis	69
2.5.1	Transient transfection of cell lines	69
2.5.2	Luciferase reporter gene assay	69
2.6	Electromobility shift assay (EMSA)	72
2.6.1	Preparation of whole cell protein extracts	72

2.6.2	Protein quantification	72
2.6.3	Preparation of double stranded probes for use in EMSA	72
2.6.4	EMSA	73
2.7	Chromatin immunoprecipitation (ChIP)	75
2.7.1	ChIP assay	75
2.8	Comparative DNA sequence analysis	78

CHAPTER 3

Bioinformatic analysis of the *msl* promoter and associated regulatory domains

3.1	Introduction	79
3.2	Results	85
3.2.1	Comparative sequence analysis of the rat <i>msl</i> genomic interval	85
3.2.2	Identification of putative cis regulatory modules (PCRM) within the 5' conserved flanking sequence	89
3.2.3	Identification of conserved TFBS within annotated PCRM utilising rVISTA	95
3.2.4	MatInspector based interrogation of the isolated rat PCRM	101
3.2.5	Manual filtering and integration of MatInspector and rVISTA predicted TFBS	102
3.3	Discussion	118

CHAPTER 4

Cardiac specific functional and epigenetic analysis of the *msl* conserved *cis* regulatory domains

4.1	Introduction	126
4.2	Results	129
4.2.1	Quantitative analysis of cardiac specific expression of <i>msl</i> <i>in vivo</i> and <i>in vitro</i>	129
4.2.2	Isolation and functional cardiac characterisation of proximal 5'-flanking regulatory domains	133
4.2.3	Cardiac specific regulatory activity is associated with the distal regulatory domains, UP2 and UP3	140
4.2.4	Cardiogenic differentiation-dependent activation of <i>msl</i> expression: Potential role for the proximal promoter domain	144
4.2.5	Epigenetic regulatory processes contribute to <i>MS1</i> cardiac specific expression	150
4.2.6	The UP2 distal domain as a potential calcineurin sensitive stress dependant regulatory enhancer	157
.1	Discussion	164

CHAPTER 5

GATA4 modulates cardiac specific expression of *msl* through distinctive mechanisms at the proximal promoter and distal UP3 enhancer domain

5.1	Introduction	172
5.2	Results	175
5.2.1	GATA4 targets the <i>msl</i> proximal promoter domain	175
5.2.2	<i>In vivo</i> binding of GATA4 at the proximal -127/+60 interval	180
5.2.3	GATA4 targeting at the UP3 domain is required for enhancer activity	184
5.2.4	GATA4 modulates endogenous <i>msl</i> expression <i>in vitro</i>	190

5.2.5	GATA4 modulates endogenous <i>msl</i> expression in embryonic and adult murine hearts <i>in vivo</i>	195
5.2.6	GATA4 associated pathological phenotypes correlate with <i>msl</i> transcriptional dysregulation	200
5.3	Discussion	203

CHAPTER 6

Regulatory characterisation of the transcriptional mechanisms governing skeletal muscle myocyte stress 1 expression

6.1	Introduction	208
6.2	Results	211
6.2.1	Quantitative analysis of relative <i>msl</i> mRNA in adult cardiac and skeletal muscle	211
6.2.2	Analysis of <i>msl</i> mRNA during myogenic differentiation	214
6.2.3	Identification and myogenic specificity of the <i>msl</i> promoter	218
6.2.4	Identification of myogenic factors that can module the <i>msl</i> promoter	221
6.2.5	Site directed mutagenesis of the <i>msl</i> promoter	224
6.2.6	<i>In vitro</i> binding of myogenic proteins to E1 and E2	227
6.2.7	Direct binding of MyoD to the endogenous UP1 and PP domains during myogenic differentiation	231
6.2.8	Temporal dynamics of Histone Deacetylase (HDAC) enzyme recruitment to the endogenous UP1 and PP domains during myogenic differentiation	235
6.3	Discussion	240

CHAPTER 7	
General discussion	247
Appendix	255
REFERENCES	263

Abbreviations

ABLIM	actin-binding LIM
ABRA	actin binding Rho activator
ANF	atrial natriuretic factor
ANG-II	angiotensin 2
ATRA	all-trans retinoic acid
BMP	bone morphogenic protein
BNP	b-type natriuretic peptide
BSA	bovine serum albumin
CAMK	calmodulin dependant kinase
CBP	CREB binding protein
CDM	cardiogenic differentiation media
CHD	coronary heart disease
ChIP	chromatin immunoprecipitation
CLB	cell lysis buffer
CNS	conserved non-coding sequence
CRM	cis regulatory module
CTGF	connective tissue growth factor
CVD	cardiovascular disease
DIG	digoxigen
DMEM	dulbecco's modified eagle medium
DMSO	dimethyl sulfoxide
EDTA	ethylenediaminetetraacetic acid
EMSA	electromobility shift assay
ERK	extracellular regulated kinase
ET-1	endothelin 1
GM	growth media
GPCR	g-protein coupled receptor
GRN	gene regulatory network
GSK-3beta	glycogen synthase 3 beta

HAT	histone acetyltransferase
HDAC	histone de-acetylase
HG	hyperglycemia
HPLC	high performance liquid chromatography
IP	immunoprecipitate
JNK	c-jun n-terminal kinase
KLF	kruppel like factors
LB	luria broth
LVH	left ventricular hypertrophy
MAPK	mitogen activated protein kinase
MEF2	myocyte enhancer factor 2
MRF	myogenic regulatory factor
MRTF	myocardin related transcription factors
MS1	myocyte stress 1
NFAT	nuclear factor of activated T-cells
NRVM	neo-natal rat ventricular myocytes
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PCRM	putative cis regulatory module
PI3-K	phosphoinositide 3-OH kinase
PP2B	protein phosphatase 2B
PWM	position weight matrix
REST	neuron restrictive silencing factor
ROCK	rho-associated kinase
RT-PCR	reverse transcription polymerase chain reaction
SAPK	stress activated protein kinases
SDM	site directed mutagenesis
SHR	spontaneously hypertensive
SRF	serum response factor
STARS	striated muscle specific activator of Rho signalling
TAD	transcriptional activation domain
TBE	tris boric acid EDTA

TF	transcription factor
TFBS	transcription factor binding site
TG	transgenic
TGF- β	transforming growth factor beta
TSA	trichostatin A
TSS	transcription start site
UPS	ubiquitin proteasome system
WT	wild type

List of Publications and Abstracts

Publications

Ounzain S, Dacwag CD, Samani NJ, Imbalzano AN, Chong NW. Comparative *in silico* analysis identifies bona fide MyoD binding sites within the Myocyte Stress1 gene promoter. BMC Molecular Biology, 2008 May, 19;9:50

Abstracts

Ounzain S, Peterson R, Menick DR, Samani NJ, Chong NW . Expression of Myocyte Stress 1, a novel gene involved in cardiac development and hypertrophy is regulated by evolutionary conserved GATA motifs. British Cardiovascular Society Annual Congress 2007, Glasgow, Scotland. Heart Journal, 2007;Vol 93 (Abstract Supplement 1)

Ounzain S, Peterson R, Menick DR, Samani NJ, Chong NW. Myocyte Stress 1 (MS1): a novel nexus that integrates the SRF and GATA4 cardiac gene regulatory networks. 7th Federation of European Biochemical Societies Young Scientist Forum 2007, Vienna, Austria.

Ounzain S, Peterson R, Menick DR, Samani NJ, Chong NW . Expression of Myocyte Stress 1, a novel gene involved in cardiac development and hypertrophy is regulated by evolutionary conserved GATA motifs. European Society of Cardiology Annual Congress 2007, Vienna, Austria.

S Ounzain

University of Leicester

Regulatory characterisation of the novel gene, myocyte stress 1.

Abstract

Myocyte stress 1 (*msl*) is a striated muscle actin binding protein required for muscle specific activity of the myocardin related transcription factor (MRTF)/serum response factor (SRF) transcriptional pathway. Previous work in our group demonstrated that cardiac *msl* is transiently up-regulated after pressure overload suggesting a possible role in the initial signalling of the hypertrophic response. Subsequent studies have supported this and demonstrated that *msl* plays an important role in cardiac development and physiology. To date, little is known about the molecular mechanisms that govern striated muscle specific expression of *msl*. In order to delineate *msl* regulation and function, a strategy of comparative *in silico* analysis coupled with experimental characterisation was used. *In silico* analysis identified four genomic intervals of potential regulatory function designated PP, UP1, UP2 and UP3. Using *in vitro* and *in vivo* approaches, important cardiac regulatory roles for these domains were defined. The PP domain represents the basal promoter and is required for all regulatory contexts. This domain serves to integrate context specific regulatory signals from the distal UP2 and UP3 domains. Within the heart the cardiac transcription factor GATA4, and the calcineurin signalling pathway confer cardiac regulatory function on the PP, UP2 and UP3 domains. Within skeletal muscle, MyoD binding sites within the PP and UP1 domain were identified, which mediate temporal induction of *msl* during myogenesis. Both cardiac and skeletal regulatory processes were dependent on epigenetic phenomena with histone acetylation being a major determinant for *msl* expression. Collectively, these findings demonstrate that *msl* transcriptional regulation is mediated by the complex interplay of context specific regulatory domains and binding factors. Therefore through *msl*, important striated muscle gene regulatory networks (GRNs) (GATA4, Mef2 and MyoD GRNs) can integrate with SRF, thus exquisitely controlling biological processes in muscle. It is proposed that dysregulation of *msl* expression may result in pathological phenotypes. Therefore, the insights obtained here may allow for the therapeutic manipulation of *msl* expression in pathological settings and potentially lead to effective palliation of such phenotypes.

Chapter 1

Introduction

1.1 General Introduction

“ the heart of creatures is the foundation of life, the prince of all, the sun of their microcosm, from where vigour and strength does flow.

- William Harvey, *De Motu Cordis*, 1628”

The heart is arguably the most widely studied organ in physiology and medicine and it was in the influential *De Motu Cordis* (On the Motion of the Heart) by William Harvey that the basic principles of cardiac development and function were delineated. Since Harvey’s time, physiologists have further documented the workings of the heart in intricate detail, however, it is only over the last ten to fifteen years that a transition towards a better understanding of cardiac function (and dysfunction) at the molecular and genetic level has taken place (Olson, 2004). This transition has and will continue to provide deep mechanistic insights into heart development and acquired adult disease and has subsequently fuelled new opportunities for therapeutic palliation and prevention of cardiac disease.

1.2 Cardiovascular Disease

1.2.1 Epidemiology

Despite this massive progress in our present understanding of cardiac function, cardiovascular disease (CVD) remains the number one cause of death in the developed world. Diseases of the heart and circulatory system are currently the main causes of death in the United Kingdom (UK). In 2002 CVD was responsible for 238,000 UK deaths, with more than one in three people (39%) dying as a result of CVD (Peterson et al 2004). The primary forms of CVD are coronary heart disease (CHD) and stroke with CHD representing half of all deaths from CVD . Premature death, which is defined in

the UK as death before the age of 75, is also caused primarily by CVD with 67,000 cases recorded in 2002 (BHF Annual Statistics, 2006).

1.3 Heart Failure and Left Ventricular Hypertrophy

1.3.1 Heart Failure

Heart failure, defined as the inability of the heart to pump sufficient blood to meet the body's metabolic demands, is the most common endpoint of CVD. It represents a clinical syndrome characterised by shortness of breath and fatigue at rest or with exertion owing to structural and functional abnormalities of the heart resulting in dyspnoea and/or oedema. Heart failure is divided into two broad categories: systolic heart failure (impaired ventricular contraction) and diastolic heart failure (impaired ventricular relaxation). It is typically induced by a number of common disease stimuli including myocardial infarction or ischemia as a consequence of CHD, long standing hypertension, myocarditis due to an infectious agent, stenosis and valvular insufficiency; congenital malformations, familial and hypertrophic dilated cardiomyopathies, and diabetic cardiomyopathy (Klein & Gheorghade, 2003; Lips et al, 2003).

1.3.2 Left ventricular hypertrophy

At the cellular level of the cardiomyocyte, the disease stimuli described in section 1.3.1 primarily induce a phase of ventricular hypertrophy, specifically within the left ventricle (LVH), in which individual myocytes grow, thereby increasing cardiac pump function and decreasing wall stress. This phase is classically referred to as 'compensated hypertrophy' (Haider et al, 1998; Rothermel et al, 2005). However, in the long term there is a maladaptive transition from this 'compensated' state to a 'decompensated' phenotype which predisposes individuals to heart failure, arrhythmia and sudden cardiac death (Rothermel et al, 2005).

1.3.3 Physiological hypertrophy

Not all hypertrophy of the heart is bad ‘per se’. In broad terms there are two types of cardiac hypertrophy, physiological and pathological which are summarised in Figure.1 below (Olson & Schneider, 2003). Physiological hypertrophy of the heart occurs within three contexts: during normal post-natal development of the myocardium (also known as eutrophy) which is the period in which the neo-natal myocyte has withdrawn from the cell cycle, during female pregnancy and finally in response to excessive physical exercise (Chien & Olson, 2002). During post-natal development, the human heart grows through hypertrophy in proportion to the size of the body. Typically the left ventricle has a mass in grams which is three to four times the body mass in kilograms. Therefore as a young child grows you see a 10-fold increase in the left-ventricular mass with this physiological hypertrophy both essential and beneficial (Lorell & Carabello, 2000).

During pregnancy the increased cardiac demand (increase stroke volume and cardiac output) placed upon the mother by the embryo leads to the mother developing a compensatory increase in left ventricular mass. This, however, regresses during the post partum period (Lorell & Carabello, 2000). Elite trained athletes who specialise in sports such as weight lifting and wrestling display a specific concentric (1.3.4) hypertrophic response while those who specialise in long distance running or cycling display eccentric hypertrophy (1.3.4) (Lorell & Carabello, 2000). Again this hypertrophy is in response to increased demand on cardiac output.

1.3.4 Classification of pathological LVH

The most proximal initiating stimuli for pathological LVH can be classified into stretch-sensitive/biomechanical mechanisms, or neurohumoral mechanisms that are associated with the release of numerous peptide growth factors, chemokines, cytokines and hormones.

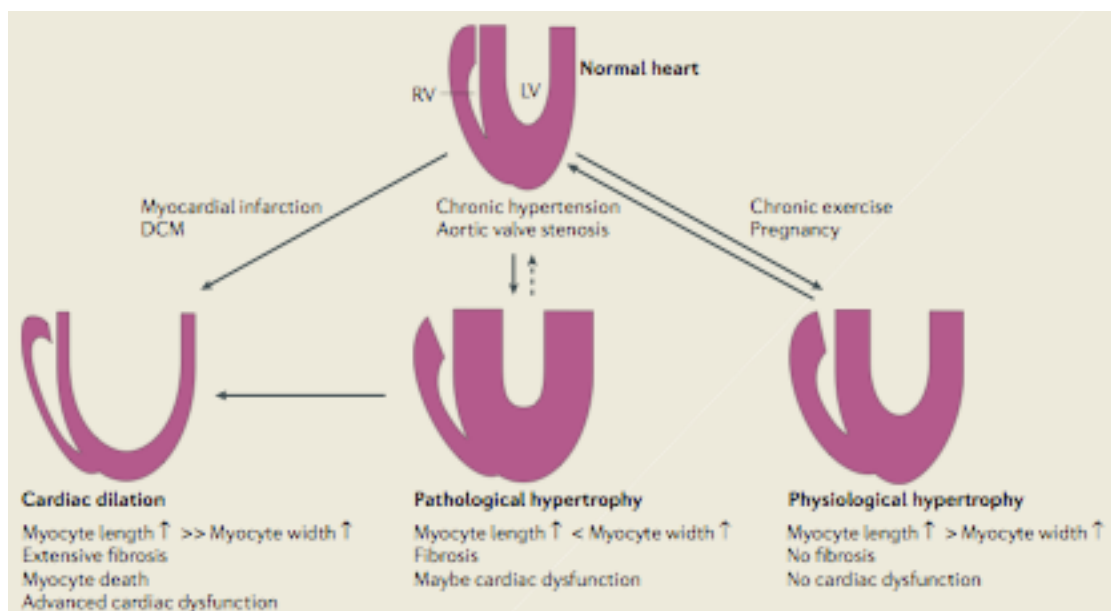


Figure 1. Different forms of cardiac hypertrophy and their associated characteristics. DCM, dilated cardiomyopathy; LV, left ventricle; RV, right ventricle (Molkentin et al, 2006)

The primary factor which mediates the activation of these diverse stimuli is an increase in hemodynamic load imposed upon the cardiomyocytes of the left ventricle. Within the '*in vivo*' whole heart context, cardiomyocytes enter hypertrophy to compensate and accommodate this extra cardiac workload. As a consequence of the ventricular cardiomyocytes withdrawing from the cell cycle in the peri-natal period the only plausible mechanism by which the heart can increase cardiac output is through promoting myocyte hypertrophy (increase in cell size) rather than increasing cell number (hyperplasia).

The increase in haemodynamic load which induces the LVH phenotype can either take the form of pressure overload or volume overload, with each form inducing a specific characteristic change in myocyte morphology (Gerdes, 1992). Pressure overload leads to a relative thickening of the ventricular wall with the myocytes laying down sarcomeres in a favourably parallel orientation thereby increasing overall myocyte cross sectional area. On the contrary to this, volume overload stimulates a relative thickening of all the myocytes with sarcomeres being organised in both a parallel and series orientation. These two types of remodelling events are typically referred to as concentric and eccentric remodelling respectively (Gerdes, 1992).

1.3.5 Molecular basis of LVH

The major cellular hallmark of LVH is an increase in cardiomyocyte size and mass which is accompanied by increased protein synthesis and a re-organisation of the sarcomere and associated contractile apparatus. These changes in cellular phenotype are coupled to a complex process of transcriptional re-programming. This genetic re-programming is ultimately responsible for the gene expression changes that underpin the development of LVH and also the subsequent progression from a compensated to a de-compensated phenotype, with the endpoint being congestive heart failure (Lips et al, 2003).

Gene expression changes include increased levels of paracrine/autocrine mediators such as transforming growth factor-beta (TGF- β), connective tissue growth factor (CTGF),

angiotensin II (AngII) and endothelin 1 (ET-1). In addition one observes transcriptional re-induction of fetal isoforms of contractile proteins [cardiac muscle alpha actin and myosin light chain 2 ventricular (mlc2v)) (Chien et al, 1991; Sugden & Clerk, 1998a), re-expression of stress markers including atrial natriuretic factor (ANF) and B-type natriuretic peptide (BNP) that are only otherwise expressed in fetal development and changes in expression of key regulatory proteins that modulate intracellular ion homeostasis. Specific changes include the down regulation of sarcoplasmic reticulum calcium ATPase (SERCA2a) and para-sympathetic and sympathetic receptors. In addition you observe the variable up-regulation of the $\text{Na}^{2+}/\text{Ca}^{2+}$ exchanger and down-regulation of Kv4.3 which causes a pro-longed action potential and therefore subsequent reduction in the transient of outward current (I_o) density (Lorell & Carabello, 2000). In summary, the hypertrophic response, which is induced by either an acute or chronic insult to the heart, is characterised by substantial changes in myocyte size, fetal gene expression, sarcomeric re-organisation and altered Ca^{2+} handling with the combined effect of these changes ultimately resulting in contractile dysfunction and heart failure (summarised in Figures 2 and 3).

1.4 *Signalling pathways in cardiac myocytes*

The initiating stimuli which stimulate cardiomyocyte hypertrophy converge on a finite array of intracellular signal transduction pathways which subsequently propagate the signal and modulate gene expression and the growth response. These intracellular signalling pathways function through two principal mechanisms: (1) Protein kinase cascades in which the signal is transmitted across the cytoplasm, resulting in the trans-activation dependent phosphorylation of nuclear localised transcription factors (Clerk et al, 2007); (2) Phosphorylation and dephosphorylation of cytoplasmic transcription factors which then trans-locate to the nucleus where they can modulate downstream gene activation. The main signalling pathways implicated in cardiomyocyte hypertrophy are described below (Figure 4) with the G protein coupled receptors (GPCRs), small G-proteins, mitogen activated protein kinases (MAPKs), PI3K, PKB, glycogen synthase kinase-3 and calcineurin being considered the most important proteins.

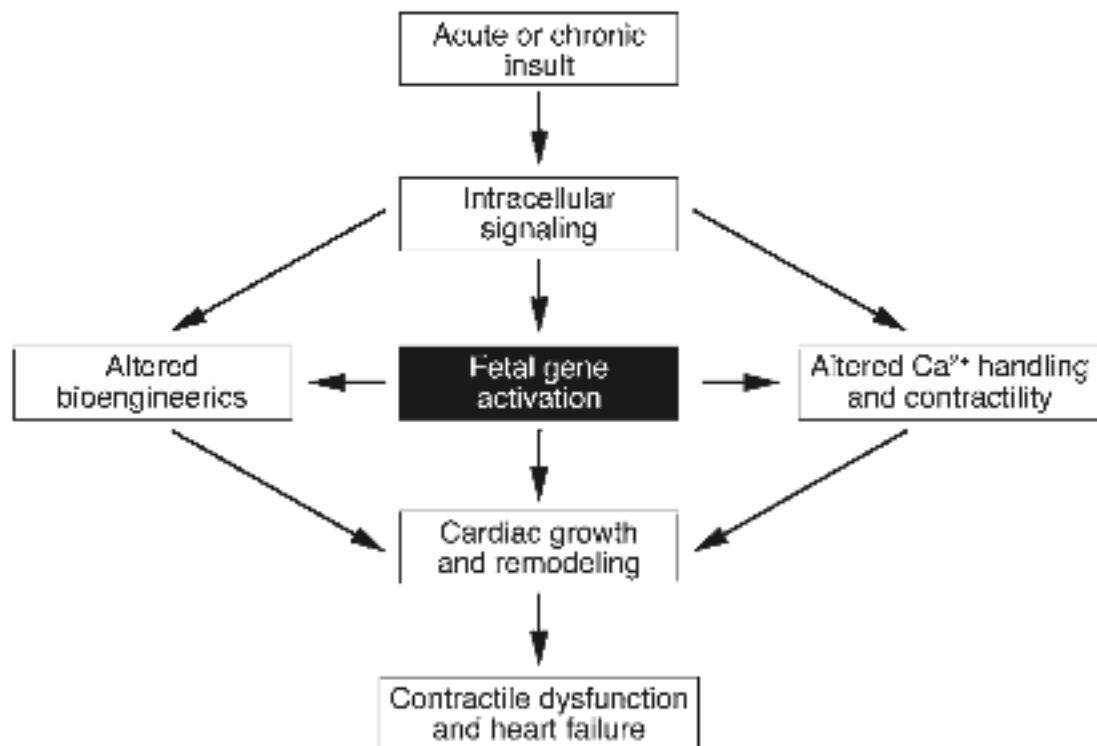


Figure 2. Summary of the abnormalities associated with the development of pathological hypertrophy and heart failure (Olson, 2005)

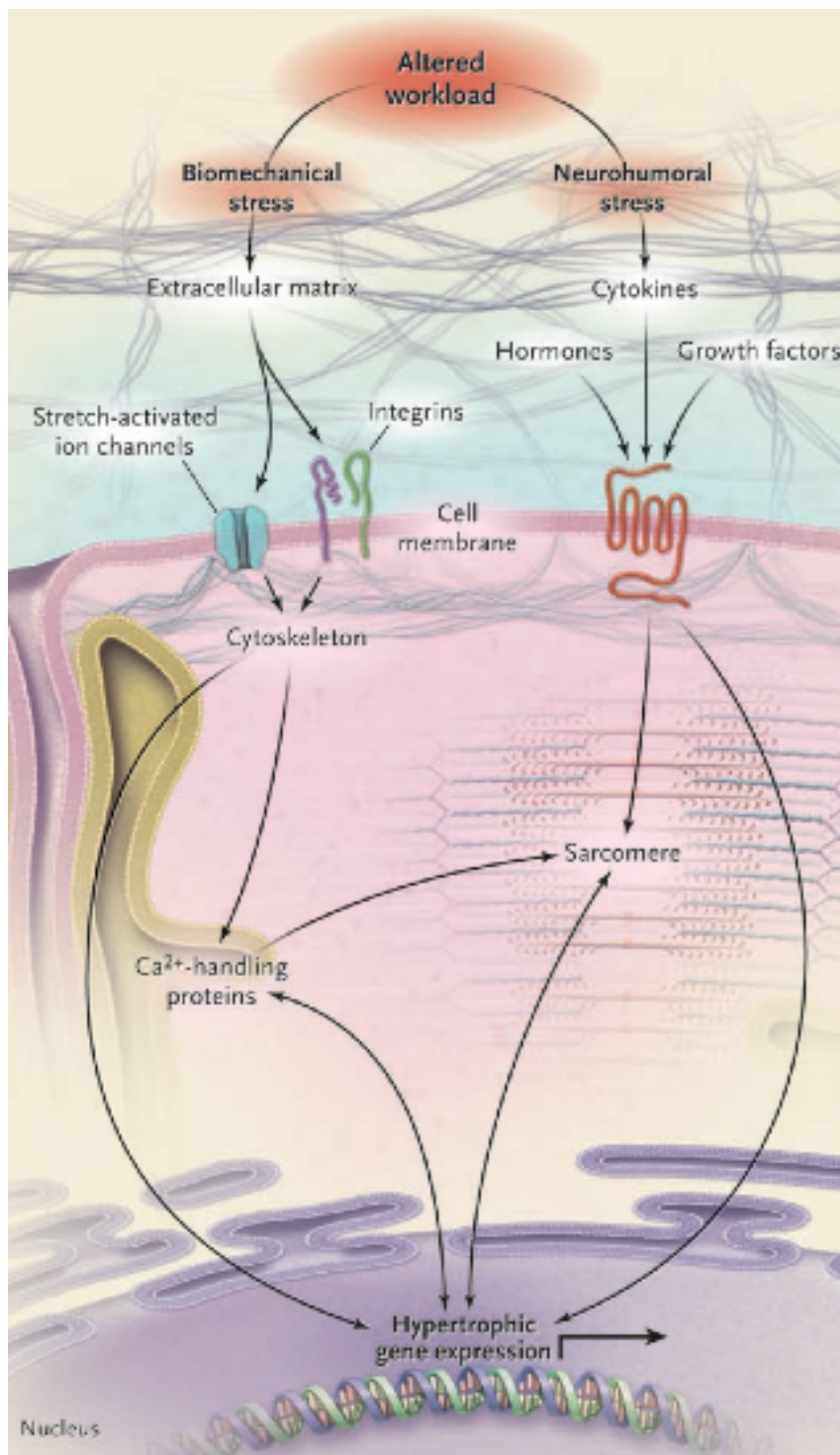
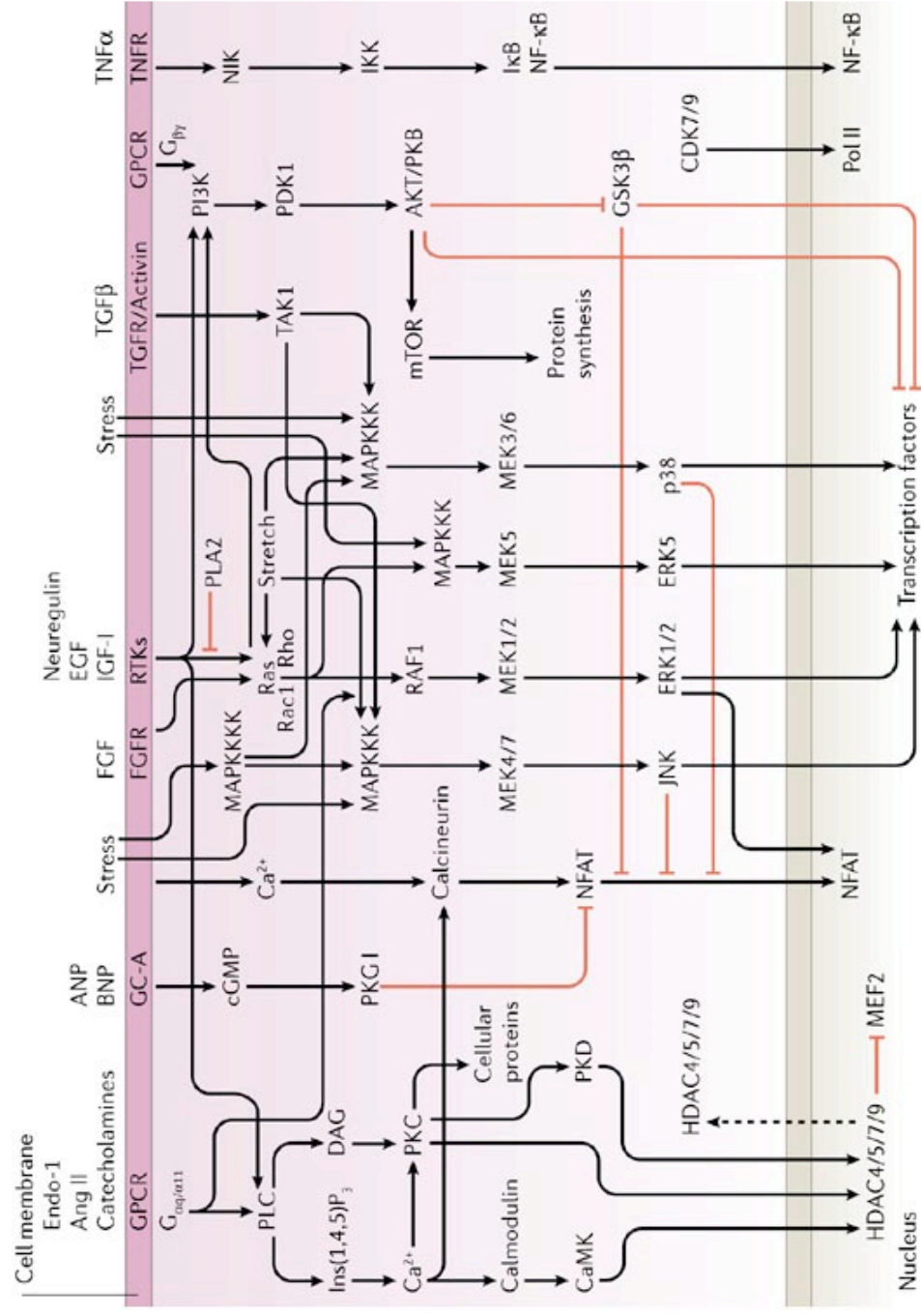


Figure 3. Summary of the cellular and molecular events triggered by alterations in cardiac workload (Olson, 2008).



Copyright © 2006 Nature Publishing Group
Nature Reviews | [Molecular Cell Biology](#)

Figure 4. Integrated schematic of signalling pathways implicated in the cardiac hypertrophic response (Molkentin et al, 2006).

1.4.1 G protein-coupled receptors

The three main functional classes of G protein-coupled receptors (GPCRs) are coupled to their cognate class of G proteins within the cardiac myocyte. The G proteins consist of individual $G\alpha$ and $G\beta\gamma$ subunits which are classified into subfamilies according to the composition of their alpha subunits, for example $G_{\alpha s}$, $G_{\alpha i}$ and $G_{\alpha q}$. The G protein subfamilies couple to specific agonist sensitive receptors: alpha-adrenergic, endothelin and angiotensin II receptors couple to $G_{\alpha q}/G_{\alpha 11}$ and have been extensively investigated and implicated in cardiomyocyte signal transduction and hypertrophy (reviewed by (Molkentin & Dorn, 2001)). Conditional cardiac specific knockdown of $G_{\alpha q}/G_{\alpha 11}$ is able to blunt the hypertrophic response in an adult animal model of pressure overload induced LVH (Frey et al, 2004). Coupled to this, constitutive cardiac specific over-expression of a dominant negative $G_{\alpha q}$ protein in adult mice was able to blunt pressure overload hypertrophy suggesting both $G_{\alpha q}/G_{\alpha 11}$ are important mediators of the hypertrophic response. Other receptors include the $G_{\alpha s}$ coupled beta-adrenergic receptors which control heart rate and cardiac contractility in response to epinephrine and norepinephrine stimulation. Conversely, the neurotransmitter acetylcholine binds cholinergic receptors which are coupled to $G_{\alpha i}$.

1.4.2 Small G proteins

Small G proteins provide a critical link between cell membrane receptors and various intracellular signalling pathways. The small G proteins comprise of five subfamilies (Ras, Rho, ADP ribosylation factors, Rab and Ran) with each family containing multiple members. These proteins play a regulatory role in modulating sarcomeric and cytoskeletal organisation, with the Ras and Rho subfamilies specifically implicated with these processes in the hypertrophic phenotype (Clerk & Sugden, 2000). Constitutively active Ras causes a significant increase in cardiac mass when over-expressed within the myocardium of adult mice with a similar phenotype mirrored in neonatal rat ventricular myocytes. The Rho family, which includes RhoA, Rac and cdc42 subfamilies, has been well characterised with respect to a hypertrophic regulatory role. Increased Rac activity *in vitro* and *in vivo* has been demonstrated to promote morphological changes associated with myocyte hypertrophy and corresponding changes in gene expression, specifically the hypertrophic markers ANF and (Clerk & Sugden, 2000; Frey et al, 2004).

Rho itself regulates numerous cytoskeletal-dependent cell functions including focal adhesion formation, myosin-based contractility, f-actin bundling and monomeric g-actin polymerisation (Yanazume et al, 2002). Within the heart Rho can stimulate the expression of the hypertrophic marker ANF and also induce myofibrillar organisation (Yanazume et al, 2002). This is as a consequence of its activation of numerous downstream protein kinases with the Rho-associated kinase (ROCK) the best characterised in the hypertrophic context (Yanazume et al, 2002). Arguably the most important downstream effector of RhoA signalling during cardiac hypertrophy is serum response factor (SRF). SRF is a nuclear transcription factor implicated as having a major role in cardiac hypertrophy (Hill et al, 1995). Although the precise mechanisms are not clear Rho appears to directly activate SRF trans-activity through a mechanism dependent on changes in actin treadmilling and abundance of various actin forms (Hill et al, 1995). In summary, the small G proteins, and specifically Rho, Rac and Ras all appear to play a role in cardiac hypertrophy.

1.4.3 Mitogen-activated protein kinases

The mitogen-activated protein kinase's (MAPKs) are well established transducers of growth and stress responses in many cell types and in particular are important modulators of the cardiac hypertrophic response (Sugden & Clerk, 1998). The MAP kinases are highly regulated protein kinases that require dual phosphorylation of their T (E/P/G)Y motif in the kinase domain to become catalytically active. On the basis of sequence homology they are classified into three major MAP kinase subfamilies, the extracellular regulated kinases (ERKs), c-Jun N-terminal kinases (JNKs) and p38 MAPKs with each of these families having a clearly defined activation cascade mediated by specific upstream MAP kinase kinase kinases (MAPK-KKs) and MAP kinase kinases (MAPK-Ks) (Figure 5). Classical hypertrophic agonists such as phenylephrine and endothelin-1 potently activate the ERKs (Sugden & Clerk, 1998), whereas the stress activated kinases (JNKs and p38) are activated more potently by common cardiac stresses including ischemia or cytotoxic agents (Sugden & Clerk, 1998). Interestingly, over-expression of MAPK phosphatase 1 (MKP-1), which inhibits all three major branches of MAPK signalling, is able to block agonist-induced hypertrophy *in vitro* in addition to pressure-overload stimulated hypertrophy *in vivo* (Clerk et al, 2007). These findings demonstrate a clear dependence on the MAPK signalling pathways in the development of cardiomyocyte hypertrophy.

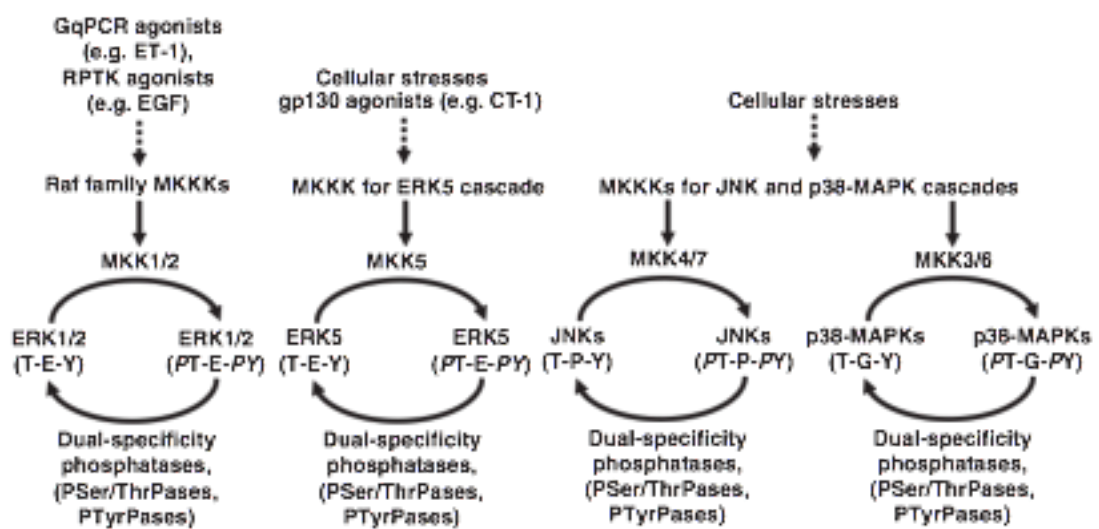


Figure 5. Summary of the mitogen-activated protein kinase (MAPK) cascades (Clerk A et al, 2007)

1.4.3.1 Extracellular regulated kinases

There are five ERK proteins identified to date designated ERKs 1-5. ERK1 and -2 (also known as p44 and p42) are the best characterised members of the family and are directly regulated by the MAPK kinases (MAPKKs), MEK1 and MEK2 (also known as MKK1 and MKK2 respectively). These are activated by multiple upstream MAPKKs which transduce stress signals directly or are coupled to regulatory effectors including the small G proteins (Ras, Rac, Rho, cdc42).

Work by Sugden and co-workers (Clerk et al, 2007) have demonstrated that ERKs 1/2 are acutely activated in a transient manner by numerous stimuli, including ET-1 and phenylephrine, in isolated cardiomyocyte preparations. They also reported that antisense mediated depletion of ERK1/2 or pharmacological inhibition of MEK1/2 attenuated the agonist stimulated hypertrophic response. Others have taken this further and shown that transgenic over-expression of MEK1, which specifically activates ERK1/2, drives the cardiac hypertrophy process (Clerk et al, 2007). GATA4 is a critical regulator of cardiac specific gene expression both during development and within the post natal myocardium. Beuno and Molkentin (2002)(Bueno & Molkentin, 2002) have shown that the phosphorylation and activation of the cardiac enriched transcription factor GATA4 is an ERK1/2 dependent process, thus further supporting a critical role for ERK1/2 in cardiac hypertrophy. In addition to GATA4, ERK1/2 can also phosphorylate and activate the ternary complex factors Elk1 and SAP1 α , which collaborate with SRF to regulate immediate early cardiac specific gene expression, an essential component on the hypertrophic response (Sugden & Clerk, 1998). The MEK5/ERK5 MAPK module is also an important contributor to hypertrophy. However, this module appears to mediate signals that result in an eccentric hypertrophy which is observed in volume overload-associated hypertrophy *in vivo* (Sugden & Clerk, 1998).

1.4.3.2 c-Jun N-terminal kinases

The JNKs, which (along with p38 MAPKs) are also commonly known as the stress-activated protein kinases (SAPKs) that mediate cellular responses to pathological

stresses and cytotoxic agents (Sugden & Clerk, 1998). The JNK class of MAPKs are directly phosphorylated by either MKK7 or MKK4, which in turn are regulated by MEKK1 and MEKK2 phosphorylation. Various upstream protein kinases including the small G proteins (Ras) phosphorylate the MEKK1/2 components (Clerk et al, 2007; Molkentin & Dorn, 2001; Sugden & Clerk, 1998b). In isolated cardiomyocytes, agonist stimulation by AngII, PE and ET-1 in addition to mechanical stretching all stimulate a rapid phosphorylation of JNK (Clerk et al, 2007; Molkentin & Dorn, 2001; Sugden & Clerk, 1998b). This activation is associated with the phosphorylation of the downstream transcription factors such as c-Jun, ATF-2, MEF2 and NFAT, all of which are implicated in myocardial growth. To support an important role for JNK, over-expression of the JNK upstream activators MEK1 or MEK7, induce gene expression changes coupled to morphological features of the hypertrophic response (Bogoyevitch et al, 1996a; Bogoyevitch et al, 1996b; Bogoyevitch & Sugden, 1996). Conversely, a dominant negative MKK4 mutant attenuates the ET-1 and pressure overload-induced hypertrophic response. In addition further evidence stems from *in vivo* studies utilising gene-targeted mice with disruption of the endogenous MKK1 gene, which display decreased JNK activity and hypertrophic sensitivity (Bogoyevitch et al, 1996a)

Although historically considered to function as a positive regulator of hypertrophy, numerous *in vivo* studies suggest the contrary and implicate the JNK pathway in negative regulation of cardiac hypertrophy. For example, it was demonstrated that over-expressing MEKK1 *in vitro* attenuated sarcomeric organisation, suggesting of a negative anti-hypertrophic effect. More significantly, cardiac specific over-expression of an MKK7-JNK1 fusion protein attenuated the hypertrophic response induced by an activated calcineurin transgene (Liang & Molkentin, 2003).

1.4.3.3 p38 mitogen-activated protein kinases

The p38 MAPK proteins contain four separate isoforms, specifically p38 α , β , gamma and δ . p38 α and β are abundantly expressed in the heart and numerous studies have implicated them in hypertrophic signalling and growth, although the direct role played in these processes is not definitively understood (Molkentin & Dorn, 2001). The most

important activators of these p38 isoforms are MKK3 and MKK6 which have been shown to be sufficient to induce hypertrophy with associated increases in marker gene expression *in vitro*, specifically ANF. In addition, direct pharmacologic inhibition of p38 MAPK activity with SB203580 and SB202190 is also able to repress characteristic aspects of agonist-stimulated hypertrophy *in vitro* (Zechner et al, 1997). To further support this, adeno-viral over-expression of a dominant-negative p38 β isoform blunted the growth response of neonatal myocytes which was coupled to reduced BNP promoter activity (Liang & Gardner, 1999; Liang et al, 2000; Wang et al, 1998). More significant evidence comes from the findings that over-expression of activated MKK3 or MKK6 also drive the development of hypertrophy *in vitro* which is associated with up-regulation in hypertrophy associated stress transcripts (Nemoto et al, 1998; Wang et al, 1998; Zechner et al, 1997). These initial findings strongly supported a role for the p38 MAPKs in the signalling cascades that lead to the development of cardiac hypertrophy, at least in isolated cardiomyocyte cultures.

However, like the other stress activated protein kinase family of JNK proteins, recent studies which have utilised genetically modified mice have indicated that the p38 signalling pathway does not function as a forward regulator of the hypertrophic response *in vivo*. For example, transgenic cardiac specific over-expression of activated MKK3 or MKK6 does not induce the hypertrophic response within individual myocytes (Liao et al, 2001). It was interesting that these transgenic mice actually develop rapid heart failure associated with reduced cardiac performance, increased fibrosis and thinned ventricular walls (Liao et al, 2001). This suggests that the myocardium is not able to undergo correct developmental hypertrophy, thereby transiting into a more compromised and dilated state. Further evidence comes from a study in which transgenic mice over-expressing a dominant negative p38 kinase displayed a nearly identical hypertrophic response compared to wild type controls. This suggests that in this setting the inhibition of p38 does not antagonise adaptive myocardial growth. Collectively, these results (from various gain-of-function genetically modified mice) indicate that p38 (and JNK) do not induce hypertrophy within the context of the adult heart. They may however have more important roles in cardiomyocyte apoptosis and the promotion of dilated cardiomyopathy.

1.4.4 Phosphoinositide 3-Kinase/Protein kinase B/Glycogen synthase kinase-3

Phosphoinositide 3-kinases (PI3Ks) have been implicated in cell growth, proliferation and survival in many different cell types (Cantley, 2002). Within the context of the mammalian heart, over-expression of constitutively active PI3K leads to increase in myocardial mass while dominant negative PI3K causes myocardial atrophy. It is interesting that within these dominant negative mice, cardiac function at resting conditions was not compromised, even though the hearts were smaller. However, when these mice were subjected to a cardiac hypertrophy inducing stimulus, specifically pressure overload, cardiac function was compromised (Matsui et al, 2002).

Arguably the most important target of PI3K within the heart is the serine/threonine kinase, protein kinase B (PKB), also known as Akt. Interestingly, cardiac specific over-expression of Akt leads to the development of cardiac hypertrophy in transgenic mice without adverse effects on systolic function, suggestive of a more physiological form of hypertrophy (Matsui et al, 2002; Condorelli et al, 2002).

Akt related effects are primarily mediated by two specific downstream targets, the glycogen synthase kinase (GSK-3) and the mammalian target of rapamycin (mTOR). mTOR is involved in enhancing protein synthesis, through the p70S6 kinase and 4EBP1/eIF4E, which is a classical hallmark of hypertrophy. PKB also phosphorylates GSK-3 β which results in inhibition of its kinase activity. GSK-3 β was among the first negative regulators of cardiac hypertrophy to be identified and was found to be able to block normal and pathological cardiac hypertrophy (Walsh, 2006). This kinase is essentially 'on' in the cell until it is turned 'off' by these hypertrophy promoting stimuli and PKB activity, thereby causing the release of its substrates from constitutive inhibition. One family of GSK-3 β inhibited substrates include specific transcription factors with well documented roles in cardiac growth (Haq et al, 2000; Michael et al, 2004; Morisco et al, 2000). These targets which include c-Myc, GATA4 and β -catenin, may be particularly important in the reprogramming of gene expression that characterises both adaptive and maladaptive cardiac hypertrophy (Pikkarainen et al,

2004; Xiao et al, 2001). Arguably the most important transcription factor substrates for GSK-3 β are the NFAT family of factors. These factors are phosphorylated by GSK-3 β at amino-terminal residues which are dephosphorylated by calcineurin (Figure 6). This GSK-3 β dependent phosphorylation therefore inhibits calcineurin driven nuclear translocation of the NFATs, thereby restricting NFAT dependent target gene activation and cardiac hypertrophy. Solid evidence for this proposed mechanism comes from work utilising transgenic mice where concomitant GSK-3 β overexpression attenuated hypertrophy of calcineurin-transgenic mice (Antos et al, 2002). In summary PI3K/PKB signalling contributes to the development of cardiac hypertrophy through the phosphorylation and subsequent inhibition of GSK-3 β .

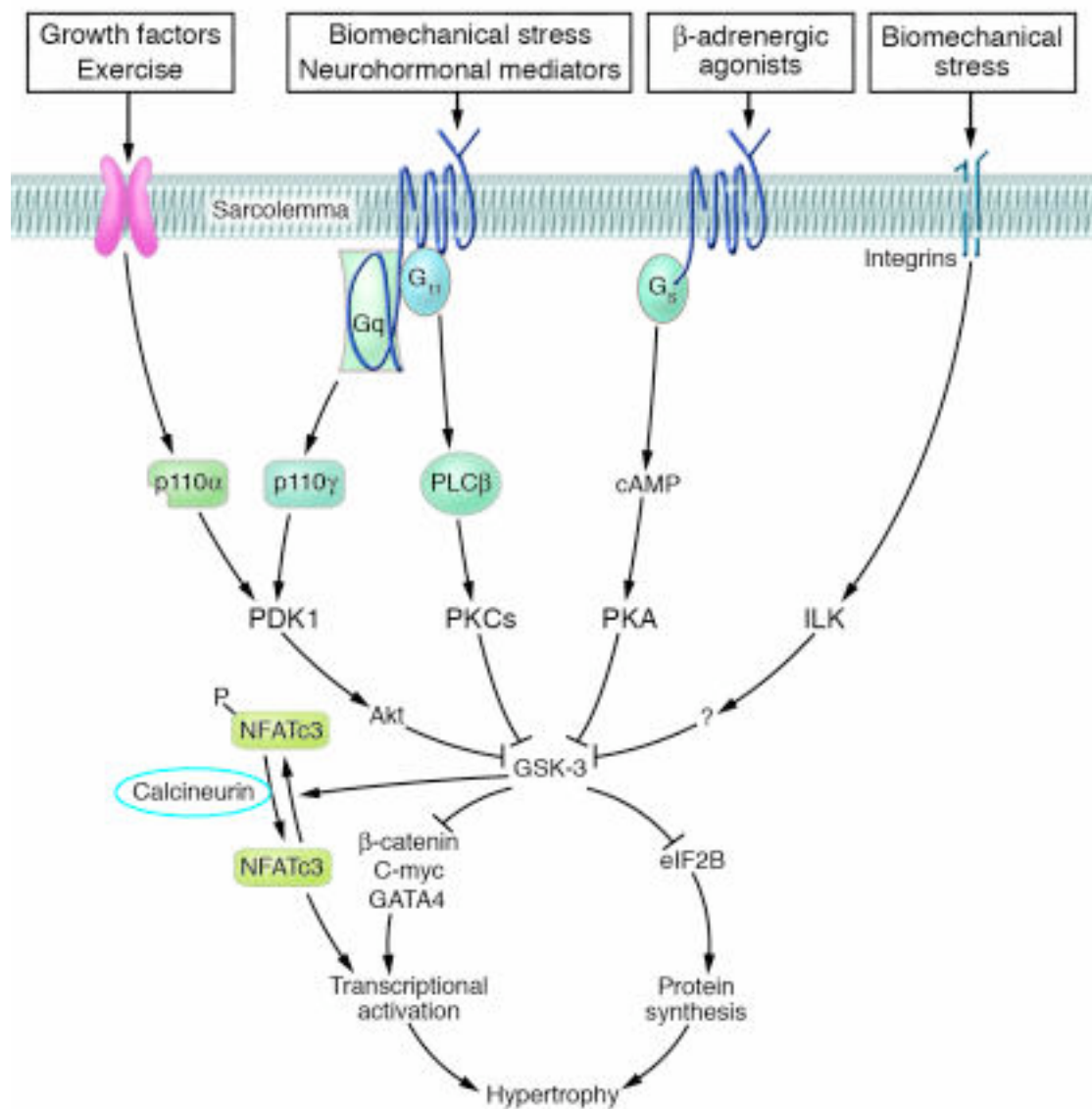


Figure 6. GSK-3 integrates multiple hypertrophic signal transduction pathways (Dorn II et al, 2005).

1.4.5 Calcineurin

Intense research has focused on the calcium-dependent signalling pathways implicated in cardiac hypertrophy (Olson, 2004). One calcium-dependent pathway that has received considerable attention encompasses the calcium/calmodulin-activated protein phosphatase, calcineurin (PP2B). Calcineurin is a serine/threonine-specific protein phosphatase that is exquisitely sensitive to sustained elevations in intracellular calcium. Calcium dependent activation facilitates binding to its primary downstream effector, the transcription factor NFAT (nuclear factor of activated T cells). In an unstimulated context NFAT transcription factors are typically hyperphosphorylated and therefore sequestered in the cytoplasm. However, upon stimulation, these factors rapidly translocate to the nucleus via a mechanism dependent on calcineurin-mediated dephosphorylation (Molkentin, 2004)(Figure 7).

Cardiac specific activation of NFAT or its upstream activator calcineurin is sufficient to induce a robust hypertrophic response in transgenic mice models (Molkentin, 2004). Conversely, inhibition of these components through gene targeting has convincingly demonstrated that this pathway is necessary for the full cardiac hypertrophic response in numerous rodent animal models (Bueno et al, 2002a; Bueno et al, 2002b). It is important to note that functionally significant cross-talk exists between the calcineurin signalling pathways with members of the MAPK family of signalling kinases. For example, mice expressing an activated calcineurin transgene displayed enhanced JNK and ERK activation. In addition, isoproterenol induced ERK activation was also dependent on calcineurin activity in isolated cardiomyocytes (Molkentin, 2004). It is of interest that recent data suggests that the NFAT/calcineurin module, unlike other pathways (for example PI3K/PKB/GSK-3 β), may be uniquely activated in pathological forms of hypertrophy, and during normal physiological hypertrophic phenotypes (Wilkins & Molkentin, 2004).

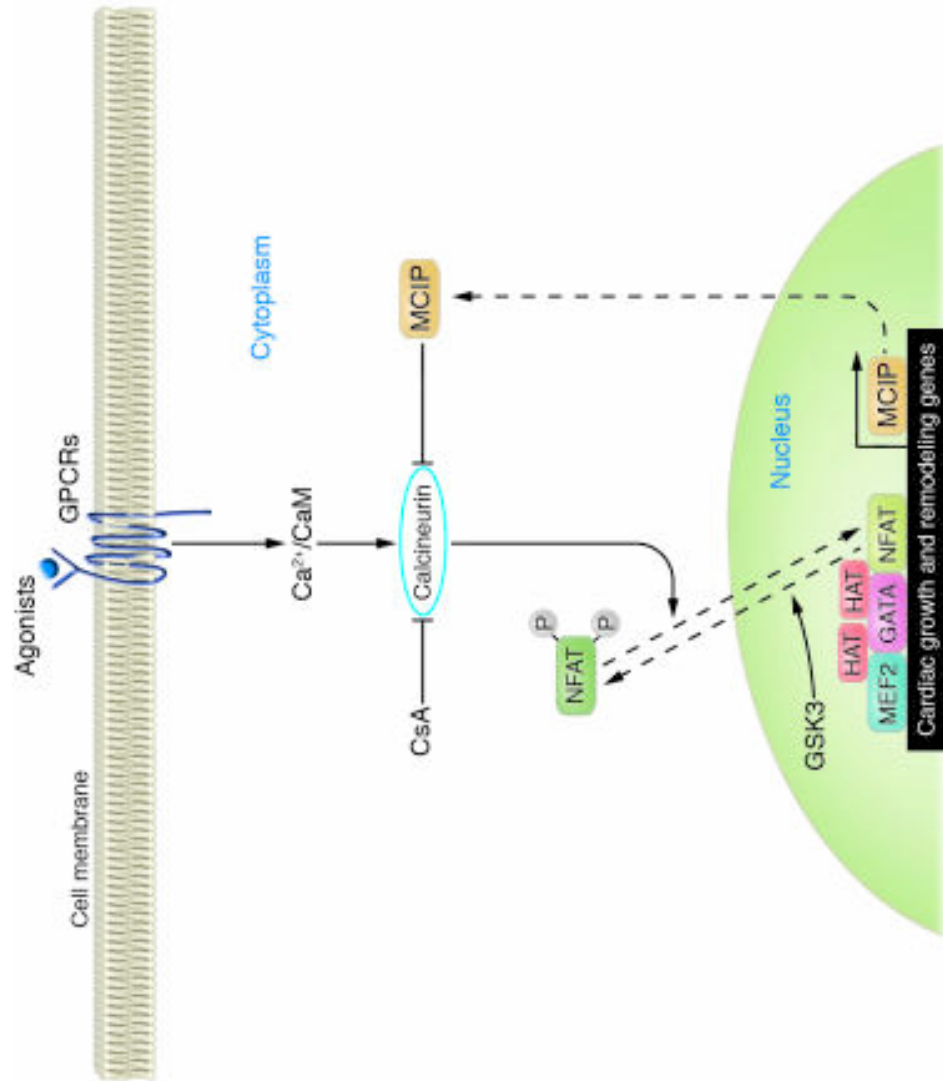


Figure 7. Calcineurin as a central regulator of signal-dependent modulation of cardiac hypertrophy (Olson et al, 2005).

1.5 *Transcription factors in cardiac hypertrophy*

The signalling pathways described above ultimately control cardiac hypertrophy through their regulatory modulation of cardiac gene transcription. Transcription of a particular gene is governed by the complex interplay of multiple transcription factors and associated co-factors which bind to specific DNA elements within the proximal 5' promoter elements and associated distal *cis* regulatory elements. The promoter, and associated distal regulatory elements, are therefore a key point for signalling pathway integration whether this is as a consequence of direct phosphorylation/dephosphorylation of transcription factors downstream of protein kinase/phosphatase signalling pathways, or a result of altered concentrations of transcription factors in the nucleus. Throughout the last decade a finite number of transcription factors have been identified which are implicated in cardiac hypertrophy and particularly important for the direct transcriptional 'transduction' of the signalling pathways discussed above (Oka et al, 2007). It is of interest to note that the majority of the factors characterised also have functionally important roles in cardiac development. This paradigm therefore suggests that adult heart hypertrophy is regulated through the re-employment of developmental transcription factors. The current understanding of these hypertrophy driving transcription factors, focusing on those with defined roles in cardiac development, differentiation and maturation, will be discussed below.

1.5.1 *GATA binding protein 4*

Within vertebrates six GATA family transcription factors have been identified and they can be separated into two subclasses based on their expression profiles. GATA-4, -5 and -6 are expressed in various endoderm and mesoderm derived tissues including the heart, liver, gonads and gut (Molkentin, 2000), while GATA-1, -2 and -3 are primarily restricted to the hematopoietic lineages. All the GATA family members bind the nucleotide sequence element (A/T)GATA(A/G) through a highly conserved dual zinc finger DNA binding domain. They also contain a potent transactivation domain in addition to other domains that also contribute to transcriptional activation through specific interactions with transcriptional co-factors (Molkentin, 2000). Although

GATA-5 and -6 have been implicated in cardiac context specific gene expression, GATA4 has been extensively studied, with numerous studies indicating this factor as an essential regulator of cardiac development and differentiation, as well as an important regulator of cardiomyocyte survival and hypertrophic growth (Molkentin, 2000; Pikkarainen et al, 2004).

In one of the first studies exploring a role for GATA4 within cardiac specific development, tetraploid embryo complementation was employed utilising *Gata4*^{-/-} embryonic stem cells. The subsequent embryo displayed hypoplastic ventricles and a complete loss of the proepicardium which resulted in lethality (Watt et al, 2004). Using *Cre-LoxP*-based technology, GATA4 was specifically knocked out in the embryonic heart (using a *Nkx2.5-Cre* knock-in allele), with these embryos again displaying hypoplastic ventricles and embryonic lethality (Oka et al, 2006; Pu et al, 2004). In humans, recent association studies have demonstrated that a heterozygous mutation in GATA4 is strongly associated with congenital abnormalities in cardiac septation, thereby exemplifying the developmental importance of GATA4 (Garg et al, 2003). Collectively these studies in genetically manipulated rodent models supplemented with the findings in human cohorts with congenital cardiac abnormalities underscore the importance of GATA4 in regulating developmental and differentiated gene expression in the heart.

GATA4 is also expressed in the adult heart where it functions as a key transcriptional regulator of many well characterised cardiac genes including β -myosin heavy chain (β -MHC), atrial natriuretic factor (ANF), b-type natriuretic peptide (BNP) and sodium for calcium exchanger (NCX) (Molkentin, 2000; Pikkarainen et al, 2004). A direct role for GATA4 in mediating differentiated cardiac gene expression is supported by the observation that anti-sense GATA4 mRNA expression was able to inhibit the basal expression of many cardiac-restricted genes in isolated cardiomyocyte cultures. In addition to basal cardiac specific expression, GATA4 also mediates inducible gene expression in response to many of the classical hypertrophic stimuli including endothelin-1, phenylephrine, pressure overload and isoproterenol (Hasegawa et al, 1997; Herzig et al, 1997; Liang et al, 2001; Morimoto et al, 2000). A direct regulatory

role in cardiac hypertrophy is also inferred from work where adenoviral driven over expression of GATA4 in NRVMs is sufficient to induce cardiomyocyte hypertrophy (Liang et al, 2001). Conversely and more significantly, expression of anti-sense mRNA or dominant negative GATA4 was able to completely block the GATA4-directed transcriptional responses and features of cardiomyocyte hypertrophy induced by endothelin-1 and phenylephrine *in vitro* (Charron et al, 2001; Liang et al, 2001).

Recent work has demonstrated GATA4's necessity in mediating cardiac hypertrophy *in vivo*. Using conditionally targeted GATA4-loxP allele in conjunction with two different heart-specific *Cre*-expressing transgenic lines, mice were generated that had reduced GATA4 specifically within the cardiomyocytes (70% and 95% reduction). These mice were viable into young adulthood but interestingly as these mice aged they developed cardiac dilation and heart failure, which the authors suggest is a product of increased apoptosis (Oka et al, 2006). This is not surprising with others demonstrating that germline heterozygous targeted GATA4 mice showed greater cardiac apoptosis following doxorubicin treatment. More relevant to cardiac hypertrophy, mice with cardiac specific conditional GATA4 deletion exhibited attenuated myocardial growth following prolonged pressure overload or following exercise stimulation (Oka et al, 2006). It was, however, of particular interest that this loss of Gata4 did not affect post-natal growth of the heart (developmental hypertrophy) therefore suggestive that GATA4 function in the adult heart is specific to the regulation of maladaptive and adaptive growth.

The mechanisms through which the classical hypertrophic agonists stimulate GATA4 driven cardiac hypertrophy are also beginning to be unravelled. For example, a number of these stimuli have been shown to enhance GATA4 transcriptional activity through phosphorylation: pressure overload, angiotensin II, endothelin-1, phorbol esters and isoproterenol induced specific phosphorylation of GATA4. This event resulted in increased DNA binding and/or transactivation potential of GATA4 (Hasegawa et al, 1997; Hautala et al, 2001; Kerkela et al, 2002; Kitta et al, 2001; Liang et al, 2001; Morimoto et al, 2000; Morisco et al, 2001). These phosphorylation events are mediated by the MAPK signalling pathways. Stimulation of isolated myocytes or the whole heart

with appropriate hypertrophic agonists results in phosphorylation of GATA4 at serine 105, with phosphorylation at this residue potentially enhancing DNA binding activity and transcriptional activity of GATA4. This is via the direct activation of ERK1/2 and p38 MAPK (Charron et al, 2001; Liang et al, 2001). Due to the fact that both these MAPK receive signal inputs from diverse upstream signalling pathways suggests that serine 105 in GATA4 serves a role as a key convergence point in regulating the cardiac hypertrophic response.

In addition to positive post-translation regulation, GATA4 is also negatively regulated by GSK-3 β -mediated phosphorylation. This event is able to reduce both basal and stimulus-induced nuclear localisation of GATA4 thereby suppressing its transcriptional activity. In summary, it is clear that GATA4 is a key transcriptional regulatory of basal and stress inducible cardiac specific gene expression. GATA4 also serves as a key transcriptional convergence point in the adult heart where multiple stress-dependent signalling pathways modulates its function as a mechanism of controlling the cardiac hypertrophic response.

1.5.2 Myocyte enhancement factor 2

Myocyte enhancement factor 2 (MEF2) was first characterised as a muscle enriched DNA binding activity from differentiated skeletal myotubes (Gossett et al, 1989). This muscle enriched DNA binding activity is encoded for and consists of homo- and heterodimers of four separate transcripts designated Mef2-A, -B, -C and -D (Black & Olson, 1998; Molkentin & Olson, 1996). Dimers of these MEF2 proteins bind to the consensus MEF2 motif, CTA(A/T)₄TAG, which is present in the proximal and distal regulatory domains of most skeletal and cardiac muscle structural genes characterised to date (Black & Olson, 1998; Molkentin & Olson, 1996). The MEF2 family are also distantly related with another key muscle regulatory factor, serum response factor (SRF), with both of these containing the MADS-box (MCM1, Agamouse and Deficiens, SRF) domain. The MEF2 factors are generally ubiquitously expressed in the adult vertebrate organism, however specific regulatory functions for these factors have been

demonstrated in specific tissue and organ types, including neuronal cells and striated muscle.

Similar to GATA4, the MEF2 transcription factor family are essential regulators of cardiac development, including morphogenesis, differentiation and maturation. For example, murine targeted deletion of MEF2C led to an early embryonic lethal phenotype which was associated with a down-regulation of cardiac specific genes, which presumably caused the absence of the right ventricle and a cardiac looping defect (Bi et al, 1999; Han et al, 1997). On the other hand, loss of MEF2A led to severe disorganisation of the myofibres which resulted in right ventricular dilation. Although MEF2A and -C murine mutants presented a clear cardiovascular phenotype, final characterisation of MEF2 family function in cardiac development may be impossible given that all four MEF2 family members are expressed in the developing heart and would therefore all have to be inactivated simultaneously. However, in order to overcome these limitations associated with functional redundancy, transgenic mice have been generated that express a MEF2-dominant negative protein, specifically in the heart. Not surprisingly, these mice displayed an early postnatal lethality that was associated with ventricular wall thinning, hypoplasia of cardiomyocytes and chamber dilation (Kolodziejczyk et al, 1999). This finding, together with the MEF2 specific knock out findings supports an important role for MEF2 family proteins in the proper differentiation and postnatal development of the heart.

MEF2 family proteins have also emerged as central regulators of cardiomyocyte hypertrophy in the adult heart. A number of lines of evidence have emerged supporting this hypothesis. Firstly, MEF2 DNA-binding activity is increased by a number of classical hypertrophic stimuli including pressure and volume-overload *in vivo* (Molkentin & Markham, 1993; Nadruz et al, 2005), and myocyte stretching *in vitro* (Nadruz et al, 2005; Shyu et al, 2005). The second and most compelling evidence comes from studies demonstrating that cardiac specific mild over-expression of MEF2A or MEF2C is sufficient to induce ventricular dilation and contractile dysfunction. More significantly, these mice displayed greater hypertrophic growth following pressure overload (Xu et al, 2006). However one should note the myocytes from these transgenic

mice showed a relative increase in length, as opposed to an increase in cross-sectional area, suggesting that MEF2 over-expression did not drive a classic hypertrophic phenotype *per se*.

Supplementary to these direct lines of evidence, numerous other indirect findings have further implicated MEF2 as a central regulator of adult heart disease and potentially the pathological hypertrophic response. As will be further discussed in Section 1.7.2.1, histone de-acetylase (HDACs) enzymes, specifically the Class II HDAC family, are important regulators of gene expression that function in co-ordination with MEF2 family factors in the regulation of muscle specific gene expression (Lu et al, 2000; Miska et al, 1999; Sparrow et al, 1999; Wang et al, 1999; Xu et al, 2006). The calcium/calmodulin-dependent protein kinases (CaMK) is able to module the interaction of MEF2 with the Class II HDACs. Hypertrophy initiated calcium-dependent pathways result in CaMK activation, which subsequently leads to direct phosphorylation of class II HDACs, thereby stimulating their nuclear extrusion and permitting MEF2 factors to activate muscle and hypertrophy associated gene expression (Lu et al, 2000; McKinsey et al, 2000). Through the use of an exquisite MEF2-dependent murine reporter transgene, it was demonstrated that these factors are the primary mediators of both class II HDAC and CaMK-dependent cardiac hypertrophy. More specifically this reporter demonstrated that MEF2 directly responded to activated CaMK and calcineurin in the heart. Of note, these pathways are strongly implicated in cardiomyocyte hypertrophy.

In addition to being downstream of CaMK and calcineurin, MEF2 factors are also directly activated by cardiac hypertrophy signalling effectors including big-MAPK-1 (BMK-1/ERK5)(Kato et al, 1997). ERK5, which is downstream of and activated by MEK5 (Kato et al, 1997; Marinissen et al, 1999), directly phosphorylates MEF2 factors therefore increasing their activity. In support of this, activated MEK5 transgenic mice showed a phenotype reminiscent of Mef2 transgenic models, which included ventricular dilation, reduced contractile function and activation of the hypertrophic gene expression programme. In summary these findings suggest that MEF2 family proteins are likely to function as important and central regulators of adult cardiac disease

responses. This is especially true for dilated cardiomyopathy and to some extent adult hypertrophy.

1.5.3 Nuclear factor of activated T cells

The nuclear factor of activated T cells (NFATs) factors and their upstream signalling regulator, calcineurin, have been implicated as critical transducers of the cardiac hypertrophic response. As described previously, stimulus-dependent elevations in intracellular calcium activates calcineurin, which subsequently binds cytoplasmic NFAT and de-phosphorylates the N-terminal regulatory domains. This exposes the NFAT nuclear localisation signal thereby stimulating its nuclear translocation where in collaboration with other nuclear factors, including GATA4, it trans-activates downstream gene expression, thereby stimulating the hypertrophic response (Wilkins & Molkentin, 2004). Upon nuclear trans-location the four calcineurin-modulated family members, which are NFATc1-, -2, -3 and -4, bind an ambiguous consensus sequence (G/A)GAAA either alone or in association with AP-1 and GATA4 (Hogan et al, 2003; Wilkins & Molkentin, 2004). It is important to note that the N-terminal regulatory domain is enriched with multiple phosphorylation sensitive serine and threonine residues. These residues are sensitive to negative-acting phosphorylation by a diverse group of signalling kinases, resulting in cytoplasmic sequestering and thus inhibited NFAT activity. One can therefore describe NFAT as a factor similar to GATA4 as these factors both function as signalling convergence points whereby multiple hypertrophic signalling pathways integrate to modulate the appropriate transcriptional output.

NFAT factors are important regulators in cardiac development (Schulz & Yutzey, 2004). Indeed, cardiac restricted over expression of a dominant negative NFAT mutant was sufficient to result in thinned atria and significant down-regulation of cardiac structural genes (Schubert et al, 2003), thus suggesting that NFAT factors are particularly important in myocyte maturation. In agreement with this deduction, NFATs have also been demonstrated to be very important modulators of myocyte growth in the adult heart. In support of this, cardiac-specific increases in calcineurin activity and the over-expression of a constitutively nuclear NFAT factor, both induced substantial cardiac

hypertrophy that quickly developed into heart failure (Molkentin et al, 1998). After this initial seminal study, numerous follow on studies exemplified the centrality of the calcineurin/NFAT module as a necessary mediator of pathological cardiac hypertrophy (Wilkins & Molkentin, 2004). To further explore the specificity of this master hypertrophic signalling module, Molkentin's group generated transgenic murine lines containing an NFAT-dependent luciferase reporter. This reporter demonstrated specific myocardial activation by activated calcineurin, and was conversely repressed by the calcineurin inhibitor cyclosporine (Wilkins et al, 2004). In addition, these reporter mice also allowed the dynamics of calcineurin activation to be explored and it was shown that the calcineurin/NFAT pathway was constitutively upregulated throughout a time course of pressure overload induced hypertrophy (Wilkins et al, 2004). This was also true in the failing mouse heart model following myocardial infarction (Wilkins et al, 2004).

Many groups have also utilised inhibitory dominant-negative and gene targeting technologies to further probe the necessity and sufficiency of NFATs as central hypertrophic mediators. For example, NFATc3 knock out mice displayed significant and constitutive reduction in calcineurin-induced cardiac hypertrophy. These mice also displayed a compromised response to mount the appropriate hypertrophic compensated state following aortic banding or angiotensin II infusion (Wilkins et al, 2002). This suggested that NFATc3 was a critical intermediate in calcineurin-induced hypertrophy of the adult heart. These results, and specifically this hypothesis was supported by studies examining the effects of GSK-3beta on myocyte hypertrophy (Antos et al, 2002; Haq et al, 2000; Sanbe et al, 2003). GSK-3beta has been shown to directly phosphorylate the N-terminal regulatory domain of NFAT factors, thereby antagonising the effect of calcineurin and thus inhibiting nuclear accumulation. It is of interest that p38 MAPK and JNK MAPK activity also regulates cardiac hypertrophy, in part through an NFAT-dependent mechanism (Braz et al, 2003; Liang et al, 2003).

1.5.4 Smads

The Smad family of transcription factors are primarily associated with the transduction of transforming growth factor- β (TGF- β) super-family member signals. These ligands induce assembly of associated receptors on the plasma membrane which results in direct phosphorylation and activation of Smad proteins residing within the cytoplasm, therefore inducing their nuclear translocation and accumulation (Shin et al, 2002; ten Dijke & Hill, 2004). The Smads themselves comprise of three distinct subfamilies which have different structures and associated function: (1) Receptor-Smads, including Smad-1, -2, -3, -5 and -8, (2) Co-factor-Smad, includes Smad4, which then associates with the receptor-Smads to activate downstream genes, and (3) Inhibitory-Smads (I-Smad), includes Smad-6 and -7. These Smads serve to inhibit the trans-activating potential of the receptor and cofactor Smads (Shin et al, 2002; ten Dijke & Hill, 2004).

The Smads have been shown to be important regulatory factors in cardiac development, specifically modulating endocardial cushion formation, valve morphogenesis and cardiac lineage determination inducers, with is thought to be as a consequence of their role in inducing the expression of Nkx2-5 (section 1.5.6)(Schultheiss et al, 1997). For example, gene targeted Smad6 mice displayed severe cardiac abnormalities associated with myocardial differentiation and morphogenesis. Within the adult heart, Smad factors are critically involved in multiple aspects of cardiac pathophysiology, including the regulation of cardiac hypertrophy, the transition to heart failure and fibrosis (Araujo-Jorge et al, 2002; Cucoranu et al, 2005; Dixon et al, 2000; Hao et al, 1999; Hao et al, 2000; Wang et al, 2002). The exact nature of the roles that the Smad factors play in these phenotypes is somewhat ambiguous. For example, gene targeted Smad4 mice developed basal hypertrophy that progressed rapidly to heart failure therefore suggestive that Smad activation serves a basal anti-hypertrophic regulatory function in the adult heart (Wang et al, 2005). However, although these results suggest a negative role, TGF- β itself is thought to be a pro-hypertrophic cytokine within the myocardium. It has been demonstrated the reduction of TGF- β 1 in the mouse attenuates hypertrophic sensitivity to angiotensin II (Schultz Jel et al, 2002). Conversely, TGF- β over-expression via transgenesis induced cardiac hypertrophy however one caveat being that this effect could not be directly coupled with Smad activation, and therefore may likely to occur in a Smad independent manner (Rosenkranz et al, 2002). Although the exact

nature of Smad factor function in cardiac hypertrophy remains to be defined, it is clear that on the whole they play important regulatory functions in cardiac pathophysiological phenotypes.

1.5.5 Nkx2-5

Nkx2-5 (Csx) is a unique homeobox transcription factor which recognises the consensus sequences 5'TNAAGTG-3' and 5'-TTAATT-3'. Nkx2-5 is one of the first and arguably most important transcription factors implicated in the modulation of cardiac gene expression and heart development (Komuro & Izumo, 1993; Lints et al, 1993), with gain and loss-of function strategies implicating it with cardiac lineage commitment and subsequent heart tube looping and morphogenesis (Kasahara et al, 1998; Komuro & Izumo, 1993; Lints et al, 1993; Stanley et al, 2002). In addition, compelling evidence from humans demonstrates that patients with mutations in Nkx2-5 have congenital abnormalities typically characterised by aberrant ventricular septation and cardiac conduction abnormalities (Schott et al, 1998).

Although Nkx2-5 is well defined as being a central modulator during embryonic cardiac development, its regulatory role in the post natal heart is unclear. However, studies exist which implicate Nkx2-5 in the development of cardiac hypertrophy. For example animal studies have demonstrated that Nkx2-5 expression is upregulated in response to classical hypertrophic stimuli including aortic banding, phenylephrine and isoproterenol infusion (Saadane et al, 1999; Thompson et al, 1998). Nkx2-5 can also potentially function as a modulator of the cardiac hypertrophic response through its known ability to interact with other cardiac restricted factors including SRF (Chen & Schwartz, 1996) and GATA4 (Durocher et al, 1997; Lee et al, 1998; Shiojima et al, 1999; Takimoto et al, 2000). It is of interest that recent work also demonstrates that Nkx2-5 directly interacts with the newly identified calmodulin binding transactivator (CAMTA), which itself is capable to activate hypertrophic gene expression (ANF upregulation) and hypertrophy (Song et al, 2006). In summary, the current literature demonstrates that Nkx2-5 is a particularly important modulator of cardiac-specific gene expression and potentially plays an important co-operative role in the adult cardiac hypertrophic response.

1.5.6 Hand Family

Hand1 and Hand2 which are also known as eHAND and dHAND respectively are classical helix-loop-helix transcription factors and therefore recognise and bind the E-Box DNA response element (5'-CANNTG-3'). Both of these factors have important regulatory functions during cardiac development and specifically in the regulation of ventricular specific gene expression, with Hand1 regulating gene expression within the left ventricle and Hand2 within the right (Firulli, 2003; Riley et al, 2000; Srivastava, 1999; Srivastava et al, 1997). Although no circumstantial evidence currently exists, early work has demonstrated that Hand2 is up-regulated in a rat model of pressure overload induced cardiac hypertrophy. However, others have reported the down-regulation of both Hand1 and -2 in a mouse model of phenylephrine-induced hypertrophy (Srivastava, 1999; Srivastava et al, 1997). Therefore, although the exact nature and mechanisms of Hand action are somewhat confusing to date, one can accept that changes in cardiac specific Hand activity may play a role in adult heart disease responses although direct functional evidence is needed to support and further delineate this possibility.

1.5.7 Kruppel like factor family

The Kruppel like factor family of transcription factors (KLFs) are archetypal C2H2 zinc finger transcription factors which possess two cysteine and two histidine residues which coordinate zinc within each finger to form a conserved DNA-binding structure. This structure recognises the GC-rich consensus sequence CACCC (Kaczynski et al, 2003). There are currently 17 characterised Klf family members in mammals designated Klf1-17 (Suske et al, 2005; van Vliet et al, 2006). Many of the currently identified Klf's are expressed in the heart although the specific myocardial cell types they are expressed in are still unclear (Atkins & Jain, 2007; Feinberg et al, 2004). As of now, only Klf13 and -15 have been shown to play an important role in cardiac myocytes (Fisch et al, 2007; Uchida et al, 2000). Klf15, which is highly expressed in the myocytes of the adult heart, appears to play a negative regulatory role in cardiac hypertrophy since it is down-regulated in hypertrophy. In addition, overexpressing Klf15 is sufficient to repress the

changes in gene expression and associated morphological changes induced by hypertrophic stimuli (Fisch et al, 2007). However, it is of interest that in a recent study by Cullingford and colleagues (Cullingford et al, 2008), using a micro-array based strategy, they demonstrated the differential expression of numerous other Klf s in response to the potent hypertrophic agonist, endothelin-1. This finding implies that other members of the Klf family may also have important regulatory roles in the response of cardiomyocytes to hypertrophy inducing extracellular stimuli.

1.5.8 Serum response factor

Serum response factor (SRF) was originally characterised as a transcription factor which conferred serum sensitivity on the immediate early gene *c-fos* (Norman et al, 1988). As the founding member of MADs box domain family of proteins which includes MEF2 (Shore & Sharrocks, 1995). The MADs domain itself serves as a DNA binding domain through which SRF binds to the consensus sequence, CC(AT)₆GG, which is commonly referred to as the serum response element or CArG box. This motif is found in the proximal and distal regulatory domains of numerous cardiac, skeletal and smooth muscle specific genes in addition to the classical immediate early genes in which it was first characterised (Miano, 2003; Shore & Sharrocks, 1995).

During vertebrate embryogenesis SRF transcript is primarily detected within the striated and smooth muscle lineages with more ubiquitous expression throughout the whole embryo at later developmental stages (Belaguli et al, 1997; Croissant et al, 1996). Early work attempting to characterise the developmental role of SRF in muscle was hindered by early lethality. In these models, SRF deficiency resulted in impaired gastrulation and lethality occurring before formation of the mesoderm and therefore muscle lineages (Arsenian et al, 1998). However, in subsequent studies cardiac specific conditional SRF-gene targeted mice were utilised, and results from these studies demonstrated that SRF was required for correct cardiac development via its obligatory role as a master regulator of cardiac specific gene expression involved in sarcomerogenesis. With respect to skeletal muscle, SRF deficiency was associated with severe tissue hypoplasia,

again supportive for a crucial role for SRF for regulating striated muscle-specific gene expression during development.

SRF has also been implicated as an important regulatory modulator of the cardiac hypertrophic response. For example, hearts over-expressing a mutant SRF protein displayed severe chamber dilation, wall thinning and lethality suggestive of an important role in the development of hypertrophy (Zhang et al, 2001a; Zhang et al, 2001b). Conversely, and in support of this hypothesis, cardiac-specific over-expression of SRF was sufficient to induce robust patho-physiological hypertrophy (Zhang et al, 2001a; Zhang et al, 2001b). At the level of the individual cardiomyocyte, SRF perturbation is sufficient to induce disorganisation of the contractile apparatus associated dysregulated stress fibre formation and attenuated expression of sarcomeric genes (Balza & Misra, 2006; Parlakian et al, 2005). In addition to the down-regulation of sarcomeric gene expression, micro-array analysis of SRF null myocytes also demonstrates down-regulation of important cardiac transcription factors involved in cardiac differentiation and hypertrophy (Balza & Misra, 2006). It therefore appears, and has been suggested, that SRF expression is obligatory for the maintenance of the basal “trophic” state of the heart, and also in the induction of heart growth (hypertrophy) in response to stimulation.

On its own, SRF is a poor transcriptional activator as it does not contain a transactivation domain. Therefore SRF dependent gene activation is dependent on co-operation with other transcription factors and co-factors. For example, SRF can interact with GATA4, Nkx2-5 and the Smads-1 and -3 to synergistically activate downstream muscle specific expression (Belaguli et al, 2000; Callis et al, 2005; Chen & Schwartz, 1996; Oh et al, 2004; Qiu et al, 2003; Sepulveda et al, 1998; Sepulveda et al, 2002). In addition, SRF can also be regulated by the homeodomain only protein, Hop. Hop binds to DNA-associated SRF and mediates the direct recruitment of HDACs which therefore results in transcriptional inhibition of SRF target genes (Chen et al, 2002). To conclude, SRF is a central regulatory factor of cardiac gene expression during development and pathological hypertrophy. More specifically, it appears to serve as a platform for integrating the independent activity of numerous co-factors which allows an exquisitely

fine tuned transcriptional response to occur, which subsequently mediates adult hypertrophic growth of the heart.

1.5.9 Myocardin family of SRF co-factors

In addition to the independently acting cardiac transcription factors described above, SRF is also regulated by an extra-ordinarily powerful family of co-factors, termed the myocardin family of SRF transcriptional co-activators. Myocardin was the founding member of this family and was discovered in a bioinformatics screen for novel cardiac restricted transcripts (Wang et al, 2001). Subsequent functional characterisation demonstrated that myocardin could stimulate the transcription of CArG-dependent enhancers in a DNA-binding independent manner. It was shown that this occurred through the formation of a stable ternary complex with SRF on specific CArG-boxes. In support of this, *SRF*-null cells displayed abolished myocardin-dependent transcriptional activity (Wang et al, 2002b). It was of great excitement at the time that myocardin expression was largely confined to the cardiovascular system, specifically within the cardiac myocytes of the heart and also smooth muscle cells, where myocardin appears to have a particularly important regulatory role (Miano, 2003).

Soon after the initial identification of myocardin, Olson's group proceeded to characterise the related homologous proteins designated the myocardin related transcription factors (MRTFs). These include myocardin related transcription factor-A (MRTF-A, also known as BSAC, MKL-1 and MAL), and myocardin related transcription factor-B (MRTF-B, also known as MKL-2) (Mercher et al, 2002; Wang et al, 2002b). Contrary to myocardin, these related factors appear to be expressed ubiquitously in a broad range of embryonic and adult tissues (Mercher et al, 2002; Wang et al, 2002b). At the structural level, all members of the myocardin family share homology in multiple functional domains (see the figure 8). The N-terminal region of the family members contains a triplicate of RPEL motifs which are important in the interaction of family members with actin (Miralles et al, 2003). The interaction of these co-factors with SRF is mediated by a short peptide sequence which is composed of a glutamine-rich and basic region (Wang et al, 2001; Wang et al, 2002b). In addition to

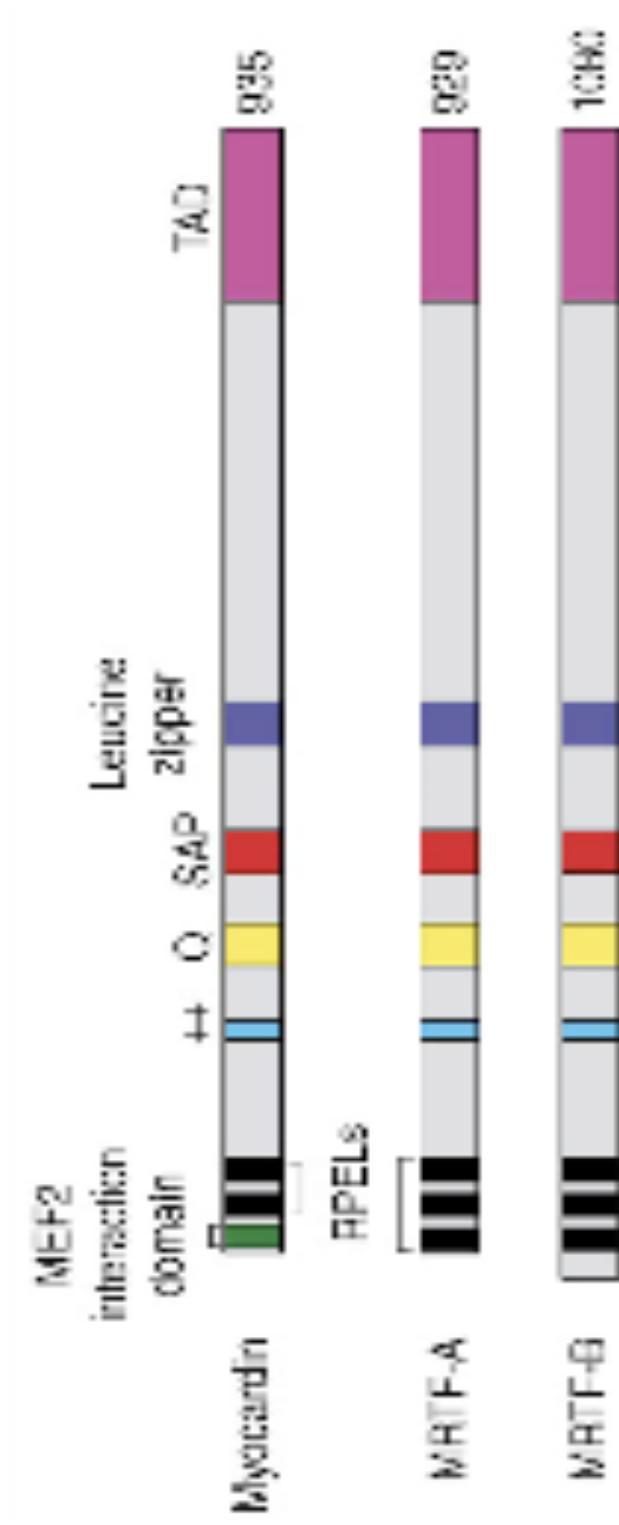


Figure 8. Structural domains and similarity of the myocardin family of SRF co-factors. ++, basic region; Q, glutamine-rich region; TAD, transcription activation domain; SAP, SAP protein domain; RPEL, RPEL protein domain motifs (Olson, 2006).

binding directly to SRF, the myocardin related family members also homo and hetero-dimerise with each other, and this is mediated by a coiled-coiled leucine zipper type motif. The C-terminal regions are the least conserved regions at the amino-acid level and these represent the transactivation domains (TADs), with truncation of these domains rendering the resultant proteins as dominant-negative mutants.

The mechanisms through which these TADs function are not fully understood, however studies with myocardin have begun to identify potential modes of action (Cao et al, 2005). It has been demonstrated that myocardin modulates SRF target genes through the targeted recruitment of chromatin remodelling enzymes. Context and stimulus dependent association with the histone acetyltransferase p300 enhances SRF trans-activity while interaction with class II HDACs represses (Cao et al, 2005). It is also of interest that although the mechanisms is not known, myocardin is also able to increase the binding of SRF to non-consensus CArG boxes, with this having functional implications for SRF transcriptional activity in smooth and cardiac muscle cells (Hendrix et al, 2005).

Given the central role SRF plays in cardiac specific gene expression during cardiac development and hypertrophy, it is not surprising to find that the myocardin family of co-factors are also important in these processes. For example, dominant negative myocardin (which will inhibit all myocardin related cofactors) can prevent cardiac gene expression in *Xenopus* embryos while conversely forced over-expression of wild type myocardin is sufficient to induce cardiac hypertrophy (Wang et al, 2001). In mammals, mice lacking myocardin do not show an overt cardiac phenotype, however this is likely as a result of compensation from the other MRTF family members. Some of these members do however have a specific regulatory function in the heart, specifically within the smooth muscle derived tissues. MRTF-B for example appears to be critical for the correct patterning of the branchial arch arteries and outflow tract (Oh et al, 2005).

Myocardin, and presumably other MRTFs (through their ability to dimerise with myocardin) have also been implicated in pathological growth of the heart in response to classical hypertrophy inducing stimuli. Functional evidence for this comes from studies

in which overexpression of myocardin in isolated myocytes is sufficient to induce the expression of cardiac fetal genes that are associated with the hypertrophic response (Wang et al, 2001; Wang et al, 2002b). Within this model, myocardin also stimulated *de novo* assembly of the sarcomere and subsequent hypertrophy (Xing et al, 2006). On the other hand, a dominant-negative myocardin mutant was able to block agonist induced hypertrophy (Xing et al, 2006). It is also interesting that GSK-3 β , which is a well characterised negative regulator of myocardial growth, phosphorylates myocardin and inhibits its transcriptional activity, again providing evidence to implicate myocardin as an important modulator of pathological cardiac growth (Badorff et al, 2005).

1.5.10 Role of the myocardin family in Rho signalling

Rho signalling and associated changes in actin dynamics, both of which are key regulators of the hypertrophic response, appear to act through the MRTF/SRF signalling pathway. Early work had demonstrated that signalling by the Rho family of small GTPases could stimulate SRF activity via an actin polymerisation dependent pathway, which was itself independent of MAPK activation and the classical ternary complex SRF co-factors (TCFs)(Sotiropoulos et al, 1999). The exact mechanism for the actin-sensitivity imposed on SRF has been proposed to be due to the sequestration of the MRTFs in the cytoplasm by direct binding to monomeric G-actin via the N-terminal RPEL motifs (Miralles et al, 2003). Therefore, actin polymerisation in response to Rho signalling stimulates the translocation of the MRTFs to the nucleus where they can synergise with SRF to trans-activate downstream gene expression.

Although initially suggested as the mechanism, until recently, significant areas of ambiguity remained regarding the exact nature of the actin-sensitivity of MRTF nuclear localisation. Firstly, mutants of RhoA have been characterised that are able to stimulate SRF activity independent of stress fibre formation (Sahai et al, 1998). This suggested that among the many downstream pathways regulated by RhoA, there may be an alternative modulator of SRF activity. Secondly, the definitive mechanisms that modulated the nuclear import of the MRTFs were not precisely defined. Finally, and arguably the most important, the muscle specific regulatory nature of these MRTF/SRF

pathways was completely unknown. It was clear that this MRTF/SRF pathway was important in muscle biology, however the MRTFs and SRF are ubiquitously expressed raising the important question of how muscle specificity of this pathway is modulated. To this end, the identification of myocyte stress 1 (*msl*) by our group in 2002 has begun to answer some of these important questions (Mahadeva et al, 2002).

1.6 Myocyte stress 1

1.6.1 Initial identification and functional characterisation

Previous work in our group had developed a novel molecular indexing approach (Mahadeva et al, 1998) that was utilised for the identification of novel transcripts that are differentially expressed during the early stages of pressure-overload induced LVH (in a rat model where the aorta is banded and constricted thereby inducing pressure overload). Using this approach, Mahadeva and colleagues (1998) identified an array of genes that were differentially expressed, in particular, acutely up-regulated post-aortic banding in a pressure-overload rat model for LVH. Classically it is these acutely induced “immediate early” genes that are responsible for the transduction and subsequent propagation of transcriptional signals that ultimately culminate in the development of LVH. One of these differentially expressed acutely induced cDNA fragments was subsequently characterised and designated *msl* (Mahadeva et al, 2002). Northern blot analysis demonstrated that *msl* (which has a full length transcript size of 1.3 kb) was acutely up-regulated in the left ventricle within one hour following aortic banding, with mRNA abundance returning basal levels by four hours. This acute transient induction occurred well before any detectable change in left ventricular structure and mass. This finding, coupled to its acute transient expression profile, led us to hypothesise that *msl* represents an immediate early gene and could potentially be involved in the initial signalling of the cardiac hypertrophic response.

The rat *msl* gene locus is on chromosome 7 with the gene itself having a simple genomic organisation consisting of two exons and a single intron. The translated product is a 375-amino acid protein with a molecular weight of 42 kilo daltons.

Interestingly bioinformatic based BLAST searches revealed that the full length protein showed no significant homology to any known protein associated functional protein domains. Tissue distribution analysis of the transcript measured by semi-quantitative PCR demonstrated that *msl* was a striated muscle restricted transcript, with high levels in the heart and skeletal muscle, but at barely detectable levels in the adrenals, brain, kidney, liver and testis. It was also of interest that within the left ventricle, *msl* displayed a developmental specific expression profile: early expression was observed in the embryonic ventricle with significant up-regulation in post-natal expression through to adulthood. This implies developmental and maturation specific regulation of *msl* within the myocardium.

Using a subtractive differential cDNA cloning screen to identify novel genes expressed in the mouse heart tube at embryonic day 8.25 (E8.25) but not in other regions of the embryo, Eric Olson's group identified STARS (striated muscle activator of Rho signalling)(Arai et al, 2002), the mouse homologue of *msl* (note: on genome databases both *msl* and STARS are collectively annotated as ABRA, actin-binding Rho activator protein). The nature of this name came from this initial study in which the authors demonstrated that STARS was a sarcomeric actin binding protein which could stimulate SRF dependent transcription, via an undefined pathway that involved RhoA signalling. Subsequently, Olson's group published a study in which the molecular nature of this mechanism was identified and characterised (Kuwahara et al, 2005). STARS appeared to activate SRF-dependent transcription by inducing the nuclear localisation of MRTF-A and -B. Although not well understood this mechanism is through the attenuation of monomeric G-actin based sequestration of the MRTFs within the cytoplasm. STARS may either compete with the actin binding RPEL motifs within the MRTFs (thereby blocking actin/MRTF interactions) or alternatively drive the MRTFs into the nucleus indirectly by stabilising the F-actin and therefore decreasing the cytoplasmic F:G actin ratio.

Recent work, again emanating from the Olson laboratory, demonstrated that members of the actin-binding LIM domain-containing (ABLIM) protein family interact with MS1/STARS, associate with actin and potentiate STARS dependent SRF-transcriptional

activity (Barrientos et al, 2007). Two novel ABLIM proteins, ABLIM-2 and -3 were identified using a yeast-two-hybrid screen as muscle specific interaction partners of MS1/STARS, in addition to the previously identified ABLIM-1. The authors also demonstrated that knockdown of these ABLIMs attenuated STARS dependent SRF signalling thereby further supporting an important role for these proteins in the modulation of actin-dependent SRF activity via MS1/STARS.

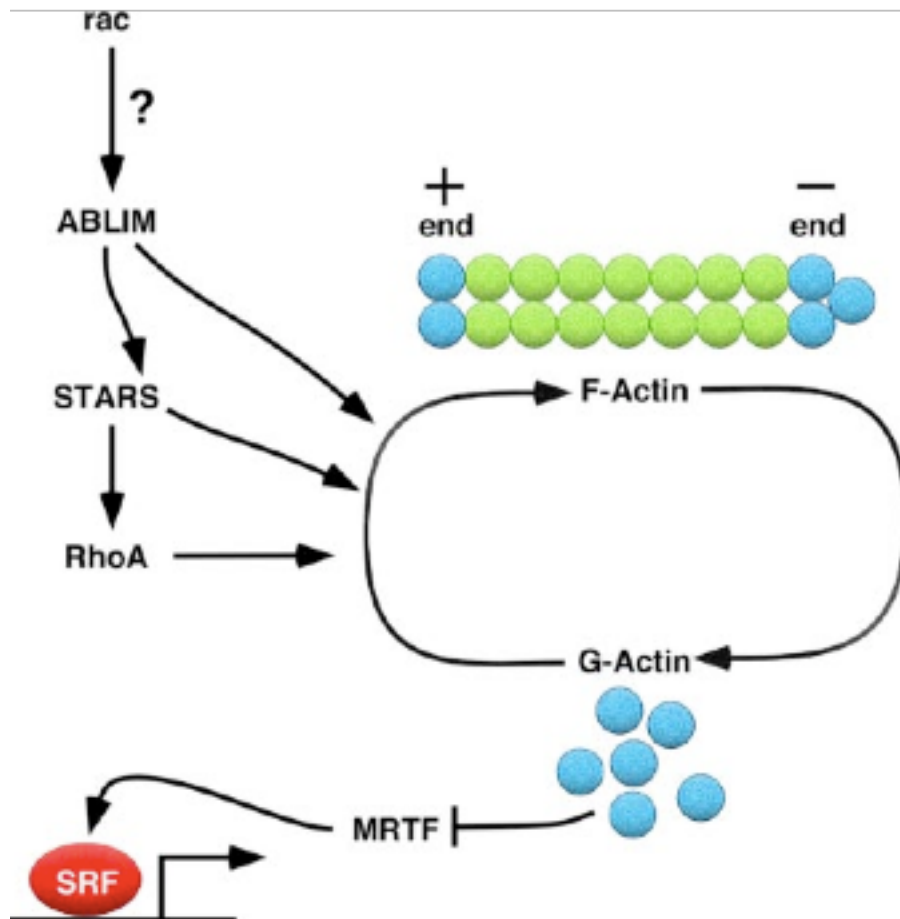


Figure 9. MS1/STARS interacts with ABLIM proteins to stimulate MRTF nuclear translocation and subsequent SRF trans-activation. (Barrientos et al, 2007)

1.6.2 Roles of *msl* in striated muscle development and disease

The immediate early expression profile exhibited by *msl* post pressure overload led our group to hypothesise that *msl* may play an important role in the early signalling that leads to the development of cardiac hypertrophy. Subsequent work in our group (Andrea L Keokemoer, PhD Thesis, 2008) has taken this early hypothesis further and demonstrated that *msl* can promote an increase in cell surface area of H9c2 myoblasts in addition to protecting against apoptosis. *Ms1* thus appears to be able to promote hypertrophy and confer cardioprotection *in vitro*, with these phenotypes associated with the up-regulation of cardioprotective and hypertrophy associated MRTF-SRF target genes. These findings were not surprising considering the well established role SRF signalling has in these processes.

In support of our *in vitro* findings, Kuwahara and colleagues have recently published an important piece of work implicating *msl* in cardiac dysfunction in both *in vitro* and *in vivo* models (Kuwahara et al, 2007). They demonstrated that STARS was up-regulated in two independent mouse models of cardiac hypertrophy as well as being substantially induced in failing human hearts. It was also shown that known SRF target fetal genes were up-regulated following *msl* transgenic over-expression *in vivo*. However, somewhat surprisingly, increased STARS expression did not induce cardiac hypertrophy *in vivo*, but when these transgenic mice were subjected to pressure overload or crossed with calcineurin transgenic mice, there was an exaggerated deterioration in cardiac function. These findings expand our *in vitro* findings in an *in vivo* model and clearly implicate dysregulation of *msl* in cardiac dysfunction and potentially the progression to heart failure (Kuwahara et al, 2007).

Additional work from our group has also implicated *msl* as a central component of striated muscle developmental pathways including differentiation and maturation. Knocking down the zebrafish *msl* homologue (*zms1*) caused a robust decrease in cardiac contractility, an enlarged atrium and curvature and shortening of the longitudinal axis as a consequence of poor skeletal myogenesis. These findings were not surprising

considering the plethora of reports implicating the MRTF/SRF axis in these processes, which in muscle is under the direct control of *msl*.

1.7 Regulation of *msl*

Despite the important role for *msl* in the regulation of the MRTF/SRF signalling axis and by association, its fundamental role in striated muscle development and disease (section 1.6.2), very little is known about its regulation. Historically gene function has been thought to reside in the protein encoded by the gene, and in this respect MS1 protein function is fairly well understood although many questions still remain unanswered. However, most genes require a network of other genes to exert their biological function, and therefore one can state that gene regulation itself is an intrinsic component of gene function. Therefore, in order to fully understand the function of *msl*, and therefore understand the nature of its role in striated muscle development and disease, one must fully understand the regulatory processes governing both its expression and activity. Gene expression, and subsequent regulation of the gene product's activity can be regulated at multiple levels; transcriptional control, post-transcriptional regulation, translational control and finally post-translational modifications. As the first step in the regulation of gene expression, transcriptional control is arguably the most important regulatory step with most cellular defects products of a defect in transcriptional control. Therefore it is important, that as the first step in the regulatory analysis of any gene (in this thesis, *msl*), one must fully elucidate the transcriptional mechanisms governing its context specific expression.

1.7.2 Transcription

Transcription is an essential process, which is required for abstract genetic information (in the form of linear nucleotide base sequence) to become a physical reality (in the form of a biologically functional protein). Transcription impinges upon all biological processes including growth, development and maturation. The transcription of DNA is carried out by the RNA polymerases, with eukaryotes having 3 RNA polymerases, which share the task of transcribing the nuclear genes. RNA polymerase II (pol II)

transcribes the majority of genes and this process requires the presence of a basal promoter. The basal promoter is composed of the core promoter, which contains the TATA box, the transcriptional start site (TSS) and the upstream regulatory elements, which are responsible for the context specific regulation of the gene (Roeder, 2005). The basal promoter interacts with distal regulatory elements to mediate the transcriptional output. Transcriptional regulatory mechanisms currently fall into three recognised categories;

- i) Transcriptional regulation in *cis*, mediated by proximal and distal *cis* regulatory elements.
- ii) Regulation in *trans*, mediated by transcriptional factors binding to *cis* regulatory elements, RNA interference and microRNA's.
- iii) Regulation on the basis of the modification state of the DNA (eg CpG methylation) and how it is packaged (eg histone acetylation and methylation).

With respect to striated muscle transcription, histone modifications and their modulatory effects on chromatin have a central role in transcriptional control.

1.7.2.1 Chromatin regulation

Chromatin describes the packaging of DNA in eukaryotic cells where nucleosomes are generated in adjacent arrays which consist of 146 bp of DNA wrapped around an octamer of four core histone proteins (H2A/H2B dimers and H3/H4 tetramer). These assemblies are then compacted into a more complex higher order structure which serves to limit DNA access to regulatory proteins including transcription factors, thereby allowing the modulation of gene expression (Jenuwein & Allis, 2001). Regions of chromatin that are actively transcribed are loosely associated with histones in a relaxed array which this process of relaxation under the direct regulatory control of histone protein tail post-translational modifications (de Ruijter et al, 2003).

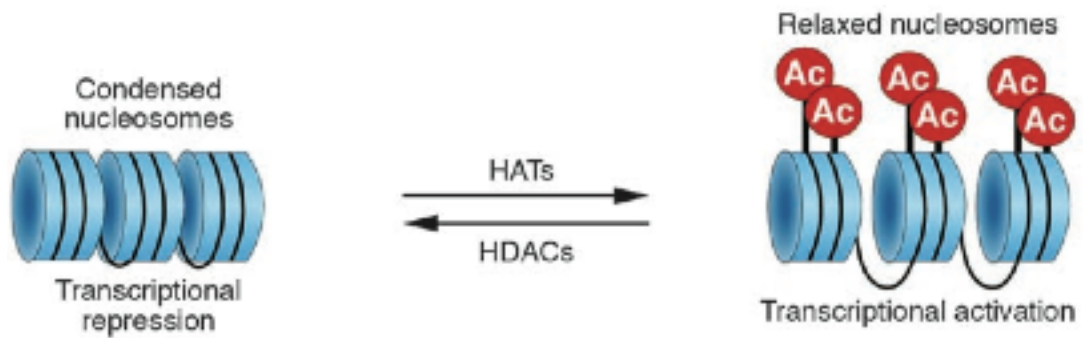
Of all these post-translational modifications, the best characterised is histone acetylation which is carried out by the histone acetyl transferases (HATs) (de Ruijter et al, 2003; Jenuwein & Allis, 2001). The best characterised HATs with respect to cardiac specific gene modulation are p300 and CREB binding protein (CBP) that interact with multiple cardiac-restricted transcription factors (Backs & Olson, 2006; Yanazume et al, 2003). In contrast, the histone de-acetylases (HDACs) remove histone acetyl moieties thereby promoting chromatin condensation and transcriptional repression (Figure 10 A). HDACs are classified into three subclasses, Class I, -II and -III, although Class-I (HDACs -1, -2, -3 and -8) and -II (HDACs-4, -5, -6, -7, -9, -10) are the most extensively characterised in mammalian cells. Targeting of both HATs and HDACs is achieved through indirect tethering to DNA regulatory elements by DNA sequence specific transcription factors and associated co-factors.

Chromatin remodelling events mediated by the HDACs have been strongly implicated in the development of cardiac hypertrophy. For example, many Class II HDACs serve to integrate various pro-hypertrophic signalling pathways through phosphorylation. Specifically, PKC, PKD and CaMK have all been identified as genuine Class II HDAC kinases that are activated by pro-hypertrophic signals (Zhang et al, 2002). Subsequent phosphorylation results in nuclear export of the Class II HDACs which abrogates HDAC dependent repression at target promoters.

In vivo, calcineurin and pressure overload induced cardiac hypertrophy are strongly associated with enhanced Class II HDAC phosphorylation, specifically HDAC-5 and -9 (Zhang et al, 2002). In support of an important regulatory role, constitutive over-expression of a nuclear retained form of these HDACs inhibited agonist-stimulated growth *in vitro* (Zhang et al, 2002). In addition, gene-targeted HDAC-5 and HDAC-9 mice develop spontaneous cardiac hypertrophy with age and enhanced hypertrophy in response to appropriate stimuli (Zhang et al, 2002). Class II HDACs are recruited to regulatory genes that modulate cardiac hypertrophy through physical interactions with cardiac restricted transcription factors such as MEF2 (Lu et al, 2000; Miska et al, 1999; Sparrow et al, 1999). Unlike class II HDACs, Class I HDACs appear to act in the opposite manner and positively modulate hypertrophic gene expression (Figure 10 B).

For example, inhibiting Class I HDAC activity through chemical perturbation significantly attenuates the hypertrophy response to prolonged pressure overload stimulation (Kee et al, 2006).

A



B

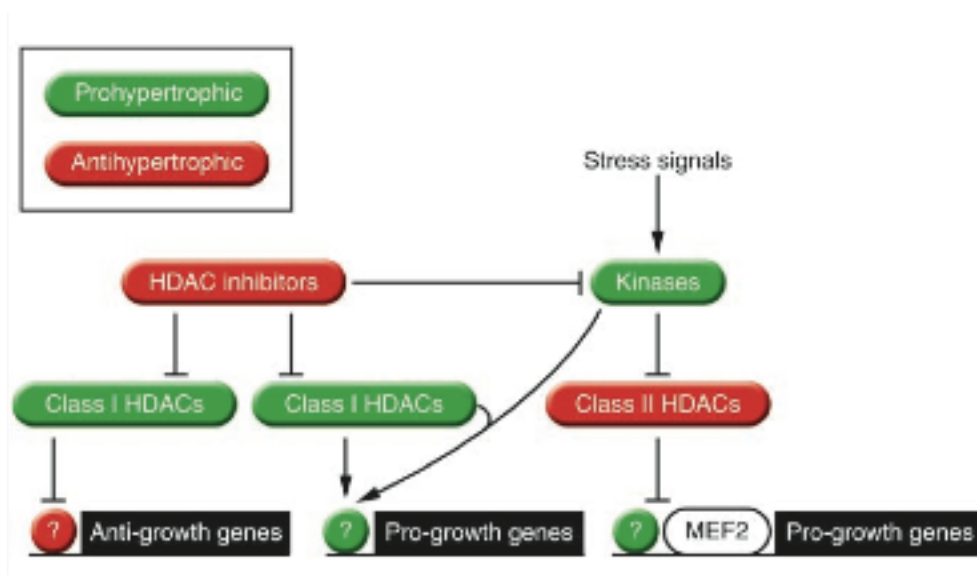


Figure 10. Regulatory roles of histone de-acetylases (HDACs) and histone acetyltransferases (HATs) in cardiac specific gene expression. (A) Schematic illustrating the actions of HATs and HDACs on chromatin. (B) Class II HDACs repress hypertrophy while Class I HDACs appear to be prohypertrophic through their independent regulatory actions on anti- and pro-hypertrophic genes (Olson, 2005).

1.8 Aims

Understanding the mechanisms governing context specific transcriptional and epigenetic modulation of *msl* is important because it will; 1) Give us a greater insight into *msl* function, remember regulation is intrinsic to function; 2) Illuminate our current understanding of the composition of the general transcriptional regulatory mechanisms/networks active in such contexts in which *msl* is differentially expressed (striated muscle development, differentiation, maturation); 3) allow the identification of potential pathways and regulatory points for therapeutic intervention and manipulation. This is of particular importance given the evidence implicating *msl* dysregulation with cardiac hypertrophy and cardio-myopathy; 4) Allow us to start building accurate gene regulatory networks furthering our understanding of these processes at a systems level.

Therefore, for the purpose of this study we propose the following specific aims;

1. To execute a thorough computational analysis of the *msl* genomic locus and landscape generating an array of potential *cis* regulatory domains and motifs for experimental characterisation.
2. Experimentally interrogate *in silico* defined CRMs, TFBS and cognate binding factors which confer robust striated muscle specific gene expression.

2.1 Cardiac specific characterisation

2.2 Skeletal muscle specific characterisation

3. Interrogate the mechanisms governing the context specific induction in *msl* expression

3.1 Developmental induction within the heart and skeletal muscle.

3.2 Post natal induction in response to cardiac stress.

Chapter 2

Materials and Methods

2.1 *Materials*

2.1.1 *Plasmids and reagents for molecular biology*

All general chemicals and reagents were of analytical grade and purchased from Sigma-Aldrich. All solutions used for RNA extraction were prepared using sterile DEPC treated water purchased from Invitrogen. The pGL3-Basic luciferase vectors, pGEM-T Easy vector, Steady-Glo™ Luciferase Assay System (Dual Luciferase) were purchased from Promega (Table 1). ABI Prism BigDye™ terminator ready reaction mix and general real-time components (optical seals, 96 well plates) provided by Applied Biosystems. Jet Pei transfection reagent was obtained from QBiogene. The QuickChange™ site-directed mutagenesis kit was obtained from Stratagene. The QIAquick gel extraction kit and plasmid Mini/Midi Prep kits were purchased from Qiagen. Over-expression plasmids were obtained from various sources as described in Table 2. The Superscript II RT-PCR kit, T4 DNA ligase, RNase-free DNase, synthesised oligonucleotides and deoxynucleotide (dNTPs) were purchased from Invitrogen. Restriction enzymes and buffers were obtained from New England Biolabs. The EMSA reagents and buffers were all components of the DIG Gel Shift Kit from Roche. All antibodies used in chromatin immunoprecipitation assays were provided by Abcam. Other reagents not outlined here were from suppliers indicated throughout this chapter.

2.1.2 *Reagents for Cell Culture*

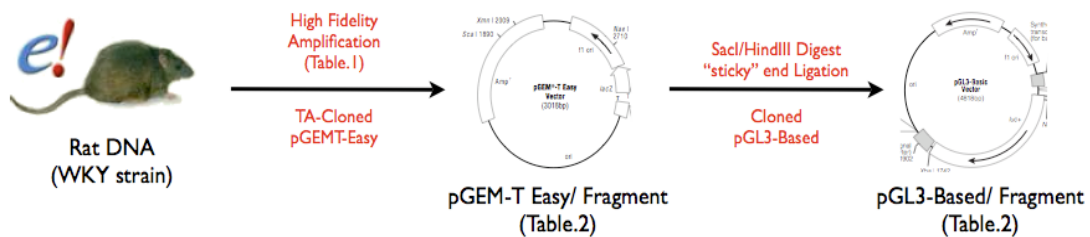
H9c2 cardiac myoblasts, C2C12 skeletal myoblasts, NIH 3T3 and Cos-7 fibroblasts were all purchased from European Collection of Cell Cultures (ECACC, EU). Cell culture media, L-glutamine, penicillin/streptomycin, fetal bovine serum and fetal horse serum were all purchased from Gibco. Cells were grown in sterile 6-well plates, 6mm and 10mm petri dishes, 25cm³ and 75cm³ flasks (Nunc).

2.2 DNA Manipulation

2.2.1 High Fidelity Polymerase Chain Reaction (PCR)

The *Rattus norvegicus* (rat) *ms1* genomic locus was located from GenBank (NC_005106) and used to design primers (using Primer 3 software) that would span and amplify the 5'-flanking region and the *in silico* defined (see Chapter 3) putative *cis* regulatory modules (PCRM, UP2 and UP3). The primers were tailed with restriction sites for *SacI* (forward primers) and *HindIII* (reverse primers)(Table 3), thereby allowing for directional cloning into the luciferase reporter vectors, pGL3-Basic and -Promoter (Table 2)(Figure.1).

Figure.1; A diagram outlining the cloning strategy (and vectors used) for the generation of promoter reporter luciferase constructs.



100ng of rat (WKY strain) genomic DNA was amplified using the Roche Expand High Fidelity PCR system and the appropriate primer pair combinations (Table 3). In addition to the template (WKY rat genomic DNA) and primers (500nM each primer)(Table 3), the reaction contained 0.2mM dNTP's, Roche Expand Polymerase buffer and 5 units of *Taq* Expand high fidelity polymerase. The reaction was typically made up to a final concentration of 50ul with milli-Q H₂O. Two drops of oil were used to seal each reaction volume and the tubes were then placed in a Techne Techgene thermal cycler. Samples were subjected to 35 cycles of amplification (45 seconds at 94°C, 45 seconds at 59°C and 90 seconds at 72°C), with a final terminal extension step performed at 72°C for 5 minutes. The amplified PCR products were electrophoresed alongside appropriate DNA ladder (100bp and 1kbp ladder, New England Biolabs) on a 2% (w/v) agarose gel (section 2.2.3).

Table 1 Cloning and reporter vectors.

Vector	Source
pGEM®-T Easy	Promega
pGL3-Basic	Promega
pGL3-Promoter	Promega
pTL-TK	Promega
ANF-pGL3-638	Junichi Sadoshima, New Jersey Medical School, USA
NFAT-RE-pGL3	Jeffery Molkentin, Cincinnati Childrens Hospital Medical Center, USA

Table 2 Protein over-expression vectors used and their source.

Protein	Vector	Source
No protein	pcDNA 3.1(+)	Invitrogen
p300	pCMV β	Toru Kita, Kyoto University, Japan
Δ p300	pCMV β	Toru Kita, Kyoto University, Japan
HDAC4	pCGN	Stuart L Schreiber, Harvard University, USA
HDAC5	pCGN	Stuart L Schreiber, Harvard University, USA
GATA-4	pCF	Mona Nemer, IRCM, Canada
Δ GATA-4	pCF	Mona Nemer, IRCM, Canada
MEF2-A	pCGN	Mona Nemer, IRCM, Canada
Δ Calcineurin	pcDL	Jeffery Molkentin, Cincinnati Childrens Hospital Medical Center, USA
MEF2-D	pcDNA 3.1(+)	Eric Olson, Southwestern University, USA
MyoD	pEMSV	Anthony Imbalzano, UMASS, USA
Myogenin	pEMSV	Anthony Imbalzano, UMASS, USA

Table 3 Sequences of oligonucleotides used for the generation of ms1 promoter and cis regulatory domain luciferase reporter constructs.

Oligonucleotide Name	Sense/ Antisense	Sequence (5'-3')
MS1/-1585	Sense	TATTCAATGCTTAGTCCTGC
MS1/-1381	Sense	CAACTGATTGATGAGCTTGTC
MS1/-1148	Sense	GAGATATATGGGAGGTTGTAG
MS1/-365	Sense	TTACAGAGGTTTAAGTGAGAGC
MS1/-300	Sense	CGGAGCTCATGATTCAATCTAGTACTTTCC
MS1/-127	Sense	CGGAGCTCAGAACACCGTCAGAGCCATAG C
MS1/+60-HindIII	Antisense	CCAAGCTTCAGGCTACCTGTTTCTTCTC
MS1/-16702	Sense	CGCTCGAGGCCCCCACGAGTTAAAGCTCA
MS1/-16432	Antisense	TGGGCTGTGTGGGTTAGGAGA
MS1/-9498	Sense	TTGCTATGGGTGGGGAAGTGG
MS1/-8608	Antisense	GGCTCGAGGGGAAGGTTGGTCAGGAAAA GG

2.2.2 *Standard PCR*

Standard PCR was used to amplify DNA fragments for diagnostic or semi-quantitative analysis (for example, diagnostic PCR of recombinant plasmids during cloning process, semi-quantitative amplification of chromatin immunoprecipitation derived DNA samples). PCR reactions were set up using 100-500ng DNA sample, 1X Abgene buffer, 1.5mM MgCl₂, 0.2mM dNTP, 500nM forward primer, 500nM reverse primer and 1.25 Units of Abgene *Taq* polymerase, in a final volume of 50µl. Primers used are listed in Table 3 and 10. The thermocycling parameters were 1 cycle for 94°C; 25-30 cycles 94°C for 45 seconds, 59°C for 45 seconds; 72°C for 45 seconds; and a final extension step of 72°C for 5 minutes.

2.2.3 *Agarose gel electrophoresis*

1X loading buffer [40% (v/v) glycerol, 60% (v/v) TE Buffer (Tris Ethylene diamine tetra-acetic acid (EDTA), 10 mM Tris-HCL, 1mM EDTA, pH 8.0] and 1X bromophenol blue were added to the DNA samples of interest. Samples were then pulse-spun in a centrifuge (Eppendorf Minifuge, 20 seconds, 12,000 rpm) and loaded onto the appropriate percentage agarose gel in 1X Tris-Acetate EDTA (TAE) buffer (400mM Tris-HCL, 20mM glacial acetic acid, 0.1mM EDTA, pH8.0) and run alongside an appropriate sized marker (Promega, 100bp and 1kbp). The gel contained ethidium bromide added at a concentration of 0.035µg/ml. The gel was typically electrophoresed in 1X TAE buffer for 1 hour at 80 V. The gel and associated migrated bands were then visualised on ultraviolet light using a Syngene gel documentation and imaging system.

2.2.4 *DNA extraction and purification from agarose gel*

The required electrophoresed DNA fragments were extracted (through the use of a scalpel) from the agarose gel and purified using the QIAquick gel extraction kit (Qiagen), as described in the suppliers handbook (Qiagen). Purified DNA was typically quantified and stored at -20°C.

2.2.4.1 DNA determination

DNA concentration was determined using the NanoDrop™ (ND-1000, Thermo-Scientific) spectrophotometer, with absorbance measured at wavelengths of 260nm and 280nm. The DNA purity was determined by the ratio of the absorbance at 260nm/280nm.

2.2.5 Ligation of isolated DNA into vector

The appropriate amount of insert DNA was placed into the ligation reaction with 100ng of the vector (pGEM-T Easy, pGL3-Based, Table.1). Typically the appropriate amount of insert is calculated so that there was a 3:1 insert to vector ratio; $(\text{vector, ng} \times \text{insert size, kb}) \div (\text{size of vector, kb} \times \text{insert:vector ratio}) = \text{ng of insert}$. The reaction also contained 3U of T4 DNA Ligase (Invitrogen), 1X TD DNA ligase buffer and H₂O, to make final volume 10µl. A control reaction is also established where the insert DNA fragment is omitted, thereby allowing one to determine the presence of re-circularised vector plasmid. Both reactions were incubated overnight at 4°C.

2.2.6 Transformation of bacteria with plasmid DNA

Typically 5µl of the ligation reaction was added to 100µl DH5α chemically competent cells (Invitrogen), mixed gently and incubated on ice (or at 4°C) for 30 minutes. Heat shock treatment at 37°C for 45 seconds was executed using pre-calibrated water bath. The reaction was then allowed to recover on ice for 2 minutes prior to the addition of 400µl pre-warmed SOC, a high nutrient broth media. The reaction was then incubated for exactly 1 hour at 37°C in a rotating incubator (at 225 rpm). Following this 1 hour incubation step, half of the reaction volume was extracted and spread onto a Luria broth (LB) agar plates containing Ampicillin (100µg/ml) for transformed cell selection. Where blue/white selection of transformants was necessary (pGEM-T Easy sub-cloning), 40µl X-Gal (20mg/ml) was added to the LB agar plates. Plates were inverted and incubated for 16 hours at 37°C.

Subsequently, single isolated transformant colonies were picked and grown in 10ml of LB (containing the appropriate selection antibiotic) for 16 hours in a rotating incubator (at 225 rpm) at 37°C. Using the QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's instructions, plasmid DNA for diagnostic analysis (diagnostic digest) was extracted from 6ml of the 16 hour growth culture. For positive clones, their remaining 4ml of culture was used to make a 20%(v/v) glycerol stock for long-term storage at -80°C and the rest was used to establish a secondary culture (approximately 10-15ml) to extract plasmid DNA for the purpose of further cloning, sequencing or transfection. The plasmid DNA was extracted using the Qiagen-tip20 Plasmid Mini Kit, following the manufacturer's instructions.

2.2.7 Restriction Enzyme Digest

Plasmid DNA was digested using 10U of the required restriction endonuclease (Invitrogen) in a final volume of 20µl, which also contained 1X appropriate restriction endonuclease buffer (Invitrogen). The digests were incubated for 2 hours at 37°C therefore ensuring complete digestion. Double digests were set up in compatible buffers where appropriate.

2.2.8 Site Directed Mutagenesis

Site-directed mutagenesis (SDM) was performed using the Quickchange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. Primers were designed using the Stratagene online SDM primer design software to introduce single or multiple changes at desired sites. The PCR reaction contained 50ng dsDNA template, 125ng sense oligonucleotide (Table D), 125ng antisense oligonucleotide (table 4), 1X reaction buffer, 1µl dNTP (10mM) and 0.05U *Pfu* DNA polymerase in a final volume of 50µl. PCR was carried out for 1 cycle at 95°C for 30 seconds, followed by 16 cycles, at 95°C for 30 seconds, 55°C for 1 minute and 68°C for 12 minutes. Following PCR, the reaction was incubated for 2 minutes at 4°. PCR products were then digested with 10 U *DpnI* restriction endonuclease for 1 hour at 37°C. This *DpnI* treated DNA (1µl) was then used to transform XL-1 Blue Supercompetent cells (Stratagene).

2.2.10 DNA Sequencing

Cloned recombinant plasmids, over-expression plasmids, and mutant plasmids (Table 5) were sequenced using the ABI Prism BigDye™ Terminator Ready Reaction Mix (ABI) and sequencing primers shown in Table 5. A PCR reaction mixture containing 8µl of the BigDye Terminator Ready Reaction Mix, 1µg of plasmid DNA and 5 pmoles of sequencing primer (Table 5) was prepared in a final volume of 20µl. Reactions were amplified (Techne Techgene thermal cycler) using thermocycling parameters that consisted of an initial 96°C for 3 minutes, followed by 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes. Thereafter, PCR products were precipitated by mixing with 50µl of 95% (v/v) ethanol and 2µl sodium acetate (pH8.0) and incubated at room temperature for 15 minutes. Precipitated DNA pellet (pelleted by centrifugation, 20,000 g for 20 minutes) was then washed with 250µl 70% (v/v) ethanol. DNA was re-pelleted after washing by a further centrifugation for 5 minutes at 20000 g, and air dried on the bench at room temperature for approximately 10 minutes. Sequencing of purified DNA was then performed using the 3730xl DNA Sequence analyser (Applied Biosystems). The data was analysed using Chromas (University of Leicester, UK) and Sequencher DNA analysing sequence software.

Table 4 Sequences of oligonucleotides used in site-directed mutagenesis to create MS1 promoter mutant constructs. Mutated or introduced nucleotides are in bold.

Oligonucleotide Name	Sense/ Antisense	Sequence (5'-3')
ΔTATA-Fw	Sense	CACCCTTTCACACCCTGCTTCTGTTTAAATCCCAGGC AACTC
ΔTATA-Rv	Antisense	GAGTTGCCTGGGATTTAAACAGAAGCAGGGTGTGAA AGGGTG
ΔE1-Fw	Sense	CACTGAACAGGTGCTGTTTCTCTGTC GTTA AGACTTA TCCTTTCAGTTCTCTTAAAA
ΔE1-Rv	Antisense	TTTAAAGAGAACTGAAAGGATAAGT CTTA ACGACAG AGAAACAGCACCTGTTCACTG
ΔE2-Fw	Sense	CTTCCACCCTGGC GCG GGGAGAAGAAAGGAG
ΔE2-Rv	Antisense	CTCCTTTCTTCTCCCCGCGCCAGGGTGGAAAG
ΔE3-Fw	Sense	CAAGGAAAACATAAAGCTAAGC GCG GGATTCAATCT AGTACTTC
ΔE3-Rv	Antisense	GAAGTACTAGATTGAATCCCGCGCTTAGCTTTATGTT TTCCTG

Table 5 Sequencing primers

Oligonucleotide Name	Sense/ Antisense	Sequence (5'-3')
RVprimer3	Sense	CTAGCAAAATAGGCTGTCCC
GLprimer2	Antisense	CTTTATGTTTTTGGCGTCTTCCA
M13	Sense	GTTTTCCTCAGTCACGAC

2.3 *Cell Culture*

2.3.1 *Maintenance of cell lines*

All of the cell lines (H9c2, C2C12, NIH 3T3 and Cos-7, see section 2.1.2) were routinely cultured as monolayers in Dulbecco's modified Eagles medium (DMEM) containing glutamax-1, which was supplemented with 100 units/ml streptomycin, 100 units/ml penicillin and 10% (v/v) foetal bovine serum (for C2C12 see section 2.3.3.2), in a humidified atmosphere at 37°C composed of 5% CO₂. H9c2 and C2C12 cells were typically passaged upon reaching 70-80% confluence thereby ensuring the preservation of myoblast cellular phenotype. NIH-3T3 and Cos-7 cells were passaged upon reaching confluence. To passage, cells were washed once with pre-warmed PBS and incubated with a trypsin-EDTA solution (0.2% trypsin, 1mM EDTA) for 2-4 minutes. The flasks were then gently agitated to disrupt cell adhesion, before re-suspending at a 1:5 dilution in fresh pre-warmed media.

2.3.2 *Storage of cells*

For long term storage, cells were frozen in liquid nitrogen, while for short term, cells were stored at -80°C. To store, cells were washed once with pre-warmed PBS and incubated with a trypsin-EDTA solution (0.2% trypsin, 1mM EDTA) for 2-4 minutes. The flasks were then gently agitated to disrupt cell adhesion. Before re-suspending in 1 ml cell culture media, an equal amount of cell culture media for storage was prepared was mixed with 20%(v/v) DMSO. Cell's re-suspended in this DMSO containing culture media were then aliquoted and frozen slowly at -80°C by placing in an iso-propanol containing insulated box. 24 hours later these cell aliquots were transferred to a liquid nitrogen cell bank for long-term storage.

2.3.3 *Cardiogenic and myogenic differentiation of cell lines*

2.3.3.1 *H9c2 cardiogenic differentiation*

H9c2 were grown as a stock of cells in T-75 culture flasks (section 2.3.1). Before reaching confluency cells were passaged at a lower density in petri dishes and cultured for a further day with standard media (10% FBS) under standard conditions. On the second day, cells were cultured in DMEM supplemented with 1% FBS with the culture media being changed every two days. Stimulation with retinoic acid analogues was used to initiate the cardiac differentiation programme (Menard et al, 1999). Stimulation with all-trans-retinoic-acid (ATRA from Sigma) was daily. ATRA was diluted in DMSO to make a 1mM stock solution which was stored at -20°C. Aliquots were only used once and diluted extemporaneously in 1% FCS culture media. All steps involving the preparation of ATRA and ATRA-containing solutions were carried out in the dark due to the light sensitivity of ATRA.

2.3.3.2 C2C12 myogenic differentiation

C2C12 mouse skeletal myoblasts were cultured in DMEM supplemented with 20% FBS, 2mM glutamine, streptomycin and penicillin (each at 100 units/ml). Once the myoblasts reach confluence, myogenic differentiation is achieved by the addition of differentiation media which contains 2% horse serum instead of 20% FBS. The confluent myoblasts were then left to differentiate into myotubes for up to 4 days with the media was changed every other day.

2.3.4 Trichostatin A treatment of H9c2 cells

For both trichostatin A (TSA) and doxorubicin treatment approximately 40,000 H9c2 cells were plated into each well of a six-well plate. 24 hours later these wells were approximately 60% confluent and ready for drug stimulation. Cells were incubated with 100nM or 300nM TSA (sigma) for 48 hours prior to RNA extraction. TSA was dissolved in ethanol to make a 1mM working stock which was diluted extemporaneously in 10% FBS DMEM prior to each stimulation.

2.3.5 Simulated sub-lethal ischemia/reperfusion

Sub-lethal ischemia/reperfusion was simulated using an ischemic buffer system (Punn et al, 2000). 100,000 H9c2 cells were seeded into individual 6cm plates in 10%FBS DMEM. Once 60-70% and 16 hours prior to ischemia, media was changed to 2% FBS DMEM. In order to simulate sub-lethal ischemia the media was removed and replaced with a modified Krebs buffer containing 4mM Hepes, 3.58mM KCl, 0.9mM CaCl₂, 137mM NaCl, 0.49mM MgCl₂, 2% (v/v) FBS containing specific compounds that induce the ischemic stress (16mM KCl, 20mM sodium lactate, 10mM 2-deoxyglucose and 1mM sodium dithionite). The ischemic buffer was adjusted to pH 6.5. The cells were typically made ischemic for 1 hour. Reperfusion was induced by removing the ischemic buffer, washing the cells in pre-warmed PBS and replenishing with 2% FBS DMEM medium.

2.4 RNA manipulation

2.4.1 RNA isolation from cells and tissues

RNA was extracted from adult and embryonic rat tissues using the RNeasy Maxi kit (Qiagen) by following the manufacturer's protocol. Total RNA was isolated from cell lines using the RNeasy Mini kit (Qiagen) according to the manufacturer's protocol. Briefly, media was removed and cells were washed twice in pre-warmed PBS. 350µl of RNeasy RLT lysis buffer is then added to the cells. Lysate was collected and RNA extraction was executed as described in the RNeasy Mini kit handbook (Qiagen).

2.4.1.1 RNA determination

RNA concentration was determined using the NanoDrop™ (ND-1000, Thermo-Scientific) spectrophotometer, with absorbance measured at wavelengths of 260nm and 280nm. The RNA purity was determined by the ratio of the absorbance at 260nm/280nm.

2.4.2 Deoxyribonuclease (DNaseI) treatment of RNA

1µg of total RNA was incubated with DNaseI (Sigma) according to the manufacturer's instructions (on column digestion). The DNaseI treated RNA was then reverse transcribed as described in section 2.4.3.

2.4.3 cDNA synthesis

1 µg of DNaseI treated total RNA was incubated with 0.5µg oligo dT, and 0.5mM dNTP mix for 10 minutes at 70°C in a Perkin Elmer Cetus DNA thermal cycler. The reaction was then placed on ice for 2 minutes and briefly centrifuged to remove condensation. This reaction was then supplemented with 1X First strand buffer (Invitrogen), 10mM DTT and 200 U Superscript II reverse transcriptase (RT) enzyme (Invitrogen) giving a total reaction volume of 20µl. Control reactions (minus RT) were also established that contained equal volume of H₂O instead of Superscript II RT enzyme. The reactions were then incubated in the thermal cycler at 42°C , for 55 minutes; 70°C for 10 minutes and then finally back on ice for recovery. The subsequent cDNA was typically diluted 1:5 with DEPC treated H₂O, with 1µl used for down-stream PCR amplification.

2.4.4 Semi-quantitative RT-PCR

A typical reaction would contain 1µl template cDNA (1:5 dilution from RT reaction), 500nM forward primer, 500nM reverse primer, 1X Abgene buffer, 1.5mM MgCl₂, 2mM dNTP mix and 1.25 U Abgene Taq polymerase. Mili-Q H₂O is used to make up the final reaction volume to 20µl. Typically a “no template” control was also included where H₂O is used instead of cDNA template. The primer sequences of the genes analysed by SQ-PCR and the optimised cycle number for there linear amplification can be found in Table 6. Ribosomal protein L32 (RPL32) was used as the internal control. The thermocycling parameters were 1 cycle for 94°C; (see Table 6 for the number of cycles) of 94°C for 45 seconds, 59°C for 45 seconds; 72°C for 45 seconds; and a final extension step of 72°C for 5 minutes.

Samples were then electrophoresed as described in section 2.2.3. Gel images were obtained and documented directly onto a PC and the relative mRNA abundance was determined by band quantification (using tools from the Syngene software) and corrected for RPL32, serving as the internal control.

Table 6 Sequences of oligonucleotides used for semi-quantitative PCR and optimised cycle number for each gene.

Oligonucleotide Name	Sense/ Antisense	Sequence (5'-3')	Cycle Number
MS1-Fw	Sense	GTGACAGCATAGACACAGAGGAC	28
MS1-Rv	Antisense	CACTGCTGCCACCTGCCTT	28
GATA4-Fw	Sense	CTGCGGCCTCTACATGAAGC	29
GATA4-Rv	Antisense	CTGAATGTCTGGGACATGGA	29
MEF2C-Fw	Sense	AGATAGTGTTCATGTTGCAGGTTCA	26
MEF2C-Rv	Antisense	CTCATGGCTTAGGGATGTGCTTTC	26
SRF-Fw	Sense	GCATTCACAGTCACCAACCTGC	26
SRF-Rv	Antisense	TCATTCACTCTTGGTGCTGTGGG	26
RPL32-Fw	Sense	GTGAAGCCCAAGATCGTC	19
RPL32-Rv	Antisense	GAACACAAAACAGGCACAC	19
α -1-CLTCC-Fw	Sense	GGAGAGTTTTCCAAAGAGAGG	26
α -1-CLTCC-Rv	Antisense	GATCACCAGCCAGTAGAAGAC	26
Per1-Fw	Sense	AGTACGCGCTGGCCTGTGTCAA	25
Per1-Rv	Antisense	GGCGCTTCATAACCGGAGTGG	25
MyoD-Fw	Sense	CGGCGGCAGAATGGCTACGACACC	24
MyoD-Rv	Antisense	CACTGCATTCTAGTTGTGGTTTGT	24
α -1-SLTCC-Fw	Sense	CGCGAGGTCATGGACGTGGAG	26
α -1-SLTCC-Rv	Antisense	GATCACCAGCCAGTAGAAGAC	26

2.4.4.1 Statistical analysis

Statistical analysis was carried using a one-way analysis of variance (ANOVA). Typically, a *P* value <0.05 was considered to be statistically significant.

2.4.5 Real-time quantitative RT-PCR using TaqMan® gene expression assays

Multiple reactions were typically set up as a master mix and a typical 20µl reaction would contain:

- 1µl Template cDNA
- 1µl 20X Pre-formulated assay mix (Applied Biosystems) which itself contains 0.9mM forward and reverse primers coupled to 250nM FAM™-dye labelled TaqMan® MGB probe.
- 10µl 2X TaqMan® Universal PCR master mix (Applied Biosystems)

d H₂O to a final volume of 20µl.

For TaqMan® gene expression analysis TATA-binding protein (TBP) was used as reference gene to normalise mRNA abundance between different samples. The amplification reaction was executed in a ABI Prism® 7900HT sequence detection system. The cycling conditions were 50°C for 2 minutes (holding step), 95°C for 10 minutes (holding step) and then 40 cycles of 95°C for 15 seconds (denaturation) and 60°C for 1 minute (annealing and extension step). The data was automatically sorted and analysed using the comparative $\Delta\Delta C_T$ method (Livak & Schmittgen, 2001). This method allows the quantitative determination of fold induction of gene of interest between different samples using the following formula;

$$\text{Fold induction} = 2^{-\Delta\Delta C_T}$$

$$\Delta C_T = \text{mean } C_{T(\text{gene of interest})} - \text{mean } C_{T(\text{reference gene})}$$

$$\Delta\Delta C_T = \Delta C_{T(\text{calibrator})} - \Delta C_{T(\text{unknown})}$$

The calibrator sample, which is typically represented by the non-treated or basal control sample allows for corrections to be made for inter-assay variation. It is important to note that the above formula is based on the assumption that the efficiency of the PCR reaction for both the reference internal control gene, and the gene of interest is identical with a doubling of product being achieved with every cycle.

2.4.6 Real-time quantitative RT-PCR using SYBR® Green and analysed using the Pfaffl method

Relative mRNA abundance was determined through the use of the Pfaffl method for relative mRNA quantification (Pfaffl, 2001). Multiple reactions were typically set up as a master mix and a typical 25µl reaction would contain:

1µl	Template cDNA
0.5µl	10µM Forward primer
0.5µl	10µM Reverse Primer
12.5µl	2X Power SYBR® Green PCR Master Mix (Applied Biosystems)

d H₂O to a final volume of 25µl.

For SYBR® Green Pfaffl based gene expression analysis RPL32 was used as reference gene to normalise mRNA abundance between different samples. The amplification reaction was executed in an ABI Prism® 7900HT sequence detection system. The cycling conditions were 95°C for 45 seconds, 62°C for 45 seconds, 72°C for 45 seconds and a final step at 95°C for 15 seconds and 60°C for 1 minute. The primer sequences for the genes of interest are available in Table 6. A series of cDNA dilutions were generated thereby allowing the generation of a standard curve for the reference gene and gene of interest by plotting the log of input cDNA versus C_T. The standard curve allows the efficiency (E) of amplification of each primer set to be calculated according to the equation:

$$E=10^{[-1/\text{slope}]}$$

Contrary to the relative $\Delta\Delta C_T$ method, the amplification efficiency of the reference gene and target gene do not have to be equal. This is because their actual efficiencies are incorporated into the analysis formula, thereby increasing the accuracy of the final relative fold induction value's. The relative expression is therefore determined as follows:

1. Mean C_T values are extrapolated for calibrator samples (typically non treated control) and samples for the reference gene and gene of interest.
2. Gene of interest ΔC_p = mean gene of interest. C_T (calibrator) - mean gene of interest. C_T (sample).
3. Reference gene ΔC_p = mean reference gene. C_T (calibrator) - mean reference gene. C_T (sample).
4. The values for efficiency for each genes PCR amplification is then raised to the power of the respective ΔC_p , and the gene of interest:reference gene ration is determined using the forumla:

$$\text{Ratio (fold change)} = \frac{\text{Efficiency (gene of interest)}^{\Delta C_p \text{ (gene of interest)}}}{\text{Efficiency (reference gene)}^{\Delta C_p \text{ (reference gene)}}$$

It is important to note that this formula makes the assumption that while the efficiencies of the reference and gene of interest may be different, the efficiency of amplification for each sample within a PCR reaction is the same.

2.4.7 Statistical analysis

Statistical analysis was carried using a one-way analysis of variance (ANOVA). Typically, a P value <0.05 was considered to be statistically significant.

Table 7 **TaqMan® FAM™ labelled probes used for quantitative real-time PCR**

Probe Target Gene	Species	Assay_ID
Myocyte stress 1 (ABRA)	Rat	Rn00598518_m1
Apoptotic repressor with CARD domain (nol3)	Rat	Rn01431482_g1
TATA binding protein (TBP)	Rat	Mm01277045_m1

Table 8 **Sequences of oligonucleotides used for quantitative real-time PCR**

Oligonucleotide Name	Sense/ Antisense	Sequence (5'-3')
MS1-Fw	Sense	GTGACAGCATAGACACAGAGGAC
MS1-Rv	Antisense	CACTGCTGCCACCTGCCTT
JunB-Fw	Sense	ATCAGACACAGGCGCATCTC
JunB-Rv	Antisense	TCTTGTGCAGGTCGTCCAG
Fra I-Fw	Sense	CAGGCGGAGACCGACAAGTT
Fra I-Rv	Antisense	TGCAGTGCTTCCGGTTCAA
EF1- α -Fw	Sense	AGCTTCTCTGACTACCCTCCACTT
EF1- α -Rv	Antisense	GACCGTTCTTCCACCACTGATT
RPL32-Fw	Sense	GTGAAGCCCAAGATCGTC
RPL32-Rv	Antisense	GAACACAAAACAGGCACAC

2.5 *Transient transfection analysis*

2.5.1 *Transient transfection of cell lines*

Approximately 40,000 H9c2/C2C12 cells/well or 30,000 NIH3T3/COS-7 cells/well of a 6 well plate were seeded and incubated for 24 hours prior to transfection. Transfections were carried out using the cationic lipid transfection reagent JetPei (QBiogene), according to the manufacturer's instructions. Cells were co-transfected with 0.5µg of promoter-reporter luciferase construct (Table 1 and Table 9) and with equal-molar amounts of the desired over-expression plasmid (Table 2). Total amount of DNA transfected was always kept constant (1.5µg maximum amount transfected) using an empty vector, typically pcDNA3.1. To normalise for transfection efficiency, the pRL-TK (promega) expression plasmid containing *Renilla* luciferase (20ng per well) was co-transfected. *Firefly* and *Renilla* luciferase activities were measured 48 hours post-transfection using the Dual-Glo™ luciferase assay system (Promega) as described in section 2.5.2.

All transfections within each experiment were performed in triplicate, with each experiment typically performed on a minimum of at least three separate occasions. Data from transfections is expressed as means ± standard error of mean (S.E.M) relative to promoter construct specified, and differences between samples were detected using a one-way ANOVA, with $P < 0.05$ considered to be statistically significant. For non-luciferase reporter transfections, 6 well plates were seeded as before and transfected with a maximum of 1.5µg over-expression plasmid. 48 hours post transfection total RNA was isolated from cell as described in section 2.4.1.

2.5.2 *Luciferase reporter gene assay*

Firefly and *Renilla* luciferase activities were measured 48 hours post transfection using the Dual-Glo™ luciferase assay system (Promega) and a Luminat LB9507 luminometer (Berthold Technologies). The assay was executed as described in the users' handbook (Dual-luciferase assay handbook, Promega). Briefly, cells in the 6 well plates were

lysed using the luciferase cell lysis buffer (CLB, Promega). The first luminescence from the firefly luciferase (representing promoter reporter activity) was measured by adding 100µl of LARII substrate into 40µl of cell lysate in a fresh luminometer tube. The second luminescence for Renilla luciferase (representing the internal control activity) was quantified by the addition of 50µl of Stop and Glo substrate to quench the first reaction and simultaneously initiate *Renilla* luciferase reaction. Data was then extrapolated as relative luciferase activity, the ratio of the first luminescence over the second *Renilla* luminescence.

Table 9 **ms1 promoter reporter luciferase constructs including wild type and mutant promoter constructs.**

Promoter Reporter	Insert Size (bp)
-1585/+60-Luc	1645
-1381/+60-Luc	1441
-1148/+60-Luc	1208
-365/+60-Luc	425
-300/+60-Luc	360
-127/+60-Luc	187
-16702/-16432 -Luc	270
-9498/-8608-Luc	890
-1585/+60- Δ TATA-Luc	1645
-1585/+60- Δ E1-Luc	1645
-1585/+60- Δ E1- Δ E2-Luc	1645
-1585/+60- Δ E1- Δ E2- Δ E3-Luc	1645

2.6 Electromobility shift assay (EMSA)

2.6.1 Preparation of whole cell protein extracts.

CellLytic™-M cell lysis extraction reagent (Sigma) was used to extract whole cell protein extracts from H9c2 and C2C12 myoblasts, following the manufacturer's instructions. In order to inhibit endogenous protease activity, 1X complete mini protease inhibitor cocktail (Roche) was added to the reagent.

2.6.2 Protein quantification

Protein was quantified using the Bio-Rad Protein Assay (Bio-Rad) kit as instructed by the manufacturer. This is based on the Bradford assay. Briefly, bovine serum albumin (BSA) was used to establish the following standard protein concentrations, 0.2µg/µl, 0.4µg/µl, 0.6µg/µl, 0.8µg/µl and 1.0µg/µl in a final volume of 20µl. To this 1ml of diluted Bio-Rad dye (diluted 1:5) was added prior to absorbency determination at 295nm on a nano drop spectrophotometer.

2.6.3 Preparation of double stranded probes for use in EMSA

The sequences of complimentary sense and anti-sense oligonucleotides used as labelled probes or unlabelled competitors in EMSA's are listed in Table 10. The dehydrated (lyophilised) sense/antisense oligonucleotides were re-suspended in TEN buffer (50mM Tris-HCL pH8.0, 150mM NaCl, 5mM EDTA) to a final concentration of 300ng/µl. For annealing, equal amounts of complimentary sense and anti-sense were incubated for 10 minutes at 95°C, to disrupt any secondary coiling of the oligonucleotides that may impede annealing, and then allowed to anneal by slow cooling at 15-25°C. Prior to the digoxigenin (DIG) end-labelling reaction the annealed oligonucleotides were further diluted with TEN buffer to a final concentration of 25ng/µl. Labelling reactions contained 25ng/µl annealed probe, 1X labelling buffer (roche), 5mM CoCl₂, 0.05mM DIG-ddUTP and 2.5 U terminal transferase (Roche). Reaction was mixed briefly by

pulse spin centrifugation, incubated at 37°C for 15 minutes and terminated on ice with the addition of 0.02M EDTA (pH 8).

The efficiency of the labelling reaction was determined by comparison of a spotted dilution series of labelling reaction with one of a labelled control-oligonucleotide (Roche) in a direct detection assay (described in section 2.6.4). The dilution series was 0ng/μl, 0.004ng/μl, 0.04ng/μl, 0.4ng/μl, and 4ng/μl. These dilutions were directly spotted onto Hybond+ nylon membrane and visualised by chemiluminescent detection and autoradiography (section 2.6.4).

2.6.4 EMSA

DIG-labelled probes (25ng/μl) were incubated with 5-10μg of total whole cell protein extract for 15 minutes at 15-25°C in a reaction containing 1X binding buffer (Roche), 1μg poly [d(I-C)], 0.1μg poly L-lysine and either double distilled H₂O or 200 X molar excess of the appropriate competitor probe. 1X loading dye (Roche) was added to the reactions, and protein-DNA complexes were resolved by electrophoresis on a 5% non-denaturing polyacrylamide gel (5% acrylamide, 0.06% bis-acrylamide, 40mM TrisHCL, 90mM boric acid, 2.5mM EDTA, 0.04% TEMED, 0.1% APS) in TBE buffer (40mM TrisHCL, 90mM boric acid, 2.5mM EDTA) at 150V for 1-2 hours at 4°C.

Following electrophoresis, migrated DNA-protein complexes were transferred, via contact blotting, to a positively charged nylon membrane (Hybond+, Amersham Biosciences). Thereafter the membrane was placed on Whatmann 3MM paper pre-soaked with 2X SCC buffer (Roche) and cross-linked at 120mJ for two minutes in a Stratalinker (Stratgene). Following cross-linking, DIG-labelled oligonucleotides were visualised by incubation with alkaline phosphatase-labelled F(ab)₂ anti-DIG Ab, followed by chemiluminescence reaction with 100μg/ml CSPD substrate (Roche) and auto-radiographic visualisation, following the manufacturers instructions for the direct detection assay (Roche).

Table 10 Sequences of double-stranded oligonucleotides used in EMSAs

Oligonucleotide Name	Sense/ Antisense	Sequence (5'-3')
MS1-TATA-Fw	Sense	CCTGCTTCTATTTAAATCCCAGG
MS1-TATA-Rv	Antisense	CCTGGGATTTAAATAGAAGCAGG
TATA-Consensus-Fw	Sense	GCAGAGCATATAAAATGAGGT
TATA-Consensus-Rv	Antisense	ACCTCATTTTATATGCTCTGC
MS1-GATA/-99-Fw	Sense	ATAGCACCACATCTTGGCAAGA
MS1-GATA/-99-Rv	Antisense	TCTTGCCAAGATGTGGTGCTAT
MS1-GATA/-78-Fw	Sense	ATGCCTTGCAATGTTCCACCGT
MS1-GATA/-78-Rv	Antisense	ACGGTGGAACTATGCAAGGCAT
MS1-GATA/-16600-Fw	Sense	ATCTCTTCTAGACAGGAAGCTGTTACTGTCTCT TCTAAAGTGTG
MS1-GATA/-16600-Rv	Antisense	CACACTTTAGAAGAGACAGTAAACAGCTTCT GTCTAGAAGAGAT
GATA4-Consensus-Fw	Sense	TAGCTTCCCATATCGTTTCTC
GATA4-Consensus-Rv	Antisense	GAGAAACGATATGGGAAGCTA
MS1-EBX-1556/-1550-Fw	Sense	TTCTCTGTCCACATGACTTATCCT
MS1-EBX-1556/-1550-Rv	Antisense	AGGATAAGTCATGTGGACAGAGAA
MS1-EBX-253/-247-Fw	Sense	ACCCTGGCACTTGGAGAAGAA
MS1-EBX-253/-247-Rv	Antisense	TTCTTCTCCAAGTGCCAGGGT
MS1-EBX-221/-215-Fw	Sense	AGCTAAGCACATGATTCAATC
MS1-EBX-221/-215-Rv	Antisense	GATTGAATCATGTGCTTAGCT
MyoD-Consensus-Fw	Sense	CCCTTGGAACATCTGTCGATGCTG
MyoD-Consensus-Rv	Antisense	CAGCATCGACAGATGTTCCAAGGG

2.7 *Chromatin immunoprecipitation (ChIP)*

2.7.1 *ChIP assay*

H9c2 myoblasts (4×10^6 per sample, 60mm dishes) were fixed in formaldehyde at a final concentration of 1% for 10-12 minutes at 15-25°C. Glycine (pH 2.5, 125mM final concentration) was then added to terminate the cross-linking. After 10 minutes, cells were washed with ice-cold PBS and scraped into 1 ml of PBS supplemented with protease and phosphatase inhibitors (0.2 mM leupeptin, 2μM myocrocystin, 0.3nM PMSF, 5 mM DTT and 10μM E64) and pelleted by centrifugation at 3,000 x g for 5 minutes (in the cold room, at 4°C). The supernatants were then removed and the pellets were lysed for 15 minutes in 500μl of buffer L1 (50mM Tris-HCL, pH 8.0, 10% glycerol, 0.1% NP40, 2mM EDTA). The resulting nuclei were pelleted by centrifugation at 3,000 x g for 5 minutes and re-suspended in 200μl L2 buffer (5mM EDTA, 1% SDS, pH 8.0, 50mM Tris-HCL). Samples were then sheared by sonication (4 x 30 seconds at amplitude 30). Lysates were centrifuged at 4,000 x g for 5 minutes (in the cold room, at 4°C) and the supernatants were transferred to a fresh pre-cooled Eppendorf tube for the immunoprecipitation step.

At this point, a sample (typically 35μl) was retained to determine inputs. The samples for immunoprecipitation were then diluted 1:10 in RIPA buffer (50mM Tris-HCL, pH 8.0, 0.5% NP40, 0.2 M NaCl, 0.5 mM EDTA). Extracts were precleared with 20μl of a 50% (v/v) suspension of protein G-agarose suspension in RIPA buffer for 15 minutes. Immunoprecipitations were executed for 16 hours by incubation with gentle mixing by rotation (in the cold room overnight at 4°C) with 4μg of antibody of interest [IgG (sigma), anti-acetylated H4 (Abcam) and anti-dimethyl H3K4 (Abcam)] or without antibody (negative control). Immune complexes (DNA-protein-antibody) were isolated with 100μl of a 50% (v/v) suspension of protein G-agarose (Sigma) suspension in RIPA buffer containing 1μg/ml sonicated salmon sperm DNA, by incubation with gentle mixing by rotation at 4°C (cold room) for 2 hours.

Protein G-Sepharose was recovered by centrifugation at 1,000 x g for 3 minutes and washed for 3 minutes with gentle mixing by rotation in 1 ml high salt buffer (20mM Tris, pH 8.0, 500mM NaCl, 2mM EDTA, 1% (v/v) NP40, 0.1% (w/v) SDS). Protein G-Sepharose was then recovered by centrifugation at 1,000 x g for 3 minutes and washed for 3 minutes with gentle mixing by rotation in 1 ml low salt buffer (1X Tris-HCL, EDTA, TE). Finally, the immune complexes were extracted in 250µl elution buffer (1X TE, 1% (w/v) SDS), heated for 10 minutes at 65°C and then vortex agitated and centrifuged for 1 minute at 200 x g. Supernatants were then transferred to fresh Eppendorfs. The input samples (made up to 500µl in TE) and the immunoprecipitated extracts (500µl) had their cross-links reversed by incubation at 65°C for 10 hours with 0.2M NaCl. Input and IP samples were then mixed vigorously with 500µl phenol:chloroform and left on ice for 5 minutes. A centrifugation step of 15,300 x g for 10 minutes (at 4°C) was then used to separate the aqueous and organic phases. The upper aqueous phase was transferred to an Eppendorf and incubated with 500µl isopropanol for 3 hours at -80°C. DNA was finally pelleted by centrifugation at 15,300 x g for 10 minutes, precipitated by the addition of 70% (v/v) ethanol and pelleted by centrifugation (15,300 xg, 10 minutes). The DNA pellets were then air dried for 10 minutes and re-suspended in 100µl of HPLC H₂O for the input and 50µl of HPLC H₂O for the other samples.

Typically 5µl of each input sample was separated on a 1% agarose gel (see section 2.2.3) to check DNA fragment sizes. For ChIP quantitative analysis, 2µl of sample was used for quantitative real time PCR. Products were quantified using SYBR green incorporation, with primers used from Table 10.

Table 11 Sequences of oligonucleotides used for ChIP PCR analysis

Oligonucleotide Name	Sense/ Antisense	Sequence (5'-3')
NPPA-PP-Fw	Sense	GGTTGGCTTCCTGGCTGACT
NPPA-PP-Rv	Antisense	CACCCCCACCCTAGATGTC
NPPA-RE1-Fw	Sense	TTTGGCTCCTTTCTGTCACC
NPPA-RE1-Rv	Antisense	CACACACACACACACG
MS1-PP-Fw	Sense	CGGAGCTCAGAACACCGTCAGAGCCATAGC
MS1-PP-Rv	Antisense	CCAAGCTTCAGGCTACCTGTTTCTTCTC
MS1-FLK-Fw	Sense	GAGGGGTAAATATAATCACGATACA
MS1-FLK-Rv	Antisense	AGGAAACTACAAGAAATTAATAATGC
MS1-UP3-Fw	Sense	CTCACCAAAAGAATAAAGAAGCTAGC
MS1-UP3-Rv	Antisense	TAGAAGAGACAGATAACAGCTTCCT
NCX1/-132-Fw	Sense	GTGTTGGATGAAGCGGAGAG
NCX1/-22-Rv	Antisense	AACATGGTTTGCATAGCTGCA
MS1/-127-Fw	Sense	CGGAGCTCAGAACACCGTCAGAGCCATAGC
MS1/+60-Rv	Antisense	CCAAGCTTCAGGCTACCTGTTTCTTCTC
MS1/-300-Fw	Sense	CGGAGCTCATGATTCAATCTAGTACTTTCC
MS1/-127-Rv	Antisense	CGGAGCTCAGAACACCGTCAGAGCCATAGC
MS1/-16702-Fw	Sense	CGCTCGAGGCCCCACGAGTTAAAGCTCA
MS1/-16432-Rv	Antisense	TGGGCTGTGTGGGTAGGAGA
MS1/-420-Fw	Sense	TTACAGAGGTTTAAGTGAGAGC
MS1/-227-Rv	Antisense	CGGAGCTCAGAACACCGTCAGAGCCATAGC
MS1/E1/-1526-Fw	Sense	CACATTTTATCTGGTCTAATACTG
MS1/E1/-1482-Rv	Antisense	ATTTTAAATAGAACTGAAAAGAGAAGTCA
MS1/E2/-281-Fw	Sense	TAAGGTCAAGGAAAACATAAAGCTA
MS1/E2/-190-Rv	Antisense	ACGGATATGTTCCCTCCTCTCTC

2.8 Comparative DNA sequence analysis

Comparative sequence analysis of human and rat was performed using web-based software available at the national centre for biotechnology information DCODE website (ECR browser, <http://www.dcode.org/>) and the Lawrence Berkley laboratory genome website (VISTA, <http://genome.lbl.gov/vista/index.shtml>). Orthologous sequences from *Homo sapiens*, *Macaca mulatta*, *Mus musculus*, *Rattus norvegicus*, *Canis familiaris*, *Monodelphis domestica*, *Gallus gallus*, *Xenopus laevis* and *Fugu rubripes* were obtained from the ENSEMBL (<http://www.ensembl.org>) genome data-base and aligned using CLUSTAL W available for the european bioinformatic institute website (<http://www.ebi.ac.uk>). For pattern matching transcription factor binding site analysis, MatInspector (available from Genomatix at <http://www.genomatix.de/products/MatInspector/>) was used, utilising its own position weight matrix library of transcription factor binding sites.

Chapter 3

Bioinformatic analysis of the *msl* promoter and associated regulatory domains

3.1 Introduction

The primary aim of the thesis presented here is to characterise the regulatory mechanisms governing context specific expression of myocyte stress 1 (*msl*). As the initial step of gene expression, transcription is the major point of regulatory control, and therefore understanding the mechanisms governing transcriptional control of *msl* will potentially give a major insight into its context specific regulation as a whole. The transcription of DNA is carried out by the RNA polymerases, with RNA polymerase II (pol II) transcribing the majority of nuclear genes (Levine & Tjian, 2003). The mechanism through which pol II drives the transcription of genes requires the presence of a basal promoter (Levine & Tjian, 2003). This region contains the core promoter, which comprises the TATA box (although a significant proportion of gene promoters are TATA-less), the transcription start site (TSS) and specific upstream promoter elements, which contribute to the context specific regulation of the gene.

Basal pol II transcription via core promoter elements in isolation is extremely inefficient. Significant levels of transcription of a pol II template requires the intervention of stimulatory transcription factors (TFs) that bind at specific DNA transcription factor binding sites (TFBS), which are distinct from the TATA box and initiator regions where the basal factors assemble. Many classes of transcriptional regulators, including DNA-binding transcription factors, coactivators, co-repressors and proteins that alter epigenetic modifications of DNA and nucleosomal histones, integrate and combine to influence the function of the minimal promoters thereby increasing pol II mediated transcription (Fickett & Hatzigeorgiou, 1997; Fischle et al, 2003; Levine & Tjian, 2003). These DNA regions encompassing the functional TFBS are collectively known as cis-regulatory modules (CRMs) and it is these regions that confer the unique

context specific expression patterns upon the individual gene (Levine & Tjian, 2003). The CRMs, which include enhancers, silencers and insulators, can be located in the 5', 3' and intronic regulatory regions scattered over distances of ~100kb in mammalian genomes (Adams et al, 2000; Aparicio et al, 2002; Holt et al, 2002; Lander et al, 2001) (Figure 3.1). This represents a massive genomic 'search space' which has rendered the classical experimental characterisation of CRMs almost impossible.

Ms1 with its unique developmental, cell specific and stress responsive transcriptional profile is likely to require the action of multiple CRMs. In order to achieve the regulatory characterisation outlined in Chapter 1 we will need to identify such CRMs and functionally correlate their activity with the endogenous expression profile. However the massive genomic 'search space' represents a major hurdle for the effective experimental characterisation and functional annotation of the CRMs and thus regulatory hardwiring governing *ms1* transcriptional regulation.

The key functional components of all CRMs are the TFBS themselves and therefore being able to predict potentially functional TFBS is an important first step in promoter analysis and hence regulatory characterisation. Over the past 20 years many computational approaches have been developed to identify putative functional TFBS, thereby allowing the prediction of functional CRMs and minimising the 'search space' for subsequent experimental characterisation (Ureta-Vidal et al, 2003; Wasserman & Sandelin, 2004). The myriad approaches to the computational prediction of functional TFBS are primarily based on two sequence identification models, pattern matching and pattern detection (Ureta-Vidal et al, 2003; Wasserman & Sandelin, 2004). Pattern matching utilises prior knowledge of all experimentally characterised DNA binding sites for any given transcription factor (TF). Originally the known binding sites for a given TF were represented as a consensus sequence, which is essentially a description of the most common nucleotides in each position of the TFBS. Although these consensus sequences were convenient they potentially excluded a significant subset of a binding site repertoire because of the omission of important functionally variable nucleotides within the TFBS (Roulet et al, 1998; Ureta-Vidal et al, 2003; Wasserman &

Sandelin, 2004). These variable nucleotides could be captured through the use of IUPAC strings (an alphabet of characters other than ATCG), which provides a framework in which one can convey alternative choices at each variable position (for example Y= C or T)(Appendix 1). However, although they convey more information than the classical consensus sequence, IUPAC strings do not provide quantitative information as to the relative frequencies of the alternative nucleotides they code for.

This information content can be more accurately modelled and ultimately quantified through the utilisation of position weight matrices (PWMs) (Roulet et al, 1998; Stormo et al, 1982) or position-specific scoring matrices (PSSMs). Both of these matrices are capable of incorporating pattern variability by recording the frequencies of nucleotides at each site or alternatively by assigning penalties to nucleotides that should not appear within a TFBS. These matrix-based pattern-matching techniques represent an improvement over consensus and IUPAC mapping in sensitivity (i.e they have a lower false-negative rate). However, even when using PWMs the false positive rate is extremely high, and subsequently specificity is typically measured in terms of rate of predictions: this rate varies for each TF binding model. When using the most standard model parameters TFBS have specificities of 1/500 to 1/5000bp. For example a published model for the binding of MyoD predicts one binding site in approximately every 500bp (Ureta-Vidal et al, 2003). If one then applies this model to the human genome, we would predict 10^6 MyoD TFBS, of which only 10^3 are likely to be functional (see above). This three orders of magnitude difference between true and false predictions is intolerable resulting in what Wasserman and colleagues choose to term the “Futility theorem”. This theory states that in essence all predicted TFBSs using pattern-matching models alone have no functional role (Wasserman & Sandelin, 2004).

The advent of complete genome sequences from multiple species over the last decade has led to the development of biologically motivated approaches that can overcome this 1000-fold excess in false predictions, with these new approaches encompassed within the field of comparative regulatory genomics. Comparative regulatory genomics describes the comparison of orthologous genomic sequences from multiple species in order to identify putative functional CRMs. This comparative process, also commonly

referred to as “phylogenetic footprinting”, is based on the premise that some function of regulatory elements will be conserved between species, and this functional conservation will be reflected in similar nucleotide sequences (Lenhard et al, 2003).

The computational process of phylogenetic footprinting consists of three principle components: defining suitable orthologous genomic sequences for comparison, aligning these orthologous sequences and finally visualising or identifying segments of significant conservation within these sequences. Defining suitable sequences for comparison is based on the assumption that the two sequences being aligned are derived from a common ancestor, and that the time since their divergence has enabled a significant accumulation of mutations within both genomes and that selection has occurred (Lenhard et al, 2003; Ureta-Vidal et al, 2003). It is crucial however that the genomes compared still maintain sufficient sequence similarity so that the homologous regions can be easily identified. Sequence comparisons between human and rodent, specifically mouse, which diverged ~90 million years ago (Mya), have been empirically shown to be suitable for identification of functional CRMs in many cases (Bagheri-Fam et al, 2001).

Once the appropriate orthologous sequences have been isolated, they must be aligned to identify segments of sequence similarity. There are two broadly used algorithms for such alignments: one strategy is represented by local alignments, which identify all of the possible high-quality alignments between the two sequences, regardless of order or orientation, alternatively global alignments generate a single alignment that is optimised across the entire length of both sequences (Ureta-Vidal et al, 2003; Wasserman & Sandelin, 2004). The choice to use global alignments introduces the assumption that important functional sequences have remained co-linear over evolution in the orthologous sequences analysed, which is clearly not always the case. It is difficult to determine which strategy is the most efficient and this is partly due to the lack of a bona fide ancestral sequence in which to use as the reference sequence. However one comparison of both strategies, which used test sequences derived from computationally simulated evolution from the same ‘ancestral sequence’, demonstrated that both

methods had similar accuracy with respect to sensitivity of conserved sequence identification (Ureta-Vidal et al, 2003; Wasserman & Sandelin, 2004).

Through combining PWM based pattern matching algorithms and phylogenetic footprinting it is possible to further enrich for functional CRMs thereby reducing the genomic ‘search space’ for experimental characterisation. rVISTA represents a recently developed tool that integrates clustering of predicted TFBS through pattern matching with the analysis of inter-species sequence conservation to maximise the identification of putative functional CRMs (Loots et al, 2002). It has been empirically demonstrated that in the analysis of an experimentally characterised mouse-human data-set, rVISTA was able to eliminate >95% of the approximate 58,000 predicted binding sites within the human sequence, with 88% of the experimentally verified binding sites enriched as positive regulatory candidates (Loots et al, 2002). Therefore, using the current array of computational tools and databases it is now possible to identify, with confidence, putative functional CRMs. This approach massively reduces the genomic ‘search space’ for experimental analysis by generating a focused set of experimental targets. In addition, through integrating context specific binding site information derived from the literature, it is also possible to gain a putative insight into the functional role of identified CRMs. This kind of integrated analysis allows one, based on primary nucleotide sequence alone, to predict the context in which putative regulatory CRMs are functional (caveat being that the data is inferable from the existing literature). Regulation is intrinsic to function, thus such an analysis could (when stringent enough) allow functional annotation of unknown genes.

In light of the tools discussed here, a thorough unbiased computational regulatory analysis of the *msl* genomic interval was carried out. The premise for this analysis was for the identification of putative functional CRM targets for focused experimental characterisation in subsequent analysis described in this thesis. We have also interrogated the current available literature for context specific regulatory information that we will integrate with our computational analysis. This integration would potentially allow the regulatory context of any putative CRMs identified to be predicted based purely on genomic sequence. This has the potential to give us a further insight

into the biological function of *msl*, further expanding our current understanding of its biological roles.

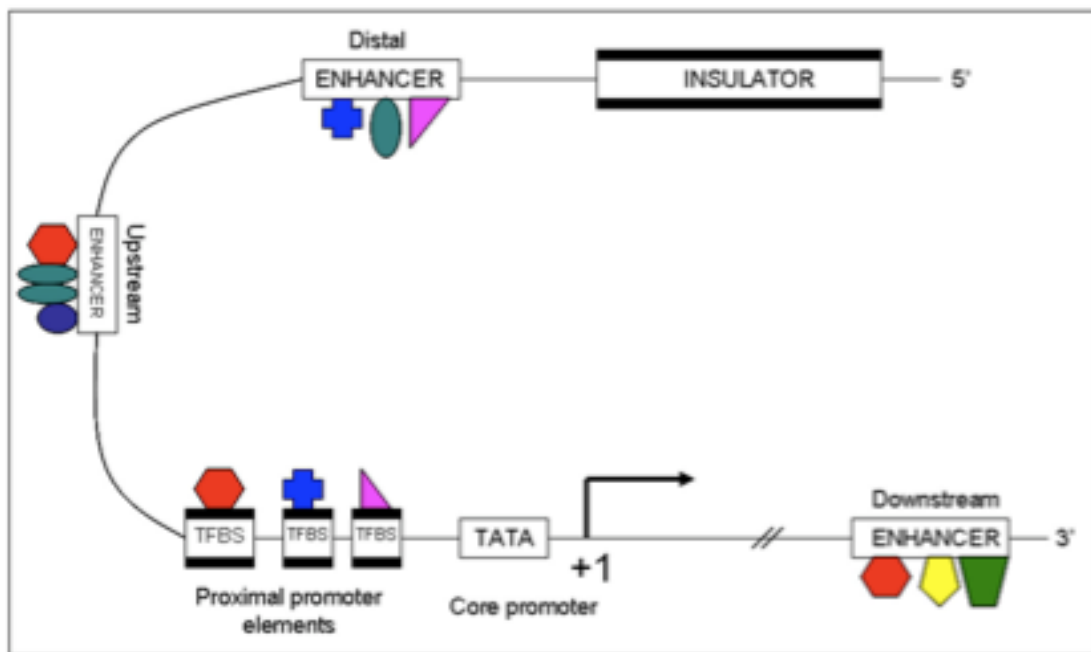


Figure 3.1. Components of transcriptional regulation. Transcription factors (represented by coloured shapes) bind to specific DNA sites (transcription factor binding sites, TFBS) that are proximal (proximal promoter elements) or distal (upstream and downstream enhancer and repressor modules) to the transcription start site. Co-ordinated and context specific binding at these cis regulatory modules (CRMs) mediates context specific transcriptional output.

3.2 Results

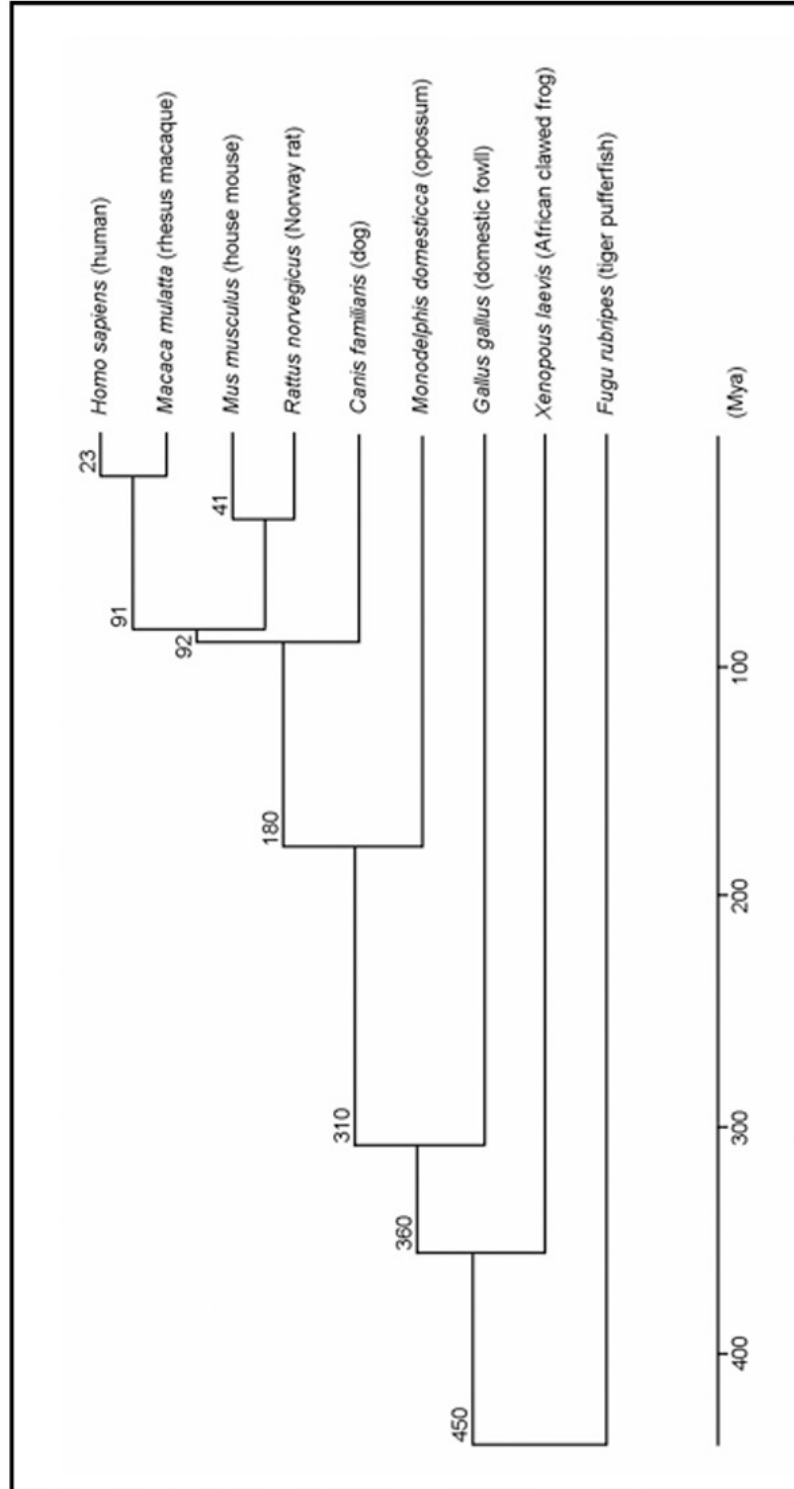
3.2.1 Comparative sequence analysis of the rat *msl* genomic interval.

As a first step in identifying the putative location of functional CRMs acting on the *msl* TSS (as defined within both the ENSEMBL and GenBank databases) we analysed a 60kbp genomic interval on rat chromosome 7 (7q31) encompassing the whole of the *msl* gene and associated 5' and 3' flanking sequence. This 60kb interval also includes the 3' region of the *Oxr1* genomic sequence.

We aligned and compared this 60kbp rat interval with the corresponding orthologous genomic *loci* from a number of other species in multiple mammalian and non-mammalian lineages. The choice of species, and therefore the orthologous *loci* utilised here, is of critical importance for the identification of functional CRMs. Comparing closely related species, for example mouse and rat (Figure 3.2.A, species diverged 41 mya), will highlight genomic sequences where divergence is most readily tolerated, by highlighting differences rather than similarities between the two species. Conversely when one compares distantly related species, for example rat and chicken (Figure 3.2.A, species diverged 310 mya), sequences under positive selection (genomic sequences constrained during evolutionary selection) are more easily identified (Ureta-Vidal et al, 2003).

In order to gain a comprehensive insight into the conservation status of the *msl* 60kbp genomic locus we compared the rat sequence with the orthologous *loci* from multiple species spanning a large evolutionary history (phylogenetic relationships representing approximately 450 million years of evolutionary time), including both mammalian and non-mammalian, lineages (Fig.3.2.A). The appropriate orthologous sequences were aligned using the MULAN alignment engine and visualisation tool available via the ECR browser (as described in the methods). MULAN utilises a widely employed technique (also used by the VISTA and Pipmaker engines) to graphically display sequence conservation profiles in reference to the base DNA sequence (our base sequence being the *Rattus norvegicus*, 7q31 60kbp *msl* genomic interval) that is linear

A



B

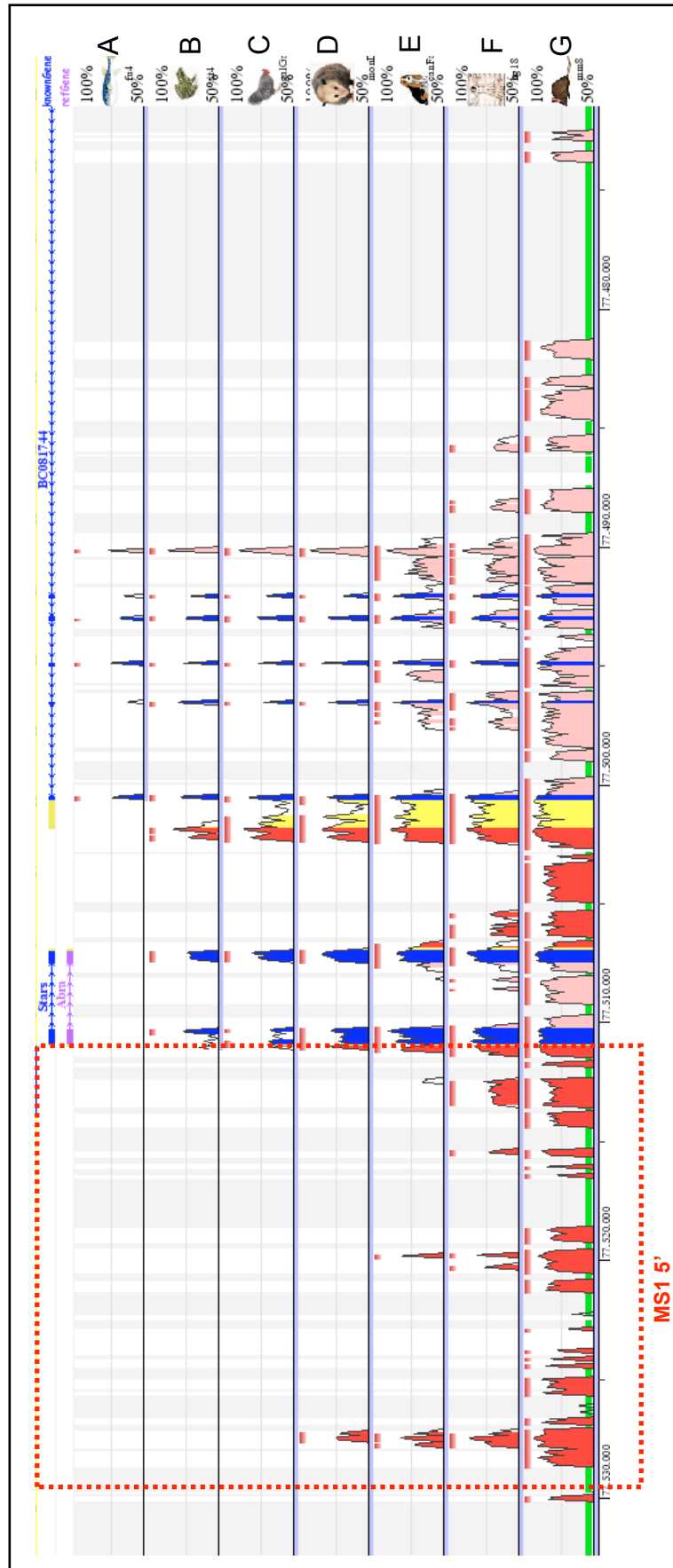


Figure 3.2 Conservation status of the *Rattus Norvegicus* MS1 (STARS) genomic locus (7q31). (A) Evolutionary relationships between the species compared are simplified and plotted in a phylogenetic tree with evolutionary distances annotated. (B) MULAN conservation analysis of the rat large MS1 locus as compared with mouse, human, dog, opossum, chicken, frog and zebrafish species. The vertical axis for each panel represents the percent sequence identity. The rat DNA sequence lies along the horizontal axis. Coloured peaks depict regions of conservation with blue representing exon's, yellow indicating UTRs, pink being non coding intronic sequence and red indicating upstream and downstream non coding sequence.

along the horizontal axis (Figure.3.2.B) while the vertical axis displays percent identity with respect to the chosen secondary sequence (Figure.3.2.A, multiple-species). Regions of significant non-coding conservation are graphically displayed as red peaks with exonic and untranslated coding sequences depicted by blue and yellow peaks respectively. Local aligners (like MULAN) enrich for all high-quality alignments between the orthologous loci, regardless of orientation or order and are therefore suited for identification of CNS between distantly related species.

In the MULAN analysis we utilised a CNS evolutionary threshold for the enrichment of sequences conserved to a level of 70% over a 100bp sliding window between orthologous intervals. Using this threshold our analysis demonstrates that within our 60kbp interval the primary area of non-coding conservation (red peaks) lies within the 20kbp 5' flanking region upstream of the *msl* TSS.

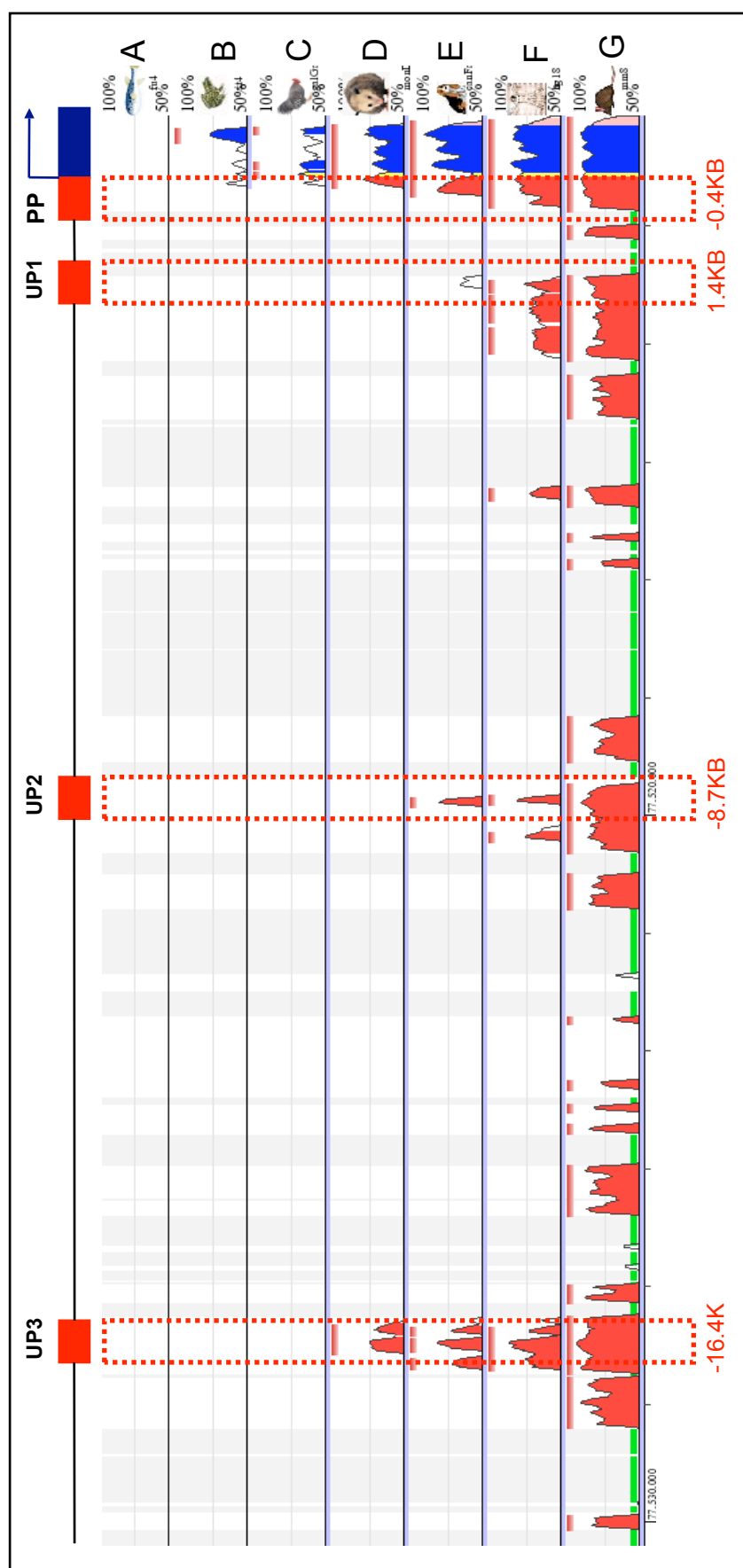
3.2.2 Identification of putative cis regulatory modules (PCRM) within the 5' conserved flanking sequence.

The preliminary comparative analysis of the large *msl* genomic interval has demonstrated that the 5' flanking sequence directly upstream of the rat *msl* TSS is highly enriched with CNS which we designate as putative cis regulatory modules (PCRM). In order to further interrogate this region we executed a focused multi-species MULAN conservation analysis utilising previously described evolutionary thresholds. This focused analysis identifies four PCRM that are conserved in multiple vertebrate lineages, specifically rat, mouse, human, macaque and dog. These four PCRM, designated PP, UP1, UP2 and UP3, can be identified in the MULAN analysis when utilising both the rat (Figure.3.3.A) and human (Figure.3.3.B) genomic intervals as the base sequence for comparison. Based on the phylogenetic relationships between these species we can assume that these four PCRM were present in the last common ancestor over 100 million years ago (mya) and have been subsequently maintained under positive evolutionary selection.

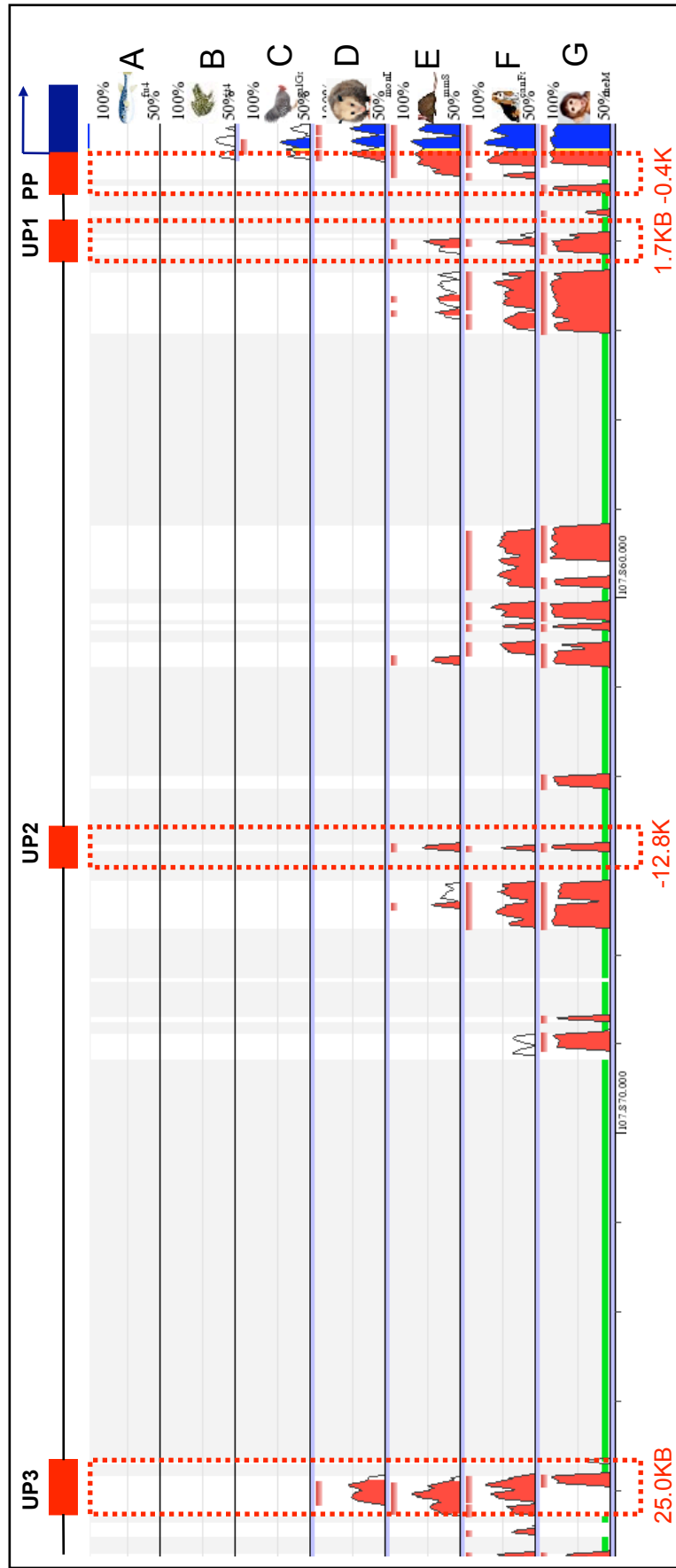
The level of constraint varies among the four PCRM identified suggesting these differences may relate to the functional importance of each PCRM, with those highly conserved likely to be of most functional importance. Thus, it was not too surprising that the PP domain, which encompass the proximal 400bp directly upstream of the *msl* TSS, is highly conserved within multiple mammalian and non-mammalian species including the frog and chicken (Figure.3.3.A and .B). The UP1 and UP2 appear to be specific to the human, rodent and canine lineages with no orthologous sequences identified in the opossum, which shared a last common ancestor with these lineages 180 mya. These domains can therefore be classified as species specific PCRM. This finding may have relevance to the functional roles of these respective domains. The only PCRM conserved within the rodent and opossum lineages is the UP3 domain, which has therefore been maintained in the 180 million years since these species diverged.

The non-coding conservation status of the PCRM (PP,UP1,UP2 and UP3) across multiple distant species implies evolutionary importance and likely regulatory function (CRM). Bona fide CRMs, including enhancer and repressor elements, can by definition act in a distance and orientation independent manner (Levine & Tjian, 2003). We therefore identified rat PCRM using CLUSTALW alignment and spatially mapped the orthologous loci in the rat, mouse and human genomes (Figure 3.3.C). Not surprisingly, considering that rodents and human diverged only 91 mya, all PCRM share a common orientation across the 5' flanking interval although the spatial distribution varies. This is exemplified by the UP2 and UP3 domains which are significantly more distal to the TSS in the human locus compared to respective *loci* in the rodent lineages. This suggests that the human locus has accumulated sequence within the 5' flanking sequence and the inter-CRM regions. None of this accumulated sequence has affected the sequence within the UP2 and UP3 domains or their relative orientation and co-linearity with respect to the TSS. It is also noteworthy that the spatial distribution of UP1 with respect to the PP has been constrained in the human and rodent lineages implicating the distance between these domains as an important parameter for regulatory function.

A



B



C

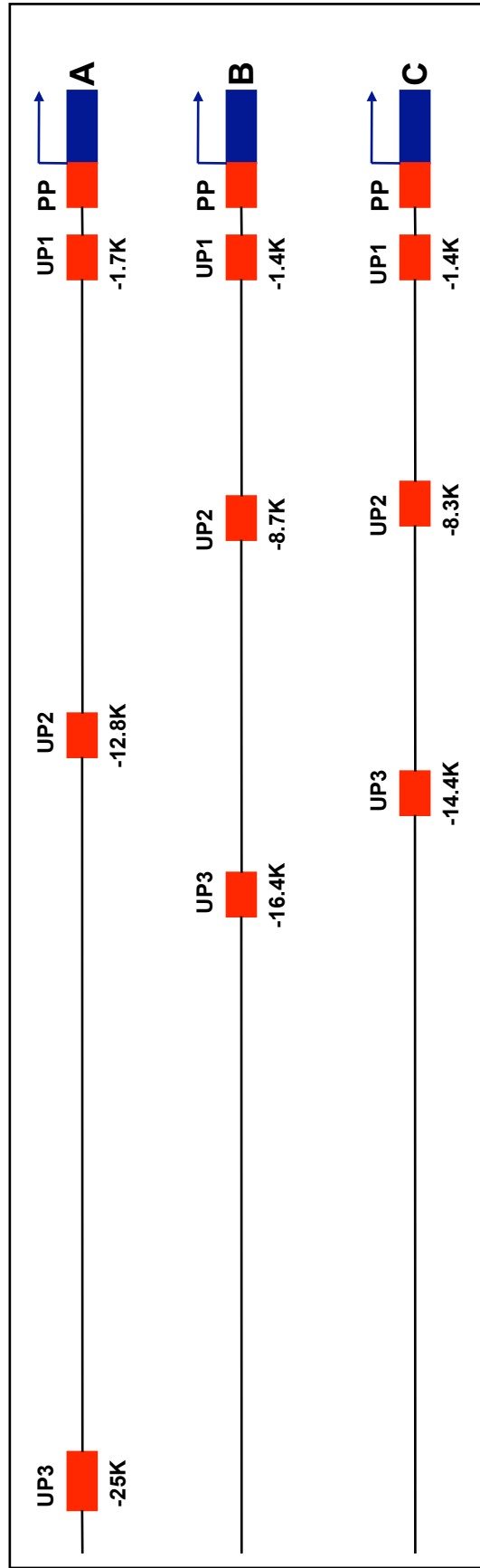


Figure 3.3. Conservation status of the *Rattus norvegicus ms1* 5' flanking conserved interval of the rat (A) and human (B) as compared with rhesus macaque, mouse, dog, opossum, chicken, frog and zebrafish species. Putative proximal and distal cis regulatory modules (PCRM) are boxed in dashed red line and their relative location with respect to the transcription start site is annotated below the box. The identified PCRM are spatially mapped in the human (A), rat (B) and mouse (C) endogenous loci.

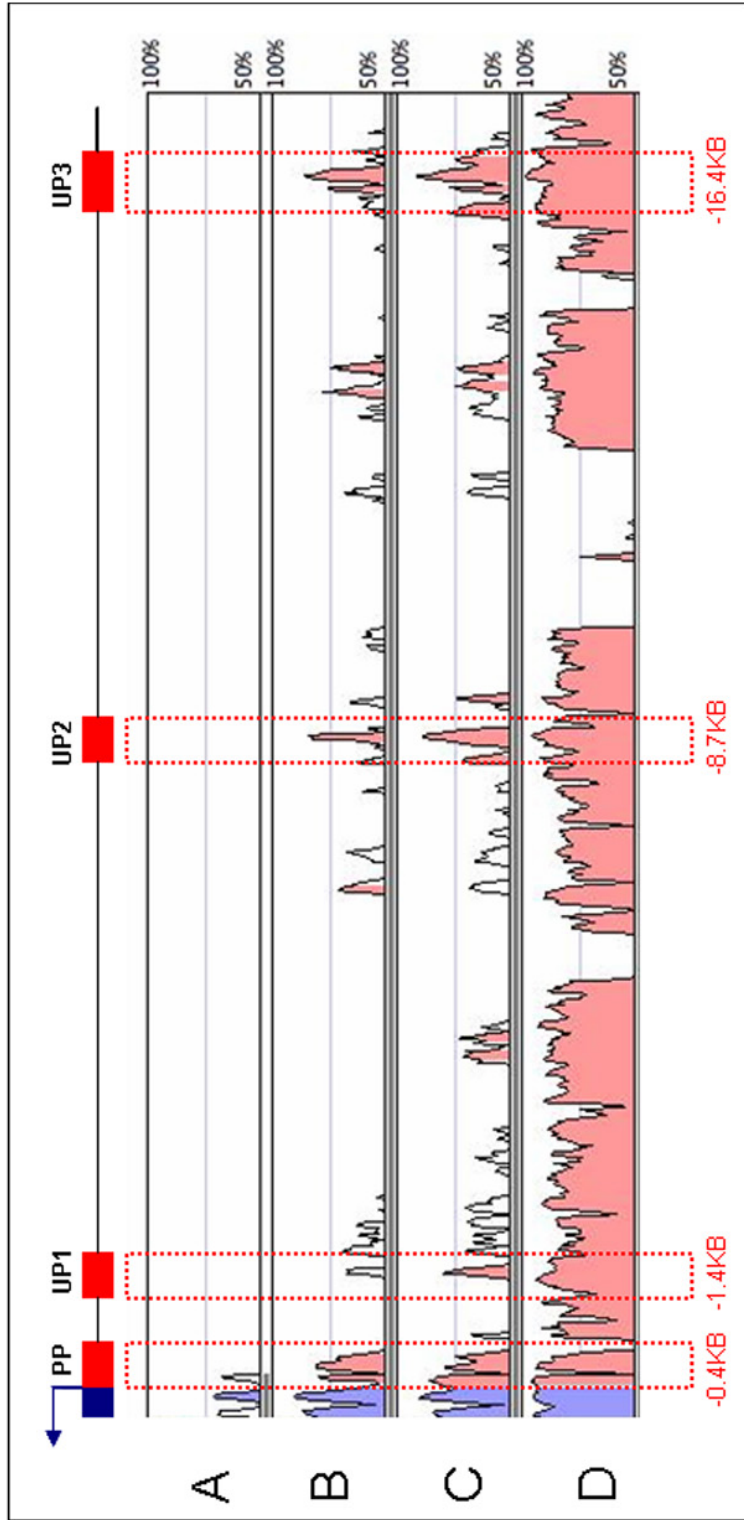


Figure 3.4 VISTA browser plot of the *rattus norvegicus* MS1 5' flanking conserved interval. Visualisation plots show conserved sequences between human(D), dog (C), opossum(B) and chicken (A) plotted against the rat base sequence using the MLAGAN algorithm. The conservation level is plotted on the vertical axis in the co-ordinates of the rat axis (horizontal axis). Conserved regions about the set conservation level of 70%/100bp are highlighted under the curve, with blue indicating a conserved exon and pink a conserved non-coding sequence. Putative proximal and distal *cis* regulatory modules (PCRM) are boxed in dashed red line and there relative location with respect to the transcription start site is annotated below the box.

Global alignment engines identifies the optimum conservation across a whole interval therefore orientation and geographical organisation of CNS/CRMS needs to be orthologous. Thus based on the present spatial mapping (Figure 3.3.C), the ms1 5' flanking interval is intrinsically suited for a global alignment based comparative analysis. The VISTA global alignment engine (<http://genome.lbl.gov/vista/>) was used to perform a multi-species analysis of the 5' flanking interval previously analysed using MULAN. The rat genomic interval was compared to the orthologous loci in the mouse, human, dog and chicken genomes (Figure 3.4). In agreement with the MULAN derived domains, four PCRMS were present in the mouse, human and dog lineages, in the same relative positions to the TSS. The PP domain is also conserved within the chicken lineage.

3.2.3 Identification of conserved TFBS within annotated PCRMs utilising rVISTA

CRMs typically function through the specific recruitment of context specific transcription factors, thus the TFBS themselves are the functional components of CRMs. In order to probe the functionality of the four PCRMs identified here, we executed a rVISTA based analysis to identify conserved TFBS within each PCRM. rVISTA 2.0 operates a pair-wise alignment strategy to identify TFBS (rVISTA utilises TRANSFAC database of TFBS) that are shared by the two input species and which are also located in an area of high local sequence conservation (80% sequence identity over a 20bp sliding window). Conserved and aligned binding sites are graphically represented as vertical lines located at the appropriate genomic position with respect to the base genome on the horizontal axis (Figure 3.5.A, -B, -C, -D, -E, -F, -G).

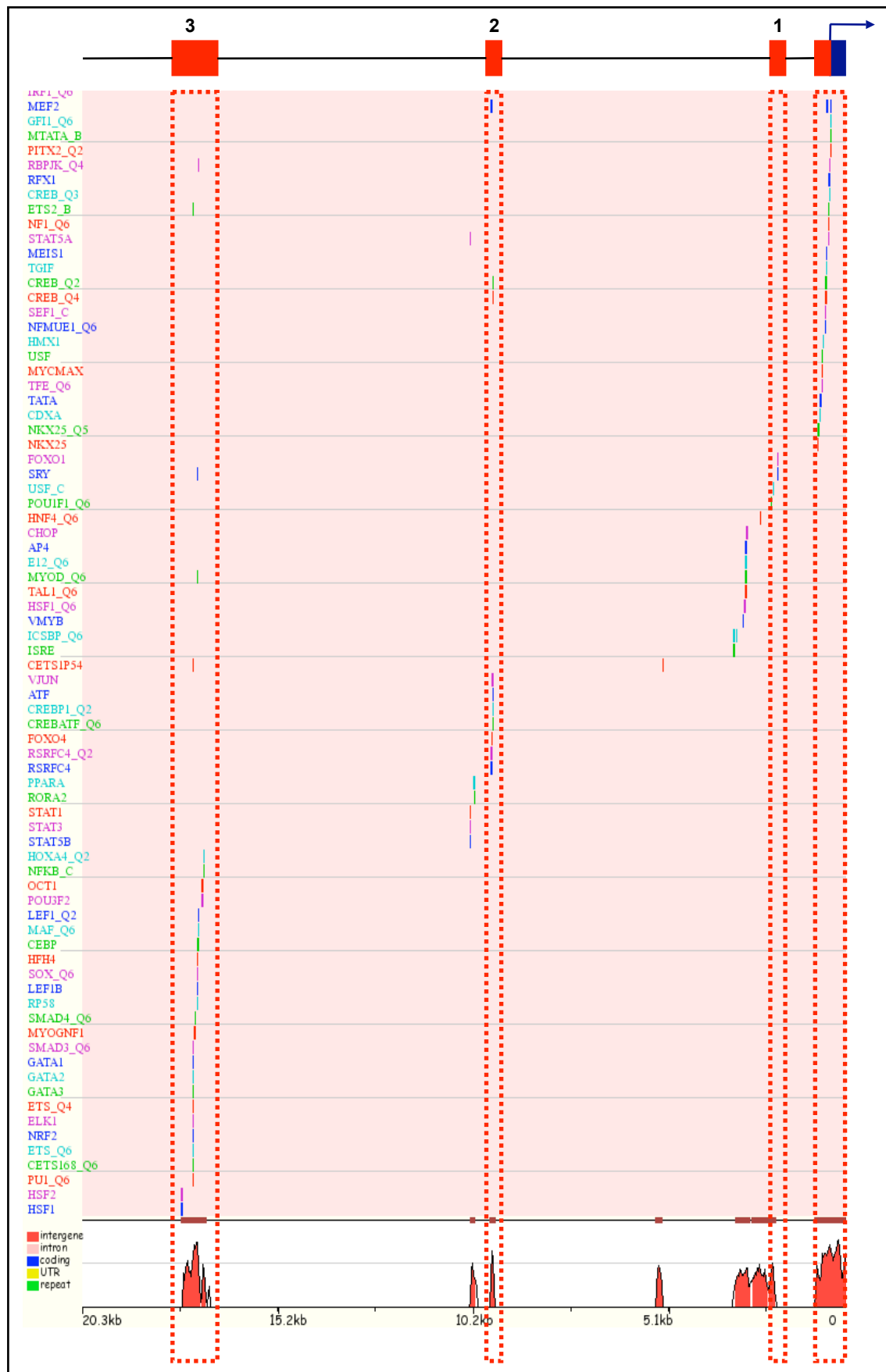
Using the ECR browser, pair-wise rVISTA 2.0 analysis of the rat 20kbp 5' flanking interval with human (Figure 3.5.A), dog (Figure 3.5.B), opossum (Figure 3.5.C), chicken (Figure 3.5.D) and frog (Figure 3.5.E) sequences were performed. The interval was submitted to TRANSFAC using default parameters (matrix similarity values of 0.75 and core similarity values of 0.7)(Fu & Weng, 2005; Knuppel et al, 1994). Figure 3.5 shows the rVISTA outputs in rat and human base sequence pair-wise alignments. The regions of evolutionary conservation are depicted with the PCRMs represented by red

peaks on the horizontal axis. The previously annotated PCRM are highlighted with a dotted red box. Predicted TFBS that are evolutionary conserved are depicted on the vertical co-ordinate by coloured lines which map to their endogenous location (either in rat Figure 3.5.A-E, or human Figure 3.5.F-G) on the horizontal axis.

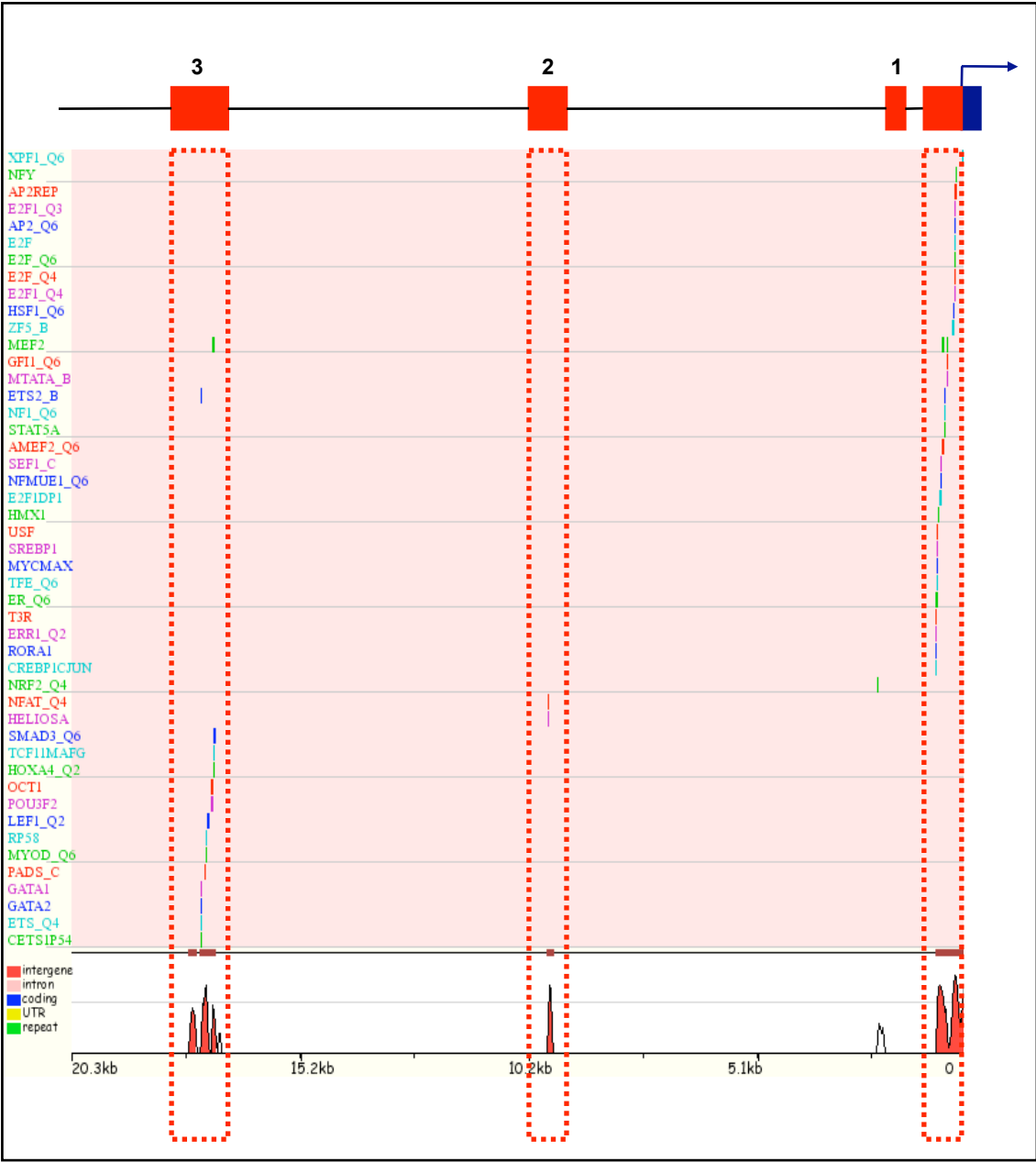
The graphical output identifies a plethora of conserved TFBS enriched within each PCRM. We speculated that common TFBS enriched within specific PCRM in two or more cross-species alignments are likely to represent functionally important TFBS. Binding sites for important muscle specific transcription factors including MEF2, SRF, NFAT and GATA4 were found to meet this criteria in that they were enriched in multiple comparisons in the various PCRM (summarised in Table 1). Multiple E-Box elements were also enriched in the PP and UP1 domains. The TATA box is a critically important motif required for transcriptional initiation (Roeder, 2005). Our output identifies a conserved TATA motif located proximally to the TSS in the PP domain. This TATA box, in conjunction with an adjacent Mef2 motif, represents the most conserved binding site identified being conserved in all pair-wise alignments, including rat/chicken (Figure 3.5.D) and rat/frog (Figure 3.5.E) alignments.

In support of the notion that bona fide TFBS will be enriched in multiple species pair-wise alignments, human/opossum (Figure 3.5.F) and human/chicken (Figure 3.5.G) 30kbp 5' flanking interval were also analysed. Common TFBS were again enriched in both the UP3 (Figure 3.5.F) and PP (Figure 3.5.F-G) domains, including the TATA box and Mef2 motifs within the PP and GATA motifs within the UP3 domain. rVISTA derived findings are summarised in Table 1 specifying the TFBS that are enriched within multiple species alignments. These motifs represent the most evolutionary conserved TFBS within each PCRM and represent prime candidates for experimental analysis.

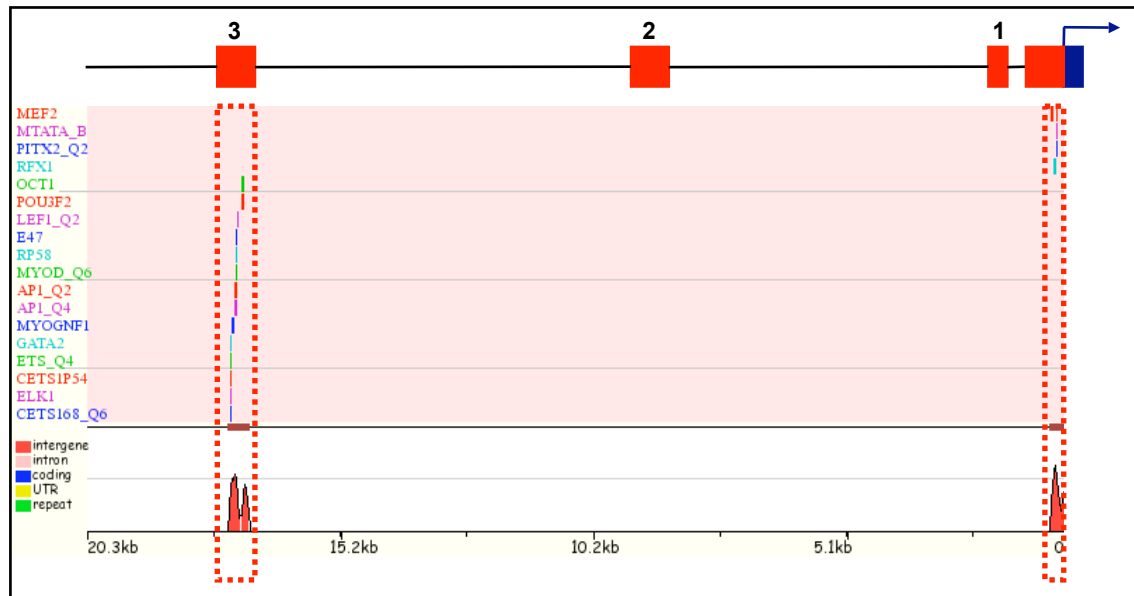
A



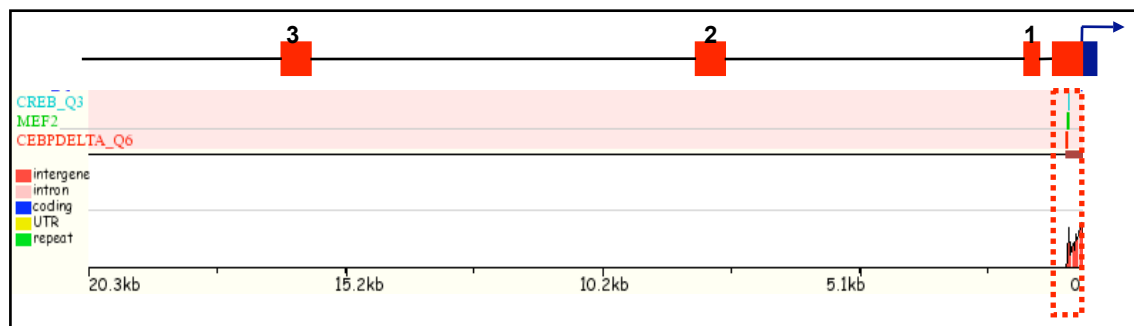
B



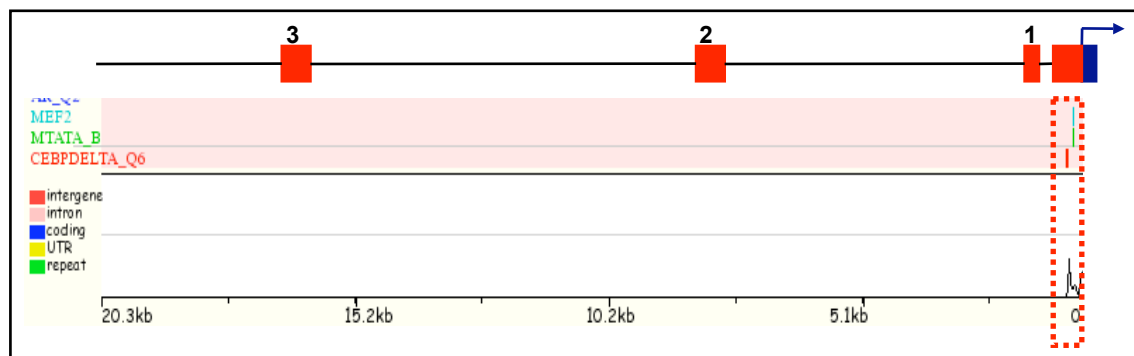
C



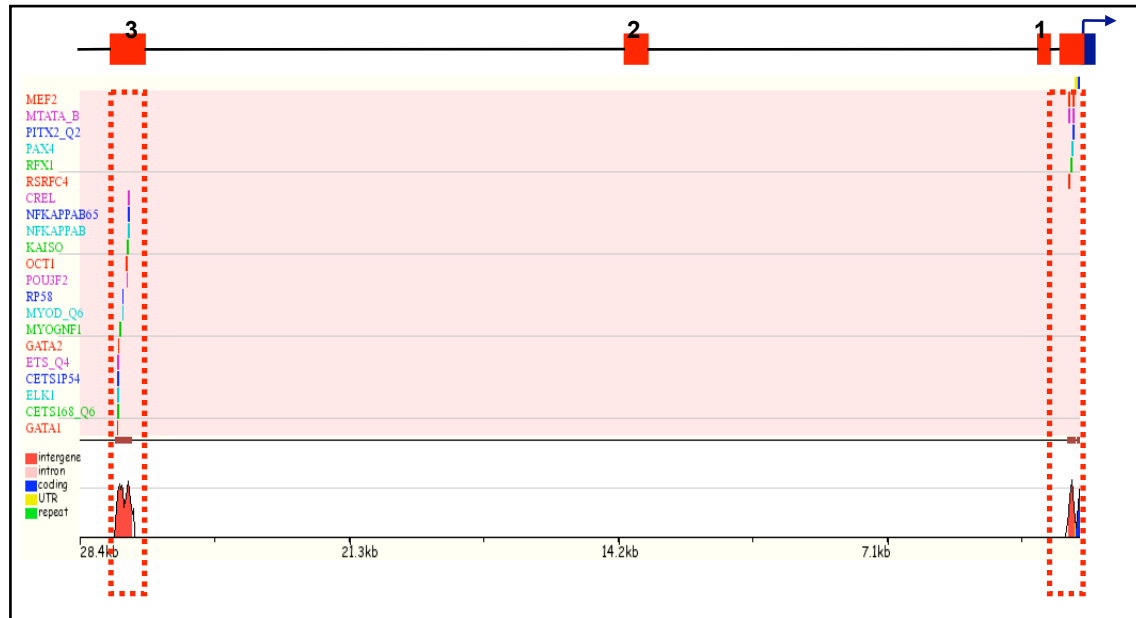
D



E



F



G

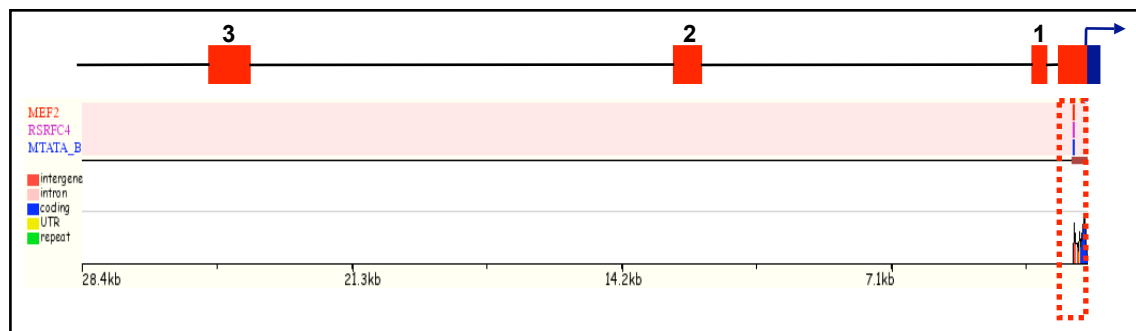


Figure 3.5. The rat (A-E) and human (F-G) ms1 5' conserved flanking intervals were submitted to the rVISTA 2.0 sequence analysis software and searched for conserved clusters of TFBS between specific species. The rat interval was compared with human (A), dog (B), opossum (C), chicken (D) and frog (E). The human interval was compared to opossum (F) and chicken (G). A vertical coloured line which corresponds to its specific TFBS indicates the position of the conserved TFBS relative to the base sequence (rat or human).

Table.1 Summary of rVISTA derived TFBS.

Proximal Promoter	UP1	UP2	UP3
TATA	E-Box	MEF2	E-Box
MEF2		NFAT	SBE
E-Box			GATA
Nkx2.5			
SRE			

3.2.4 MatInspector based interrogation of the isolated rat PCRM

The rVISTA analysis presented above enriches for TFBS using the pattern matching based Match scoring scheme (Loots et al, 2002) for matrix matching utilising PWM derived from the publicly available TRANSFAC 6.0 database, which currently contains 336 matrices (Fu & Weng, 2005; Knuppel et al, 1994). This represents only half of the matrices available in the commercial TRANSFAC (release 8.4) and MatInspector libraries. It is therefore plausible that bona fide TFBS within our PCRM may have been neglected in the rVISTA based pattern matching analysis, simply as a consequence of the limited information content of the TRANSFAC 6.0 library. We have therefore utilised an alternative pattern matching based TFBS search engine, MatInspector (Cartharius et al, 2005), to analyse the isolated PCRM domains. The current library utilised by MatInspector contains a total of 634 matrices, greatly expanding the search scope compared to the rVISTA based TRANSFAC 6.0 library (rVISTA reference). At the time of this analysis (January 2005) the MatInspector matrix library represented the most comprehensive library available in terms of specificity and sensitivity for bona fide TFBS.

The rat PCRM domains, isolated from the Ensembl genome database, were delivered in FASTA format into the MatInspector search engine, using default parameters for core and matrix similarity thresholds. The generated output file consists of a linear schematic for visualisation and a linked table which contains detailed sequence information and matrix core similarity values (Figure 3.6 and Appendix 2). On the linear schematic, each TFBS is represented by a coloured semi-circle (Figure 3.6) with the linked colour being represented in the associated table (Appendix 2).

The default analysis utilising all of the 634 matrices currently available generated 95, 26, 50 and 31 hits in the PP, UP1, UP2 and UP3 domains respectively (Figure 3.6). The number of hits is proportional to the size of the domain, which is to be expected. The linked table with colour coded TFBS information including location, matrix and core similarity values is available (Appendix 2). Although this analysis by itself does not generate a focused set of functional TFBS, it does (based on currently available matrices) provide a detailed overview of all putative binding sites based on PWM within the PCRM domains. This therefore provides us with a more complete set of binding sites with which to manually investigate for potentially functional binding sites.

3.2.5 Manual filtering and integration of MatInspector and rVISTA predicted TFBS

The futility theorem states that the majority of TFBS predicted using pattern matching algorithms will represent false positive predictions and therefore we suspect that the majority of our MatInspector derived hits in the PP, UP1, UP2 and UP3 domains will be biologically inactive (Wasserman & Sandelin, 2004). However, based on the evolutionary constrained nature of these PCRM we can be confident that a small yet significant proportion of matches will represent bona fide functional TFBS. In order to enrich for potentially functional TFBS and filter out false positive matches we formulated and executed a manual filtering process on all of the MatInspector derived hits.

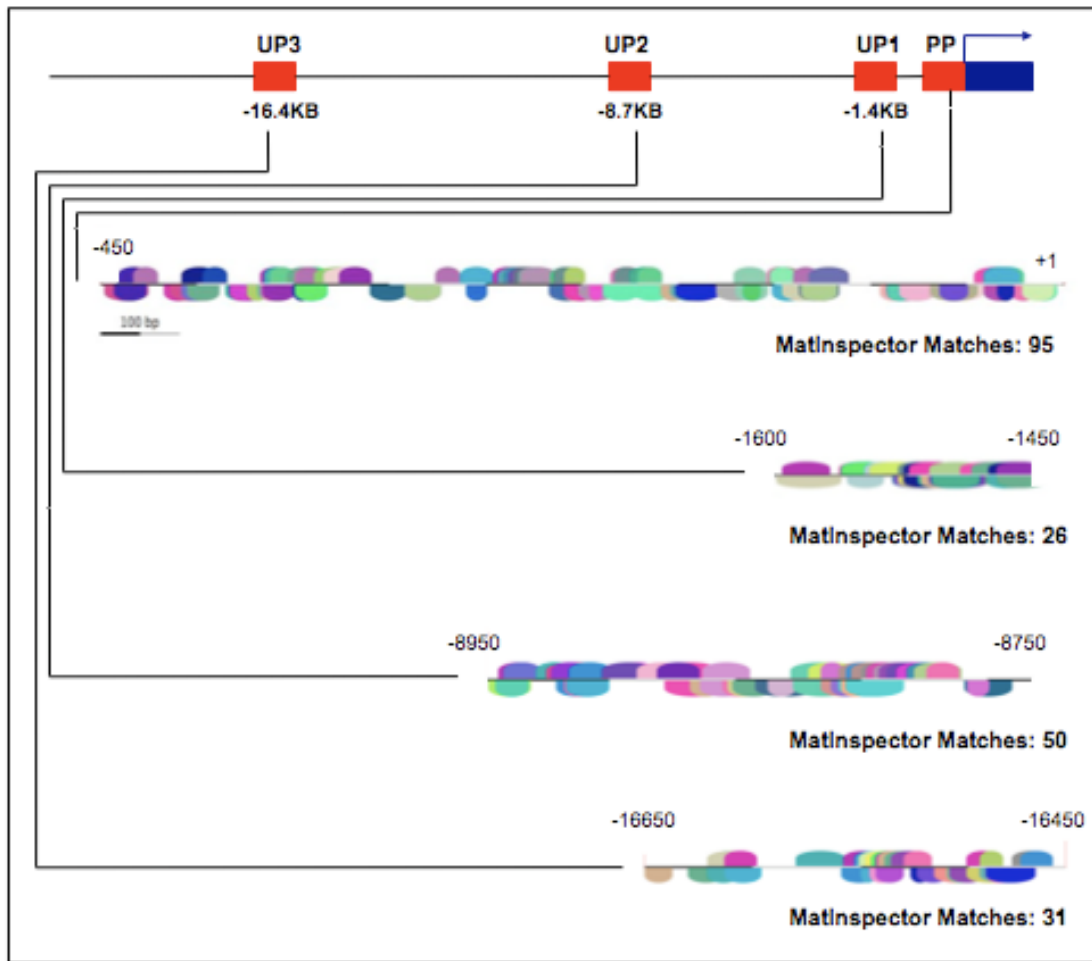


Figure 3.6. The rat PP, UP1, UP2 and UP3 PCRM sequences were obtained from the ENSEMBL genome database and submitted to MatInspector TFBS identification algorithm. Different colours denote different transcription factor families/binding sites (See Appendix 2 for detailed annotation).

The primary enrichment filter/parameter involves the exclusion of all MatInspector hits with a matrix similarity score lower than 0.8. The matrix similarity score gives a quantitative measure of how similar the test sequence is to the optimum PWM for that motif in the MatInspector library. Thus, the closer to “1”, the more conserved the test sequence is. This score would reach “1” only if the test sequence corresponds to the most conserved nucleotide at each position in the matrix. Any test sequences with scores >0.8 (85% of total hits) were considered to be biologically more significant and are then manually interrogated for phylogenetic conservation. This was executed through visual analysis of CLUSTALW rodent/human sequence alignments of the respective intervals (Figure 3.7). This filtering parameter allows us to use evolutionary constraint as an enrichment tool for functional TFBS with bona-fide motifs inherently more conserved. Single nucleotide conservation mismatches were tolerated at this step if they did not reduce the overall matrix similarity score to <0.8. This filtering step further reduced the number of MatInspector derived hits by 30%. Thus, first and secondary filtering processes excluded 45% of the initial MatInspector derived hits. However there still remained over 100 MatInspector hits within our PCRM with the majority of these likely representing false positive predictions.

The final filtering parameter was based on the integration and utilisation of biologically relevant data regarding the context specific expression of *msl*. *Msl* transcription is under tight regulatory control with robust striated muscle specificity, predominantly in cardiac and skeletal muscle. It is therefore a sound assumption that context specific regulatory factors of *msl* are likely to include factors implicitly involved in striated muscle specific gene regulation. In order to construct a library of such factors we performed a meta-analysis of all the published literature to extract information on TFBS and cognate factors with characterised roles in striated muscle gene expression. The following terms, ‘cardiac transcription factor’ and ‘skeletal transcription factor’ were inputted into the Medline search engine with these ‘search strings’ generating 5507 and 5655 hits respectively. We then manually sorted the papers and extracted motif information on any TFBS and cognate factor with a role in striated muscle determination, differentiation, growth, proliferation, stress signalling and post-natal homeostatic function. The results of this analysis were compiled into a detailed library

(Tables 2 and 3) which annotates information on the TFBS, cognate factors able to binding the site, muscle type with characterised function and references to the publication in which data was extracted.

The remaining MatInspector hits (approximately 100 hits) were examined for motifs corresponding to those present in our striated muscle library. After this final filtration step our original 200 MatInspector hits was reduced to a total of 28 hits (86% of original hits excluded), which we consider to be strong candidates as context specific regulatory TFBS. The first observation from our three-tiered manual enrichment process was that all of the rVISTA derived sites were also enriched in our 28 hits. This was of significance because rVISTA is a well characterised automated engine that is capable of significantly enriching for bona fide TFBS. Thus our manual analysis appears to have the same level of stringency and specificity as rVISTA which leads us to suggest that any additional motifs identified through our analysis may have functionality. Our analysis identified a total of 10 new motifs located within all four PCRM. These motifs included GATA, Mef2, Pbx and KLF motifs. We have schematically mapped all binding sites that were enriched through our manual and rVISTA based enrichment processes in a CLUSTALW alignment based schematic (Figure 3.7). TFBS are boxed in coloured dotted lines with each colour linked to a specific TFBS.

Figure 3.7 A, D, E and F shows rat/mouse/human CLUSTALW alignments of the PP, UP1, UP2 and UP3 domains respectively with all the enriched motifs annotated. In order to emphasise the ultra-conserved nature of the PP domain and specifically GATA, Mef2 and TATA motifs within it, we have also executed rat/chicken and rat/frog alignment with these ultra-conserved motifs annotated Figure 3.7 B and C). These results have been schematised in a linear fashion allowing for easy visualisation (Figure 3.8). The motifs, potential cognate factors and genomic location with respect to the rat *msl* TSS have also been summarised in Table 4.

In summary we have manually interrogated the PCRM and systematically filtered the MatInspector derived hits using a biologically driven approach to enrich for striated muscle specific transcription factors. This data set was then further integrated with our

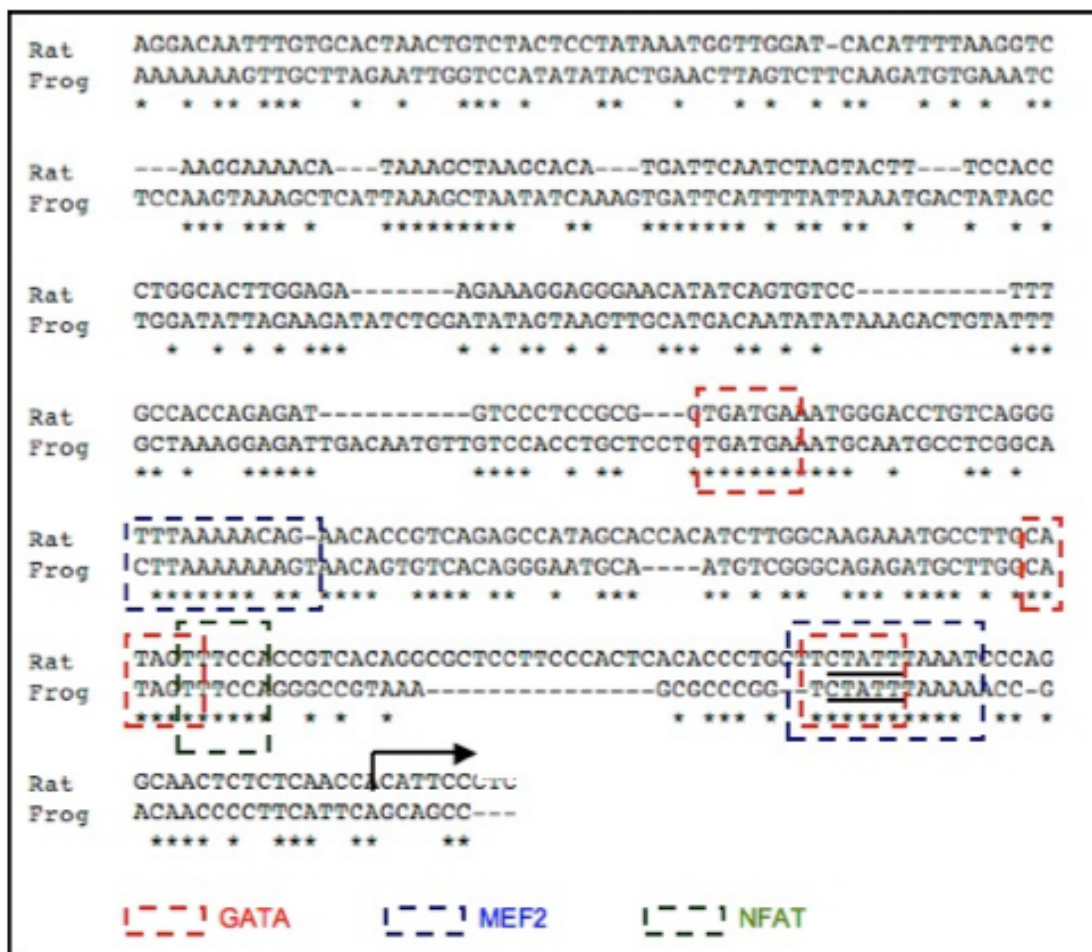
rVISTA derived predictions to generate a final predicted library of PCRM_s and encompassed TFBS (Figure 3.8).

[illegible]

B



C



D

Rat	CAGTGGCAAAGGAAAAAGAAAGGTCTATTTGTTCTTAAGGAATTATTCAATGCTTAGTC
Mouse	--GTGGCAAAGAAAAAC-AAAGGTCTATTTGTCTTCAAGACATCATTCAATGCCCAGTC
Human	CAGACCTGTTGAATGAAAAAAATTCCGTCTTACTTTAAGAAATTATTCAATTGCTGAGCC
	* * * * *
Rat	CTGCACGTTTAAATCTGGTATAATACTGAACAGGTGCTGTT-TCTCTGTCCACATGAC
Mouse	CTGCACATTTTATCTGGTCTAATACTGAATAGATGCTGTT-TCTCTGTC-ACATGAC
Human	CTGGACATTTTATCTGGTTAATACTGCATAGATGTTTCTCTTCTGTCCACATGAC
	*** * * * * *
Rat	TTATCCTTTTCAGTTCTTTAAAAATATTATTTACTTGATG
Mouse	TTCTCTTTTCAGTTCTTTAAAAATACTTATTTACTTGATG
Human	TTGTCTTTCCAGTCCTTTTAAAAAAATATTATATTTTCAGTC
	* * * * *
	!-! E-BOX !-! GATA !-! MEF2

E

Rat	AAAAGAATAATTTTATTGGAA-AACTTGAGATGGGAACAAGGAGT---AAACTAAGGAGT
Mouse	AAAACAATAATTTTATTGGAA-AAC--GATATGGGAACAGGGGGT---AAATTAAAGAGT
Human	AAACCAGAATTTTAAATGGAGTGAGTTGAAGGGTGAGCAGGAGGTGATAAATTAAGGAGT
	*** * * * * *
Rat	GCAGAATGCCTTTCTAATACACTTGGCTACTAAGTAAGGAAAATTTTACAGGATCAAAT
Mouse	GCAGAATGTCTTTCTAATATACTTGGCTACTAAGTAAGGAAAATTTTACAGGATCAAAT
Human	GTAGACTACTTTTGCAGACTCTTGGCTATTAAGCAAGGAAAATTTTACAGAATCAAAT
	* * * * *
Rat	TGAAAGTCATGGCTAAGAACTCTATATTTAGATCTTCGTAAGAAACTTGTTTAGAAAGT
Mouse	TGAAAGTCATGGCTAAGAACT--TATATTTAGATCTTCGTAAGAAACTTGTTTAGAAAGT
Human	AGAAAGTCATGCTGAGAACTCTATATTTAGATCCCTGTAAGAAGCTTGTTTAGGGAGT
	***** * * * * *
Rat	GGCGTCACAAGCTTGTGAAGTCTTGTCCCTTGAGCGTGGGGGAGGAAAGCAAGAGAAAGG
Mouse	GGTGTCAACAAGCCTGAGGAGTCTGTCCCTTGAGCGTGGGGGAGGGAAGCAAGGAGAAAGG
Human	GATGTCAACAAGCTTGTAAAGTCTGCCCCCTCAAAGTGGGGGTGAGGAGCAAGAGGAAGG
	* * * * *
Rat	ACAAAATTT-TCTTAGGACCCAGACTTCAGCAGTTTTCATTCTGGAGTCA
Mouse	ATGGAATTT-TCTTAGGGCACAGACTTCAGCAGTTTTCATTCTGGAGTCA
Human	AAGGAATTTGTGTTAGAGCACAGTTGTCAACAGCCCTGACCCTGGAGCCA
	* * * * *
	!-! MEF2 !-! NFAT !-! KLF

F

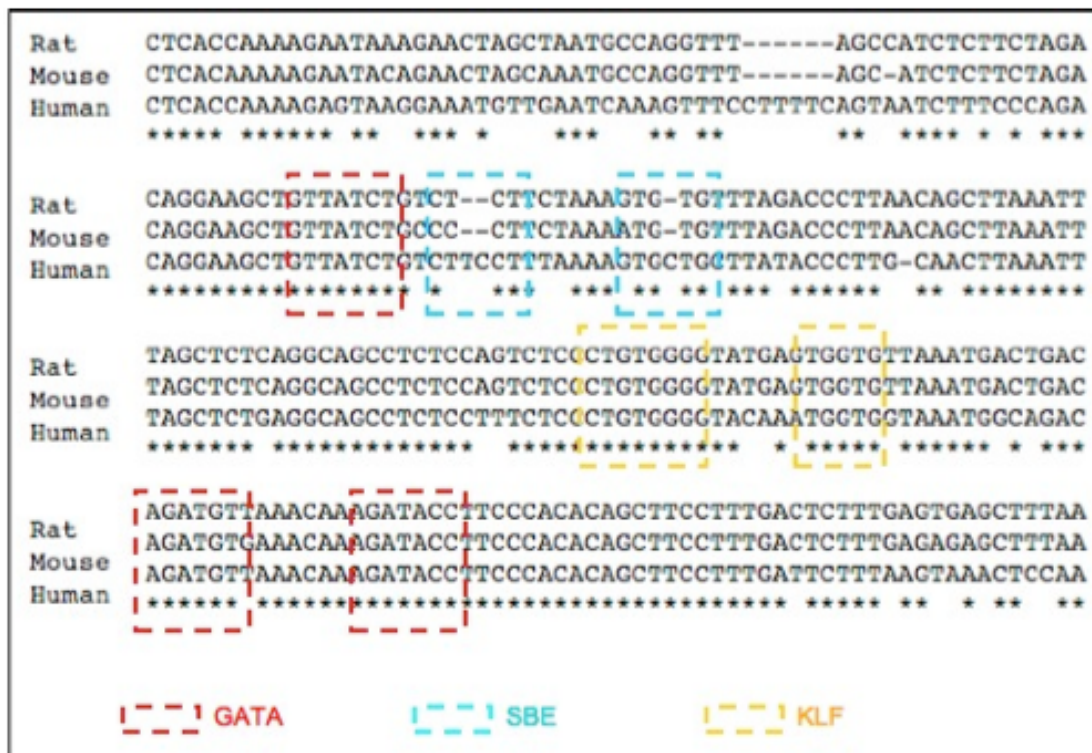


Figure 3.7. The rat, mouse, human, chicken and frog PP, UP1, UP2 and UP3 PCR sequences were obtained from the ENSEMBL genome database and aligned using the CLUSTAL W alignment tool. rVISTA and MatInspector filtered TFBS were boxed in different colour dashed lines corresponding to appropriate TFBS. Transcription start site is depicted by an arrow. (A), (B) and (C) represent the PP domain. (D) UP1, (E) UP2 and (F) UP3 domain.

Table 2 Literature survey of transcription factors and cognate TFBS implicated in cardiac specific gene expression.

Transcription Factor	Consensus TFBS (5'-3')	Muscle Type	References
GATA Family GATA4/5/6	(AT)GATA(AG)	Cardiac	(Molkentin, 2000; Peterkin et al., 2005; Peterkin et al., 2007)
SRF	CC(A/T) ₆ GG	Cardiac	(Niu et al., 2007)
MEF2 Family MEF2 A/B/C/D	CTA(A/T) ₄ TA	Cardiac	(Black and Olson, 1998; Karamboulas et al., 2006)
Nkx2.5	CAAGTG	Cardiac	(Akazawa and Komuro, 2005; Harvey, 1996)
T-Box Family Tbx1/2/3/5/18/20	GTGNNA (IUPAC)	Cardiac	(Hoogaars et al., 2007; Plageman and Yutzey, 2005)
KLF Family KLF5/13/15	CACCC	Cardiac	(Haldar et al., 2007; Lavalley et al., 2006; Nagai et al., 2005; Nemer and Horb, 2007)
Hand Family Hand1/2	CANNTG (IUPAC)	Cardiac	(McFadden et al., 2005; Srivastava, 1999; Srivastava et al., 1995)
Hey Family Hey1/2	CANNTG (IUPAC)	Cardiac	(Fischer et al., 2005; Rutenberg et al., 2006)
JMJ	SW ₄₋₆ STAWT (IUPAC)	Cardiac	(Kim et al., 2003)
Mesp Family Mesp1/2	CANNTG (IUPAC)	Cardiac	(Satou et al., 2004)
Msx Family Msx1/2	CTAATTG	Cardiac	(Chen et al., 2007)
COUP-TFII	CAGGTCACAGGTCA GA	Cardiac	(Lo and Frasch, 2001; Tian, 2003)
IrxFamily IrxF1/3/4/5	AAAACACGTGTAA	Cardiac	(Christoffels et al., 2000)
Pitx2	TAATCT	Cardiac	(Logan et al., 1998; Tessari et al., 2008)
Sox Family Sox 4/5/7/8/9/10/18	WWCAAWG (IUPAC)	Cardiac	(Lincoln et al., 2006; Zhang et al., 2005)
TEF-1	CATTCC	Cardiac	(Chen et al., 1994)
RXR	RGKTZA (IUPAC)	Cardiac	

Transcription Factor	Consensus TFBS (5'-3')	Muscle Type	References
NFAT	GGAAA	Cardiac	(Lambrechts and Carmeliet, 2004; Molkenin, 2004; Schulz and Yutzey, 2004) (Shi et al., 2000)
SMADs	CAGA	Cardiac	
NF-KB	GGGRNNYYY (IUPAC)	Cardiac	
NRSF/REST	NTYAGMRCCNNRGM SAG (IUPAC)	Cardiac	(Hall et al., 2006; Jones et al., 2005) (Kuwahara et al., 2003)
Hex	ATTAA	Cardiac	(Foley and Mercola, 2005)
Foxc1/2	GTAAATAAA	Cardiac	(Seo and Kume, 2006)
FOXO Family FOXO1/2/3	GTAAACA	Cardiac	(Evans-Anderson et al., 2008)
Islet 1	CTAATG	Cardiac	(Laugwitz et al., 2008; Lin et al., 2006)
YY1	CATNT (IUPAC)	Cardiac	
PPAR	AGGTCANAGGTCA (IUPAC)	Cardiac	(Sucharov et al., 2003) (Smeets et al., 2007; Teunissen et al., 2007)
STAT Family STAT1/3	TTCYNRGAA (IUPAC)	Cardiac	(Ananthakrishnan et al., 2005; Fischer and Hilfiker-Kleiner, 2007)

Table 2. Literature survey of transcription factors and cognate TFBS implicated in skeletal specific gene expression.

Transcription Factor	Consensus TFBS (5'-3')	Muscle Type	References
MyoD	CANNTG (IUPAC)	Skeletal	(Berkes and Tapscott, 2005; Tapscott and Weintraub, 1991; Weintraub et al., 1991)
MRF4	CANNTG (IUPAC)	Skeletal	
Myf5	CANNTG (IUPAC)	Skeletal	(Hinterberger et al., 1991; Rhodes and Konieczny, 1989)
Myogenin	CANNTG (IUPAC)	Skeletal	(Ott et al., 1991)
MEF2 Family MEF2 A/B/C/D	CTA(A/T) ₄ TA	Skeletal	(Wright et al., 1989)
SRF	CC(A/T) ₆ GG	Skeletal	{Naya, 1999}(Black and Olson, 1998)
Pax3	GTCACACA	Skeletal	(Cen et al., 2004; Charvet et al., 2006)
Id Family Id1/3	CANNTG (IUPAC)	Skeletal	{Tajbakhsh, 1997 #80}
Pbx	TGATTGAT	Skeletal	(Kemp et al., 1995)
NF-KB	GGGRNNYYY (IUPAC)	Skeletal	(Berkes et al., 2004)
NFAT	GGAAA	Skeletal	(Kramer and Goodyear, 2007)
FOXO	GTAAACA	Skeletal	(Schulz and Yutzey, 2004)
SMADs	CAGA	Skeletal	(Furuyama et al., 2003)
PPAR	AGGTCANAGGTCA (IUPAC)	Skeletal	(Kollias and McDermott, 2007)
T-Box Family Tbx1/15	GTGNNA (IUPAC)	Skeletal	(Gilde and Van Bilsen, 2003; Grimaldi, 2005)
Barx2	TAATGRTT (IUPAC)	Skeletal	(Dastjerdi et al., 2007; Singh et al., 2005)
TEF-1	CATTCC	Skeletal	
RXR	RGKTZA (IUPAC)	Skeletal	(Meech et al., 2003)

Transcription Factor	Consensus TFBS (5'-3')	Muscle Type	References
Pitx3	TAATCT	Skeletal	(L'Honore et al., 2007)
SOX Family Sox6/8/9/15	WWCAAWG (IUPAC)	Skeletal	(Hagiwara et al., 2007; Lee et al., 2004; Nie, 2006; Schmidt et al., 2003)

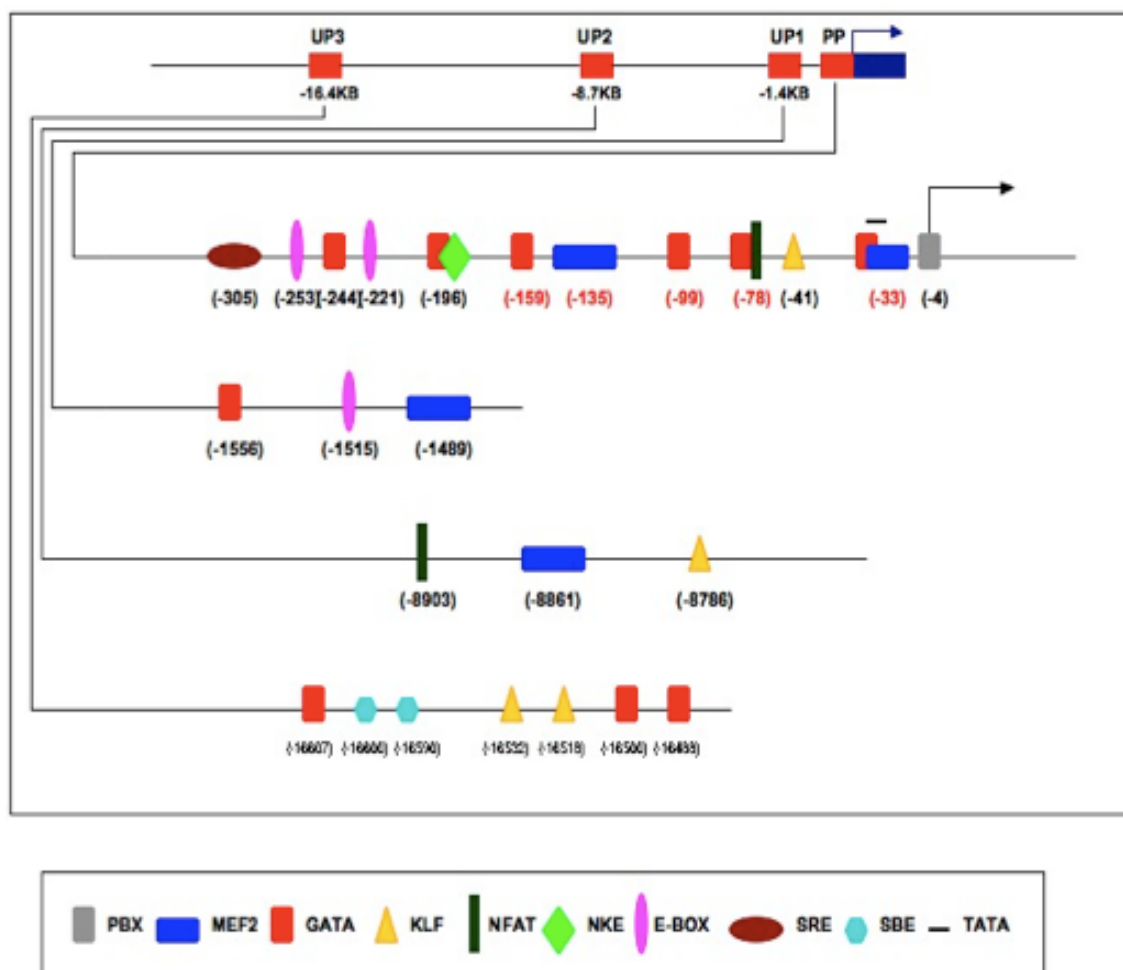


Figure 3.7. Linear schematic of our identified PCR fragment with filtered TFBS annotated using coloured shapes. The relative location of each TFBS is annotated below.

Table 4. Summary of filtered TFBS and their respective location within identified PCRM relative to the transcription start site.

Motif	Binding Factor(s)	Loci (bp)	CRM
PBE	Pbx	-4/+3	PP
MEF2	MEF2 A/B/C/D	-33/-22	PP
		-135/-124	PP
		-1489/-1477	UP1
		-8861/-8847	UP2
TATA	TBP	-30/-25	PP
GATA	GATA 4/5/6	-32/-27	PP
		-78/-71	PP
		-99/-94	PP
		-159/-153	PP
		-196/-190	PP
		-244/-238	PP
		-1556/-1550	UP1
		-16488/-16482	UP3
		-16500/-16494	UP3
		-16607/-16601	UP3
KLF	KLF5/13/15	-41/-36	PP
		-8786/-8778	UP2
		-16518/-16513	UP3
		-16532/-16524	UP3
NFAT	NFAT c3/c4	-72/-68	PP
		-8903/-8896	UP2
NKE	Nkx2.5	-195/-188	PP
E-Box	Hand 1/2, Hey 1/2, MyoD, MRF4, Myf5, Myogenin, (E-Proteins, CLK, BMAL)	-221/-215	PP
		-253/-247	PP
		-1556/-1550	UP1
SRE	SRF	-305/-288	PP
SBE	SMADs	-16590/16583	UP3
		-16800/-16595	UP3

3.3 Discussion

Understanding the context specific regulatory mechanisms governing *msl* transcription is of central importance to fully understand the functional role of *msl* in striated muscle biology. Obtaining such a mechanistic understanding would also give an intriguing insight into the general composition and function of the regulatory circuits that are involved in these associated processes, specifically striated muscle specific gene expression during differentiation, maturation and post-natal adaptation to stress. To ultimately obtain such an understanding one needs to identify and characterise the full regulatory hardwiring of the *msl* transcriptional unit, which comprises *cis* regulatory modules (CRMs) and encompassed TFBS. Historically, experimental identification and annotation of these CRMs would be almost impossible due to the limitless location of such regulatory modules, the so called genomic ‘search space’ problem (Ureta-Vidal et al, 2003; Wasserman & Sandelin, 2004). However, with the advent of whole-genome sequencing projects coupled to the development of more refined TFBS identification algorithms (and increase in TFBS matrix library content) computational annotation methods have been developed which would allow for biologically driven *in silico* prediction of CRMs ((Ureta-Vidal et al, 2003; Wasserman & Sandelin, 2004). Utilising comparative genomics, which incorporates the filtering power of evolutionary selection, it is now possible to identify, with confidence, potential CRMs therefore diminishing the genomic ‘search space’ for experimental analysis which will vastly improve and facilitate the experimental annotation of bona-fide CRMs (Bagheri-Fam et al, 2001; Loots et al, 2002).

In this chapter, the currently available set of computational tools were utilised to execute an unbiased interrogation of the *Rattus norvegicus msl* genomic interval (7q31) with the premise of identifying putative *cis* regulatory modules (PCRMs) for focused experimental characterisation in subsequent analysis. As the first step in our analysis, the MULAN browser was used to compare a large *msl* encompassing genomic interval from the rat with an array of orthologous loci derived from multiple phyla. The basis for this large interval analysis was to gain a ‘geographical’ insight into the constrained nature of the locus and identify regions of high conservation. We identified one major

region under high evolutionary constraint, a 20kbp interval located in the 5' flanking sequence directly upstream of the *msl* TSS.

The majority of functionally annotated CRM's are typically located within the 5' flanking regions directly upstream of the TSS (Fickett & Hatzigeorgiou, 1997; Levine & Tjian, 2003). We therefore suspected that our 20kbp 5' flanking interval represented a prime locus for functional CRMs acting on the *msl* transcriptional unit. Using both global and local alignment strategies, four PCRM were identified designated PP, UP1, UP2 and UP3. Within this interval the PP domain represents the most highly conserved domain, maintained throughout evolution over the last 350 million years, to the point at which rat and frog last shared a common ancestor. Due to this and its location directly upstream of and encompassing the TSS we speculate this PP domain represents the basal promoter. Numerous basal promoters share common evolutionary constraint and genomic location (Xu et al, 2006). In contrast to the PP domain, the UP1, UP2 and UP3 domains demonstrate clear species specific enrichment within the lineages analysed. The UP1 and UP2 domains are common only to the rodent, human and canine lineages which shared a last common ancestor approximately 92 mya. The UP3 domain can be found within the opossum lineage which diverged from rodent 180 mya suggesting this domain contributes to a functionally more ubiquitous aspect of *msl* genetic regulation. Other studies have demonstrated that highly conserved CRMs (like UP3) tend to be involved in developmental aspects of gene regulation (Wray, 2003; Wray et al, 2003). Many developmental morphological processes are common during the embryonic development of multiple diverse species and this explains to some extent the highly conserved nature and enrichment of specific developmental CRMs within multiple lineages (Wray, 2003; Wray et al, 2003).

Why are the UP1 and UP2 domains absent within the other lineages examined? This may reflect the stringency of the parameters used in the comparative analysis and therefore it is possible that these domains were simply not conserved to a high enough level for them to be enriched. However, when utilising the global alignment strategy, regions of low non- threshold conservation would be visually represented and we see no low-conservation peaks corresponding to the UP1 and UP2 domains. We therefore

suggest these domains may have arose during evolution after the rat/human/canine common ancestor diverged from opossum approximately 180 mya. Evolutionary adaptation at the gene regulatory level is mediated by co-utilisation of established CRM by other genes (Wray, 2003; Wray et al, 2003). This occurs through genomic rearrangements and retro-transposition events which result in CRM dispersion within the genome to alternative new loci and potentially within close proximity of other genes (Wray, 2003; Wray et al, 2003). If such CRM transposition events are evolutionary favourable, they will be positively selected and conversely negatively selected if detrimental. It is also possible, although somewhat less likely, that the UP1 and UP2 domains were present in a last common ancestor before the opossum diverged but simply not maintained within the opossum lineage branch while being conserved within the rodent branch.

Although the UP1 and UP2 domains are not conserved within the opossum and earlier lineages, it is still plausible that the regulatory contexts they mediate are conserved. Many examples exist where different species have utilised the same set of TFBS for the same regulatory output with such motifs residing within non-orthologous sequences. This is particularly common when multiple common TFBS are utilised, for example E-Box motifs (Ureta-Vidal et al, 2003; Wasserman & Sandelin, 2004). Evolution can thus utilise non-orthologous sequences to mediate common transcriptional output of orthologous genes within multiple species.

On the basis of PCRM mapped within our constrained 5' flanking interval and the emerging paradigm for modularity in gene regulation (Firulli & Olson, 1997), the following is proposed. The PP domain represents the proximal basal promoter. Typically proximal promoters alone are not able to mediate all aspects of a genes context specific regulation, for example transcriptional changes during development (Lien et al, 2002) or in response to a stress stimulus (Xu et al, 2006). They are however required for all regulatory contexts because they contain the core promoter and the site of pol II recruitment and therefore transcriptional initiation. The proximal promoter therefore serves to integrate context specific signals transduced by distal CRMs. The ANP promoter is a classical example of how the proximal promoter interacts with distal CRM

to mediate context specific regulatory transcription (Temsah & Nemer, 2005). It is likely the UP2 and UP3 domains serve such a function and mediate context specific transcriptional outputs. It was of interest that the UP1 domain, when spatially mapped within human and rodent lineages, maintained distance with respect to the PP domain. Therefore this domain does not fulfil the criteria of a classical enhancer/repressor in that distance and orientation appear to be constrained. It will be of interest to examine whether this UP1 constitutes the distal portion of the proximal promoter.

In order to identify PCRM specific TFBS and gain an insight into the potential context specific regulatory role of each domain, rVISTA 2.0 was utilised. This analysis identified an array of TFBS (Table1) that were constitutively enriched within respective domains in multiple species. The ultra-conserved PP domain contains a number of TFBS bound by factors with well characterised roles in striated muscle gene expression (Tables 2 and 3). These include multiple E-Box motifs, serum response element (SRE, bound by SRF) and a Nkx element (NKE, bound by Nkx 2.5). In addition this domain contained an ultra-conserved MEF2 motif (bound by Mef2 proteins) and a TATA box like element, which has been shown to bind TATA-binding protein *in vivo* (Lomvardas & Thanos, 2001; Soutoglou & Talianidis, 2002). The ultra-conserved nature of this TATA box like motif (TATT) coupled to its optimal spacing from the *msl* TSS (Lomvardas & Thanos, 2001; Soutoglou & Talianidis, 2002) leads us to speculate that this motif represents the bona fide *msl* TATA box and thus the location of TBP binding, Pol II recruitment and transcriptional initiation.

In addition to the PP domains, our rVISTA analysis also identified other striated muscle associated TFBS within the distal PCRM including MEF2, NFAT, E-Box, SMAD and GATA motifs. Considering the stringent striated muscle expression profile of *msl* and the unbiased nature of the rVISTA analysis, it is encouraging that the majority of rVISTA derived motifs represent binding sites associated with striated muscle specific transcription factors (Tables 1 - 3). If we were to retrospectively predict *msl* expression based solely on the TFBS identified through this rVISTA analysis, we would undoubtedly suspect a muscle specific expression profile. The current data validates the utility of rVISTA and suggests that if the appropriate whole genome sequences are

available and the PWM databases are sufficiently informative (large information content) and refined, it is possible to predict regulatory context based on sequence alone.

Due to the inherent content limitation of the TRANSFAC PWM library utilised by rVISTA we also executed a manual TFBS interrogation of all four PCRM. This manual filtering protocol combined MatInspector based PWM pattern matching with phylogenetic conservation and biologically relevant binding site information (our striated muscle TFBS library, Table 2). This strategy identified 28 TFBS within all four PCRM filtering out 86% of the original hits. These hits included all rVISTA derived motifs and additional motifs for GATA proteins, Mef2 proteins, KLF factors (Oka et al, 2007) and also a single Pbx binding motif (Pbx/Meis complex plays a crucial role in early myogenic gene expression (Berkes et al, 2004; Maves et al, 2007)). Due to the enrichment of all our rVISTA derived motifs we suspect this manual protocol has the same level of stringency and specificity as rVISTA. We therefore suggest that the additional motifs identified here are likely to have the same potential for functionality as the rVISTA derived hits, which have empirically been shown to represent prime candidates for functional TFBS (Loots et al, 2002).

The rVISTA and MatInspector filtered data-sets were combined to generate a final library of PCRM and TFBS for functional experimental analysis. In addition to focusing the experimental strategy, through a literature based interrogation of the putative TFBS and the specific clusters they form, it is also possible to gain a functional insight into the regulatory role of each PCRM. For example within the core promoter we have the ultra-conserved TATA like box which we speculate constitutes the core promoter. In addition we have six ultra-conserved Mef2 and GATA motifs within the proximal 200bp upstream of the TSS, which can be bound by Mef2 and GATA factors respectively. It is noteworthy that extensive studies of cardiac-specific gene regulatory networks implicate Mef2 binding proteins (specifically Mef2D, Mef2C) and GATA factors (4/5/6) as key transcriptional regulators within the heart (Bi et al, 1999; Black & Olson, 1998; Charron et al, 2001; Hautala et al, 2001; Liang et al, 2001; Lu et al, 2000; Molkenstein & Olson, 1996; Oka et al, 2006; Oka et al, 2007; Pikkariainen et al, 2004; Shore & Sharrocks,

1995). GATA factors are Cys4 zinc finger transcription factors with GATA4/5/6 all playing overlapping roles in the development and normal function of the heart. This is through direct transcriptional regulation and not surprisingly many cardiac gene proximal promoters (like the present *msl* PP domain) are enriched with multiple GATA motifs which mediate developmental, basal and stress inducible cardiac specific expression (Hautala et al, 2001; Kerkela et al, 2002; Morimoto et al, 2000; Morisco et al, 2001). It will be intriguing to determine the role of these GATA factors in regulating the *msl* promoter in such cardiac regulatory contexts. Examining the role of the ultra-conserved -135bp Mef2 motif will also be of interest. Members of the Mef2 family of MADS-box transcription factors are expressed at high levels in striated muscle and Mef2C is critical for correct cardiac development and associated cardiac gene expression (Han et al, 1997). Interestingly Mef2 factors and specifically Mef2C have been shown to interact with GATA factors to induce cardiac transcription (Yanazume et al, 2003)). The co-localisation of ultra-conserved Mef2 and GATA motifs raises the possibility for such co-operation at the *msl* promoter.

Apart from their role in heart-specific gene regulation, Mef2 proteins are also expressed in other tissues including skeletal muscle. Alone, Mef2 factors do not possess myogenic activity but, in combination with E-Box binding bHLH transcription factors, the myogenic regulatory factors (MRF), they drive and amplify the myogenic differentiation program (Black & Olson, 1998; Brand-Saberi, 2005; Shore & Sharrocks, 1995). With numerous E-Box and Mef2 motifs within the UP1/PP encompassing the proximal interval it would be intriguing to determine if these motifs are involved in skeletal myogenic specific activity of the *msl* promoter.

A single serum response element (SRE -305/-288) is present in the distal portion of the PP. This raises the intriguing possibility that SRF, the core component of the MS1-MRTF-SRF transcriptional axis, regulates the *msl* promoter thereby generating a positive feedback loop which supports a feedforward network motif. This may have implications during striated muscle development and adaptation to stress because such network motifs are central to the amplification and consolidation of the genetic pathways activated in such contexts. SRF is also a primary driver of cardiac and skeletal

muscle specific gene transcription and may contribute to *msl* expression in either or both of these contexts (Arai et al, 2002; Balza & Misra, 2006; Miano, 2003; Sotiropoulos et al, 1999). With respect to cardiac specific transcription the NKE element (-195/-188) and KLF binding site (-41/-36) may represent motifs of experimental interest. The NKE can be bound by Nkx2.5, a member of the NK homeobox family that plays a critical role in cardiac development and specifically cardiac gene transcription (Lints et al, 1993). In addition the KLF binding site can be bound by numerous cardiac regulatory KLF factors including KLF 5, 15 and 13, all of which have characterised regulatory function at cardiac specific promoters (Fisch et al, 2007; Suske et al, 2005). Both Nkx2.5 and KLF factors act in combinatorial manner interacting with other cardiac restricted factors, for example GATA4, SRF and Mef2C (see Chapter 1). The close proximity of these motifs with respect to the ultra-conserved GATA and Mef2 sites in addition to the SRE suggests such combinatorial interactions may also be active and functionally required at the *msl* promoter.

The significant enrichment of cardiac and skeletal specific TFBS within the proximal promoter leads to the hypothesis that the PP domain (and potentially the UP1 domain) represents the basal striated muscle specific proximal promoter. We suspect this region is critical and necessary for all regulatory contexts in which *msl* is expressed, much like the proximal promoter domains in the ANP, BNP and Ncx promoters. However distal CRM are typically used for the integration and subsequent transduction of the appropriate transcriptional output in response to the input signal (Firulli & Olson, 1997). *MsI* associated contexts include transcriptional changes in response to tempo-spatial developmental signals or post-natal transcriptional induction in response to stress stimuli. With this in mind it is of interest that the UP2 and UP3 distal PCRM contain specific clusters of enriched TFBS that have been implicated with and demonstrated to mediate the context specific regulation of other genes in striated muscle regulatory contexts. For example the UP2 domain contains NFAT and MEF2 motifs which have been demonstrated to collaborate with each other and mediate calcineurin transcriptional signalling (Beals et al, 1997; Hogan et al, 2003; Molkenstein et al, 1998). In a skeletal muscle context this signalling axis is important for the up-regulation of specific skeletal muscle transcripts associated with slow twitch muscle fibre (Schulz &

Yutzey, 2004). Kuwahara and colleagues (2007) have recently demonstrated that the *msl* proximal promoter is more active in the slow twitch muscle fibre *in vivo* (Kuwahara *et al*, 2007). In addition, cardiac specific transcriptional up-regulation in response to pressure overload or ischemia reperfusion injury (both contexts in which *msl* is differentially expressed) is critically dependent on the calcineurin/NFAT Ca^{2+} dependent pathway. It will be of great interest to determine whether the UP2 PCRM regulates *msl* expression in any of these regulatory contexts and whether it confers Ca^{2+} dependent calcineurin signalling onto *msl*.

The UP3 PCRM is the most evolutionary conserved distal PCRM and is of particular interest due to its TFBS composition, which it shares similarity with the Nkx2.5 cardiac developmental enhancer (Lien *et al*, 2002). This Nkx2.5 enhancer, which contains tandem Smad binding elements in close proximity to multiple GATA binding motifs, mediates early embryonic cardiac specific induction of Nkx2.5 (at E 7.5) through a BMP signalling pathway. Kuwahara and colleagues (2007) have demonstrated that *msl* is embryonically expressed within the linear heart tube at E8.25. This raises the possibility that the UP3 PCRM may represent a developmental cardiac specific enhancer which mediates *msl* transcriptional induction in response to development BMP/Smad signalling.

In conclusion, through comparing the TFBS composition of the current identified PCRM with well characterised CRM in the literature it is possible to predict what regulatory roles they may confer on *msl*. Clearly such an insight will allow us to generate appropriate experimental models, both *in vitro* and *in vivo*, for subsequent analysis.

Chapter 4

Functional and epigenetic analysis of conserved cardiac specific *cis* regulatory domains in the *msl* gene.

4.1 Introduction

As outlined in Chapter 1, characterising the mechanisms governing context specific expression profiles of *msl* will give a deeper insight into *msl* function, and the transcriptional regulatory processes associated with such contexts, for example, cardiac specific gene expression. The unique striated temporal-spatial transcriptional profile of *msl* expression is likely to be under the control of a complex regulatory hardwiring mechanism. In recent years a unifying paradigm for the temporal and spatial regulation of cardiac specific gene expression has emerged (Firulli & Olson, 1997). It has been demonstrated through the functional analysis of numerous cardiac specific genes (Brown et al, 2004; Lien et al, 2002; Sepulveda et al, 1998; Xu et al, 2006) that developmental and post natal context specific expression patterns are mediated by the binding of cell/tissue specific and signal sensitive transcription factors acting in a combinatorial manner at many distinct and independent *cis* regulatory modules (CRMs) (Levine & Tjian, 2003). These modular circuits serve to integrate all the appropriate signals with the proximal core promoter and subsequently modulate the transcriptional output. The integration of these signals through the appropriate TFs and regulatory domains is mediated, in a large part, through upstream intracellular signalling pathways and epigenetic regulatory phenomena (Clerk et al, 2007). This epigenetic modification of chromatin modulates transcription through creating the optimum environment for Pol II binding and transcriptional initiation (Soutoglou & Talianidis, 2002). The best characterised modifications associated with cardiac epigenetic regulation are histone acetylation and de-acetylation (Backs & Olson, 2006; Wang et al, 1999).

Regulation of histone acetylation/deacetylation has been linked to cardiac development and hypertrophy, contexts in which *msl* is differentially expressed. For example, the

HAT activity of p300 and CREB-binding protein (CBP) is required for phenylephrine induced cardiac hypertrophy (Yanazume et al, 2003). In support of this observation, inhibition of HDAC activity potentiates the hypertrophic response in muscle cells (Iezzi et al, 2004). Specifically it has been demonstrated that class II HDACs (HDAC-4, -5, -7 and -9) suppress cardiac specific gene expression in part by binding to and inhibiting the activity of myocyte enhancer factor 2 (MEF2)(de Ruijter et al, 2003; Kee et al, 2006; Lu et al, 2000; Miska et al, 1999; Wang et al, 1999; Zhang et al, 2002). Contrary to this, Class I HDACs (HDAC2) have been found to promote cardiac hypertrophy *in vivo*, presumably through specific transcriptional repression of anti-hypertrophic genes (Kook et al, 2003; Kook et al, 2002). Despite the somewhat discrepant findings with respect to the pro/anti-hypertrophic role of the various HDAC classes, it is clear histone acetylation *per se* is a critical determinant of cardiac specific gene expression. Recent studies have also demonstrated an important role for histone methylation and demethylation in the regulation of cardiac specific gene expression (Bingham et al, 2007; Bingham et al, 2006). Acting through the repressor element-1 silencing transcription factor (REST), Bingham and colleagues (2007) clearly demonstrated that differential histone methylation is required for basal and stress inducible cardiac expression of ANP and BNP.

The specific activity of TFs, co-factors and subsequent epigenetic regulatory processes acting at multiple independent CRMs will contribute to the exquisite temporal-spatial expression profile demonstrated by *msl*. With this in mind it was encouraging that the present bioinformatic analysis of the *msl* promoter and associated regulatory domains (Chapter 3) uncovered an array of highly conserved putative *cis* regulatory modules (PCRM) which we speculated mediated context specific cardiac expression of *msl*. The nature of these cardiac TFs which we predict may bind our identified domains leads us to propose that an important component of *msl* transcriptional regulation will involve epigenetic regulatory mechanisms. For example, the TFs GATA4, MEF2, NFAT and SRF, all of which have predicted binding sites within our regulatory domains, have been demonstrated to mediate cardiac specific expression through epigenetic mechanisms (Oka et al, 2007).

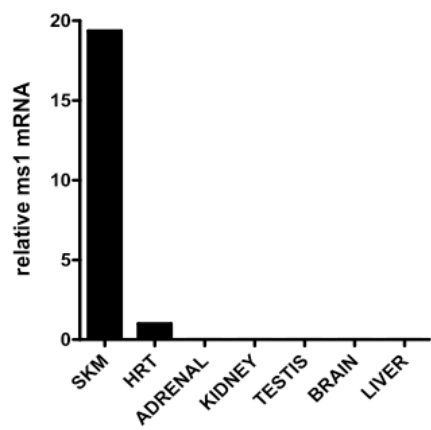
In order to gain a greater understanding of the cardiac specific mechanisms governing *msl* expression, we have proceeded to isolate and functionally analyse our *in silico* derived PCRM (Chapter 3) for cardiogenic activity. Throughout the analysis we have also investigated the epigenetic status of our identified domains as it was predicted that epigenetic regulatory processes represent major determinants of *msl* cardiac specific expression. The results derived from this present chapter will serve to illuminate our understanding of the regulatory circuitry underpinning *msl* cardiac specific expression. The epigenetic analysis also has the potential to give an insight into the composition of potential factors binding within each domain and the mechanisms for context specific expression, for example, developmental up-regulation and the transient post natal induction in response to stress. Finally these data will also allow us to determine the validity and accuracy of the current comparative *in silico* analysis with respect to predicted cardiac specific CRMs, in addition to focusing our attention in subsequent investigation for TFBS experimental analysis.

4.2 Results

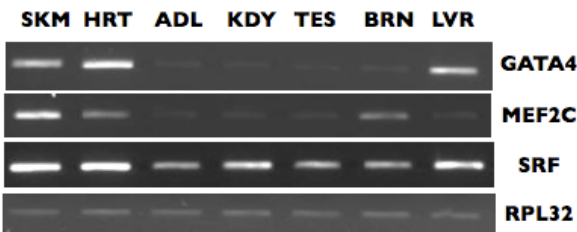
4.2.1 Quantitative analysis of cardiac specific expression of *msl* in vivo and in vitro.

During the initial identification and characterisation, *msl* expression was shown, using semi-quantitative RT-PCR (Mahadeva et al, 2002), northern blot (Arai et al, 2002) and whole embryo *in situ* hybridisation (Arai et al, 2002), to be restricted to striated muscle with significant cardiac enrichment. However due to the qualitative nature of these assays, transcript abundance of *msl* in the same set of tissues (Mahadeva et al, 2002) was measured using a quantitative assay. Using total pooled RNA (n=5) isolated from adult rat skeletal muscle (solous muscle), heart, adrenal gland, kidney, testis, brain and liver (isolated by Harin Mahadeva), quantitative real time RT-PCR of *msl* and *TBP* (internal control) using Taq Man® probes was performed as described in Chapter 2. Relative *msl* mRNA was quantified using the comparative method (Livak & Schmittgen, 2001) with expression in the heart normalised arbitrarily to one. Consistent with the published work, *msl* expression is restricted to striated muscle with significant expression in skeletal and cardiac muscle (Figure.4.1.A). Our quantitative analysis indicated a 20-fold increase in relative abundance of *msl* in skeletal versus cardiac muscle. To validate the striated muscle specificity of the RNA samples used, semi-quantitative RT-PCR of *GATA4*, *MEF2C*, *SRF* and *RPL32* (internal control) was performed as described in Chapter 2 (Figure.4.1.B). *GATA4*, *MEF2C* and *SRF* all represent cardio and myogenic transcription factors which are typically more abundant in striated muscle. The high relative abundance of all three muscle transcription factors within heart and skeletal muscle validates the striated muscle nature of the RNA samples examined.

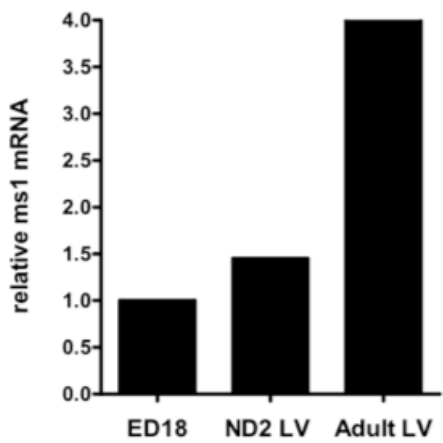
A



B



C



D

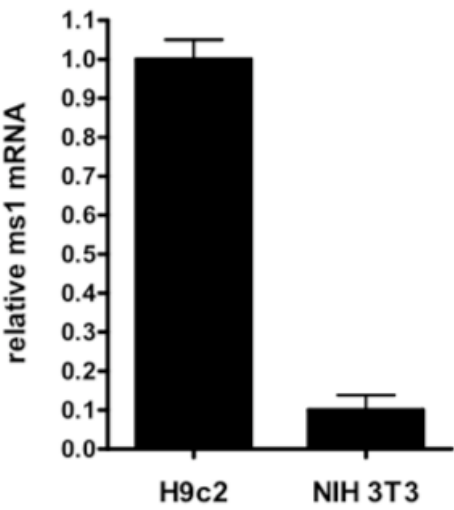


Figure 4.1. Relative tissue, developmental and cell specific ms1 mRNA abundance. Total RNA (5 separate preparations which were then pooled) was isolated from adult rat skeletal muscle, cardiac muscle, adrenals, kidney, testis, brain, liver in addition E18 whole hearts and 2 day neonate left ventricles and subjected to reverse transcription followed by quantitative (A, C) and semi-quantitative (B) PCR with primers specific for MS1 (A), GATA4 (B), Mef2C (B), SRF (B), TBP and RPL32. Relative expression in the adult heart for ms1 was arbitrarily set at 1. Total RNA was also isolated from H9c2 myoblasts and NIH3T3 and subject to quantitative real time RT-PCR (D). Expression level in H9c2 myoblasts was arbitrarily set at 1. Values are represented as means \pm SEM of at least three independent RNA extractions.

Ms1 expression during cardiac embryonic and post-natal development was also measured. Using pooled RNA (n=5) isolated from the left ventricles of 18-day-old embryos, 2-day-old neonates and 12-week-old adult Wistar rats (tissues isolated and RNA extracted by Harin Mahadeva), quantitative real time RT-PCR was executed. Relative *ms1* expression in the embryonic ventricle was arbitrarily normalised to one. This analysis indicated a clear maturation dependent increase in *ms1* expression (Figure. 4.1.C) with a 1.5 and 4-fold increase in relative *ms1* transcript abundance in the neonatal and adult ventricles respectively.

To characterise the cardiac specific mechanisms governing this strict developmental and cell restricted expression, we required an appropriate cardiac *in vitro* cell model and the H9c2 cell line was chosen. These cells, which are derived from the embryonic rat ventricle, represent an immortalised cardiac myoblast cell line capable of proliferation and differentiation. They express cardiac specific isoforms of L-type voltage operated Ca^{2+} channels and sarcolemmal ATPase splice variants characteristic of the normal heart as well as displaying cardiac specific electrophysiological properties (Hammes et al, 1994; Sipido & Marban, 1991). More significant to the current study, these cells have been effectively used for the characterisation of cardiac specific gene regulatory mechanisms and epigenetic phenomena, the contexts in which we want to examine *ms1* expression (Bingham et al, 2007; Bingham et al, 2006; Kaneda et al, 2005).

In order to confirm the utility of these cells for cardiac specific regulatory characterisation, real time RT-PCR was performed on total RNA isolated from the H9c2 myoblasts and the non-cardiac NIH 3T3 fibroblast cell line. *Ms1* expression in H9c2 was normalised to one and encouragingly, *ms1* transcript was ten times more abundant in H9c2 than in the NIH 3T3 non-cardiac fibroblasts (Figure.1.D). This suggests common cardiac specific mechanisms are driving *ms1* restricted expression *in vitro* and hence validates the use of these cells as an appropriate cardiogenic model for *ms1* experimental characterisation herein.

4.2.2 Isolation and functional cardiac characterisation of proximal 5'-flanking regulatory domains.

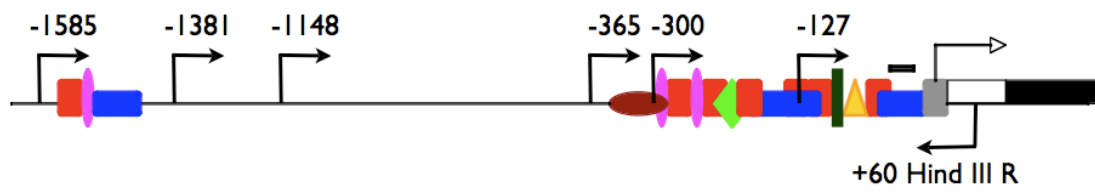
The bioinformatic comparative *in silico* analysis (Chapter 3) led us to speculate that the proximal 5'-flanking interval, in particular the UP1 and PP domains, is a crucial determinant for *msl* cardiac specific expression. This was, in part, as a result of the high number of cardiac restricted TFBS enriched within these conserved domains. In order to test this hypothesis *in vitro*, we proceeded to isolate and clone the proximal 1.6kbp 5'-flanking region, encompassing the UP1 and PP domains, directly upstream of Luciferase in a promoter reporter vector system (pGL3-Basic). The premise behind this assay and the detailed description of reporter construct generation are described in detail in Chapter 2. Briefly, specific combinations of various forward and common reverse primers (+60-Hind III)(Figure 4.2.A) were used to amplify 5'-flanking intervals for subsequent cloning into the pGL3-Basic luciferase reporter plasmid. The forward primers were specifically designed to encompass specific 5'-flanking intervals, thereby allowing a series of truncated reporter constructs to be generated in which specific hypothesised functional domains (UP1 and PP) and TFBS have been truncated (Figure 4.2.A). For example the -1381/+60 reporter does not contain the UP1 domain. In addition the -300/+60 and -127/+60 reporters truncate within the highly conserved PP domain resulting in loss of specific TFBS. We have also generated -1148/+60 reporter to give us an insight into the role of any non conserved functional domains located with the inter-domain genomic region. The truncated reporters were restriction digested with SacI/Hind III and electrophoresed on an agarose gel for visualisation to confirm integrity and validate size of the truncated promoter fragment insert (Fig 4.2.B). These reporters were also sequenced as described in chapter 2 to ensure no sequence aberrations were incorporated during the PCR amplification step.

The truncated reporters were transiently transfected into the cardiogenic H9c2 and non cardiogenic NIH 3T3 and COS-7 kidney cell lines, as described in the methods. In agreement with the *in vivo* and *in vitro* cardiac restricted expression profile of the *msl* promoter, all the truncated reporter constructs were significantly more active in the cardiogenic H9c2 with respect to the non-cardiogenic COS-7 and NIH 3T3 (Figure

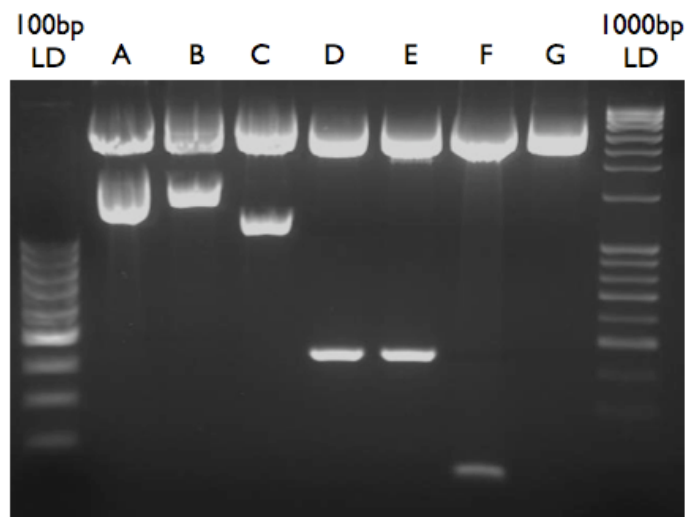
4.2.C). The relative luciferase activity of the -1585/+60 reporter was normalised to one. The respective reporter is approximately 75% less active in the COS-7 and NIH 3T3, which suggest that cardiac restricted factors activate the *msl* promoter through domains and motifs within the proximal flanking interval, as predicted *in silico*. However deletion of the -1585/+60 to -300/+60 reporter does not result in attenuated cardiac specific reporter activity. This implies that all the *cis* information required for cardiogenic activity in H9c2, lies within the proximal 300bp upstream of the TSS. This suggests the UP1 domain, in this context, is not of functional importance. However, in support of our *in silico* derived hypothesis, truncating within the ultra conserved PP domain, leads to attenuated cardiac specific reporter activity. Indeed the -127/+60 fragment is approximately 50% ($p < 0.05$) (Figure 4.2.C) less active than the -300/+60 reporter, with this attenuation only observed in the H9c2 and not the NIH 3T3 or COS-7 cells. This cardiac specific decrease in promoter activity is therefore likely caused by the deletion of key cardiac TFBS located within the -300 to -127 genomic sequence interval.

The PP domain, which encompass (and overlap) the TSS and therefore represents the location of the core promoter, with the key component being the TATA box, or alternative sites of TATA binding protein association (Roeder, 2005). Our *in silico* analysis identified a putative TATA box (TATT) located 40bp upstream of the TSS which represents optimal inter-spacing for *bona fide* TATA boxes (Smale & Kadonaga, 2003). In order to test this hypothesis we mutated this putative TATA box within the -1585/+60 reporter using site directed mutagenesis exchanging the adenine at the second position for a guanine, thereby abolishing the TATA box. This mutation resulted in a dramatic 95% decrease in promoter activity in H9c2 cells (Figure 4.3.A. black bar) with similar dramatic decrease in activity in NIH 3T3 fibroblasts and COS-7 cells (Fig 4.3.A, checked and clear bars). This ubiquitous abolishment of reporter activity in all cell types suggests this is not a cell specific effect and indicate this motif may represent a genuine TATA box.

A



B



C

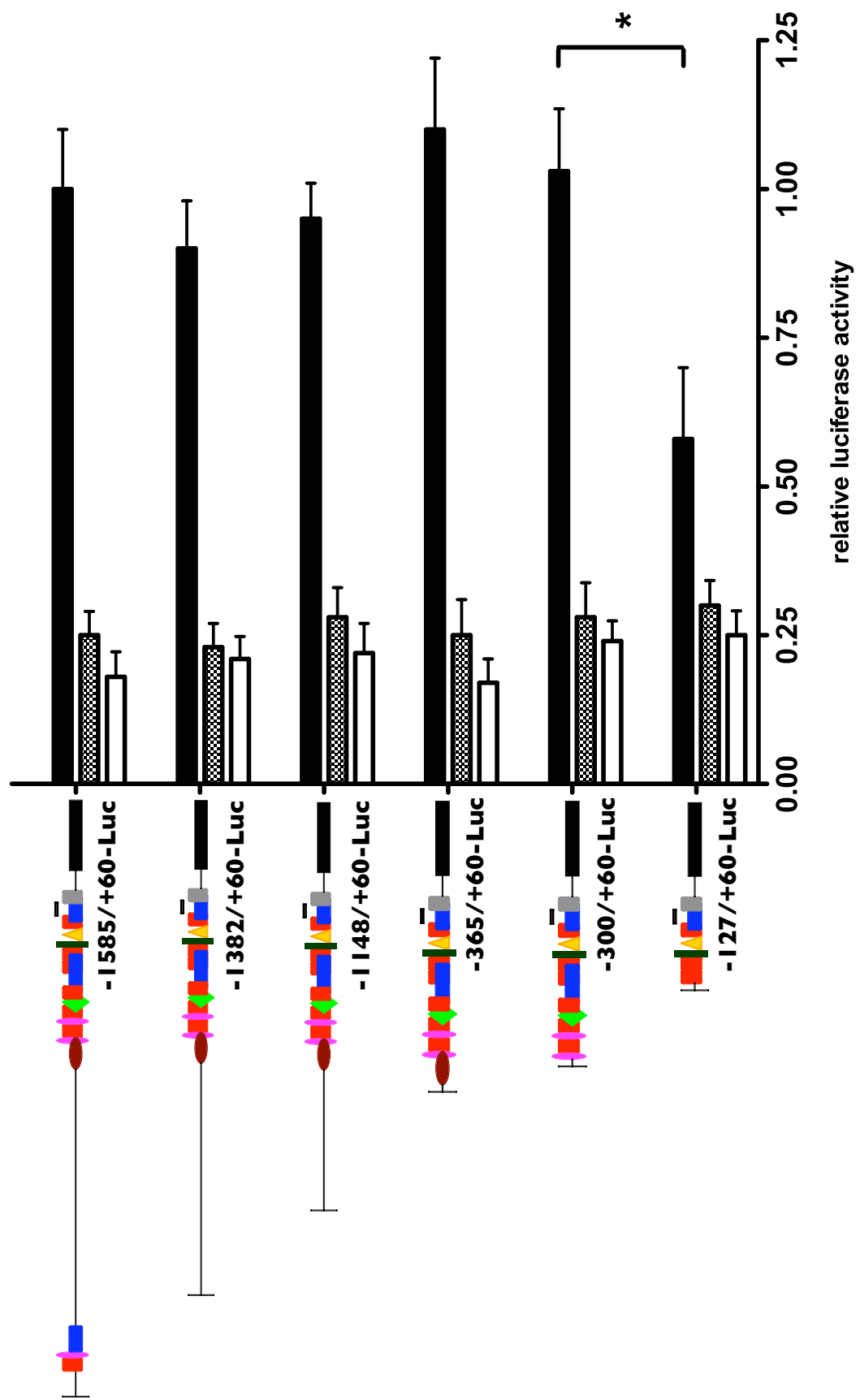
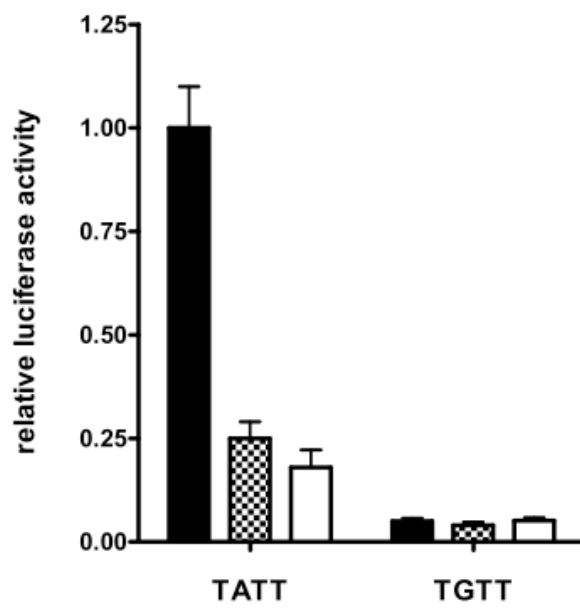


Figure 4.2. Cell specific activity of the ms1 proximal promoter. (A) Schematic representation of the primer locations used for the generation of truncated ms1 proximal promoter reporter luciferase constructs. (B) Generated luciferase constructs were restriction digested and visualised via agarose gel electrophoresis for confirmation of correct orientation and insertion. A; -1381/+60 B; -1585/+60 C; -1148/+60 D; -365/+60 E; -300/+60 F; -127/+60 G; pGL3-B. The truncated promoter reporters were transiently transfected into H9c2 myoblasts (solid black bar), NIH3T3 fibroblast (grey dotted bar) and Cos-7 cells (white bar). 48 hours post transfections at approximately 80% confluence cells were harvested for luciferase activity. Promoter activity relative to pGL3-B alone was determined for the truncated reporter constructs in each cell line. The relative activity of the truncated promoters (vs pGL3-B) in each cell line was then determined with relative activity of the P-1585/+60 reporter in H9c2 myoblasts assigned arbitrary value of 1. The results are expressed as mean \pm SEM of at least three different separate transfections in triplicate for each reporter construct. Statistically significant differences are indicated by * $p < 0.05$.

A



B

+	+	+	+	Labeled Probe
-	+	-	-	Unlabeled self
-	-	-	+	Unlabeled consensus

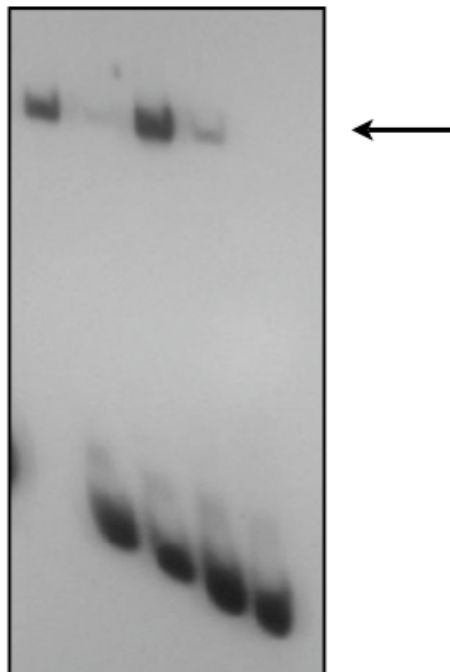


Figure 4.3. Proximal ms1 TATA like sequence represents *bona fide* TATA box. (A) The putative TATA box sequence was submitted to site directed mutagenesis and transiently expressed in H9c2 myoblasts (solid black bar), NIH 3T3 fibroblasts (dotted grey bar) and Cos-7 cells (white bar) and assayed for luciferase activity. TATA mutation resulted in a ubiquitous >95% reduction in luciferase activity with respect to the wild type P-1585/+60 reporter. The results are expressed as mean \pm SEM of at least three different separate transfections in triplicate for each reporter construct. (B) DIG-labelled oligonucleotide probe for the MS1 TATA like sequence was incubated with whole cell protein extracts made from H9c2 myoblasts. Competition experiments were performed using a 200-fold excess of unlabelled self to test for specificity and a 200-fold excess of unlabelled TATA consensus probe. Arrows indicate the resulting band shifts.

To further characterise the biological importance of the proximal TATA box like sequence, we executed an electrophoretic mobility shift assay (EMSA). Briefly, we synthesised a specific oligonucleotide encoding the TATA box like sequence and adjacent nucleotides present at the endogenous *msl* promoter. These complimentary oligonucleotides were annealed and end labelled with digoxigen (DIG) to generate double stranded labelled probe for subsequent EMSA analysis. This DIG-labelled probe was then incubated with whole cell protein extracts from H9c2 myoblasts and a molar excess (X200) of complimentary cold unlabelled probe or cold consensus TATA box. As shown in Figure 4.3.B incubation of the *msl* TATA box probe with H9c2 whole cell extract resulted in the appearance of a single DNA-protein band shift. Co-incubating this reaction with a 200 fold molar excess of unlabelled complimentary probe ablated the appearance of the shifted band, demonstrating the DNA-protein band shift is the product of specific protein/DNA-sequence specific interaction, and not a non specific protein interaction. This shifted band was also ablated through incubation with 200 fold molar excess of unlabelled consensus TATA box probe. These findings are in agreement with our mutagenesis analysis and it is concluded that this TATT sequence represents the bona-fide *msl* TATA box, and therefore site of TBP binding and transcriptional initiation.

4.2.3 Cardiac specific regulatory activity is associated with the distal regulatory domains, UP2 and UP3.

In addition to the proximal interval encompassing the UP1 and PP domains, our *in silico* analysis identified two distal domains of particular interest, UP2 and UP3. Both of these distal loci were enriched with composite clusters of cardiac specific TFBS and this led us to speculate that these domains contributed to the cardiac specific expression of *msl*. Classically, distal CRMs, like UP2 and UP3 act as enhancer or repressor modules mediating context specific activation or repression of the proximal promoter, which, with respect to *msl* is likely the PP domain.

In order to test this hypothesis *in vitro*, we proceeded to isolate and clone the distal UP2 and UP3 domains directly upstream of the SV40 basal promoter in a luciferase based

enhancer/repressor reporter system (pGL3-Promoter). Unfortunately, due to the inability to amplify and clone promoter fragments larger than 10 kbp, it was not possible to test the functionality of the UP2 and UP3 (located 8.7 and 16.4 kbp upstream of ms1 TSS respectively) domains in their native genomic context, co-linear with the ms1 PP. However, bona-fide enhancer/repressor regulatory modules can, by definition, have common regulatory effects on any given ubiquitous proximal promoter. Thus, through cloning these modules upstream of the SV40 basal promoter, we can probe and gain an insight into the cardiac regulatory activity of these distal modules and determine whether they mediate enhancing/repressing effects.

A detailed description of reporter construct generation is described in Chapter 2. Briefly, specific forward and reverse primers designed to encompass the UP2 (-9498-XhoI/-8608-HindIII) and UP3 (-16702-XhoI/-16432-HindIII) (Figure 4.4.A) were used to amplify the respective genomic intervals for subsequent cloning into the pGL3-Promoter luciferase reporter plasmid. These enhancer/repressor reporter plasmids were restriction digested with XhoI and HindIII and electrophoresed on an agarose gel (Figure 4.4.B) for visualisation to confirm integrity and validate size of the appropriate insert.

These validated reporters, in addition to the SV40 only reporter, were transiently transfected into the cardiogenic H9c2 (Figure 4.4.C) and non-cardiogenic NIH 3T3 (Figure 4.4.D) cell lines as described in Chapter 2. The relative luciferase activity of the SV40 only reporter was normalised to one. Within the cardiogenic environment (Figure 4.4.C) the UP2 and UP3 domains mediate clear regulatory effects on the SV40 proximal promoter. The UP2 module repressed the activity of the SV40 basal promoter by approximately 40% (* $p < 0.05$), suggesting in this cardiomyoblast context, UP2 has repressive activity. Conversely the UP3 module, in this cardiac context, enhances basal SV40 promoter activity by approximately 75% (** $p < 0.01$).

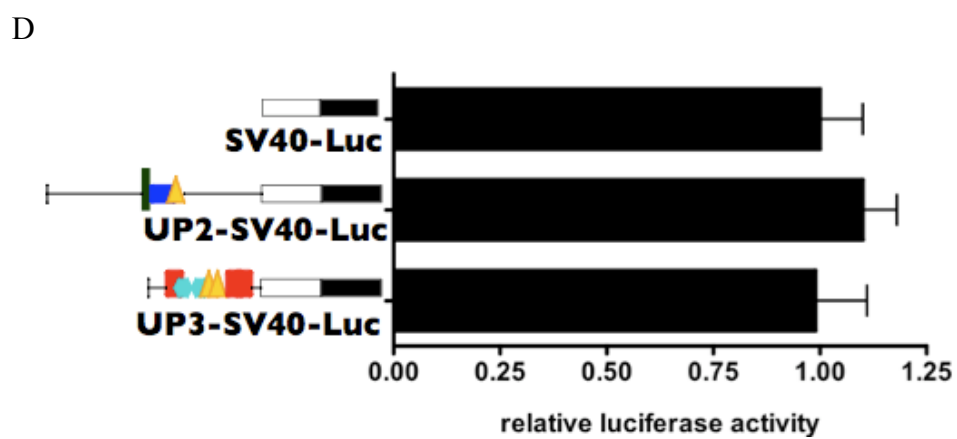
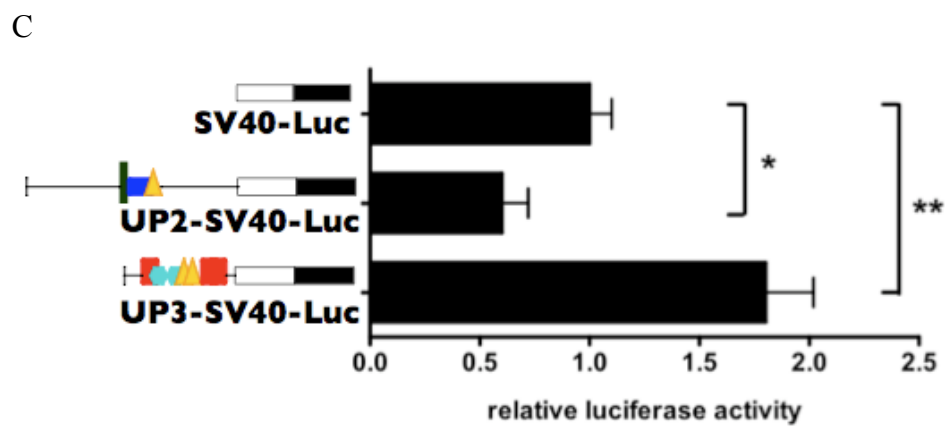
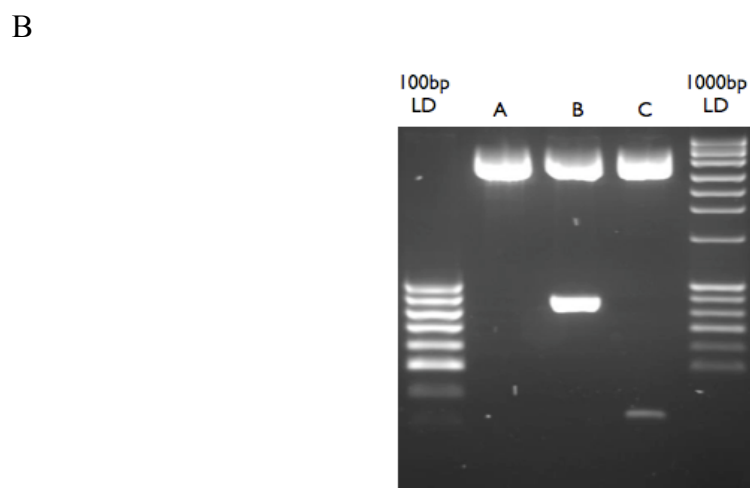
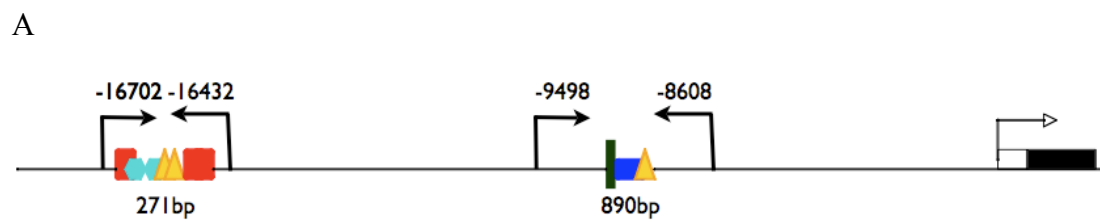


Figure 4.4. Cell specific activity of the ms1 distal domains UP2 and UP3. (A) Schematic representation of the primer locations used for the generation of the distal domain reporter luciferase constructs. (B) Generated luciferase constructs were restriction digested and visualised via agarose gel electrophoresis for confirmation of correct orientation and insertion. A; pGL3-P, B; -9498/-8608, C; -16702/-16432. The domain reporters were transiently transfected into H9c2 myoblasts (C) and NIH3T3 fibroblast (D). 48 hours post transfections at approximately 80% confluence cells were harvested for luciferase activity. Domain reporter activity relative to SV40 basal promoter was determined with SV40 activity arbitrarily set at 1 in both cell lines. The results are expressed as mean \pm SEM of at least three different separate transfections in triplicate for each reporter construct. Statistically significant differences are indicated by * $p < 0.05$, ** $p < 0.01$.

These respective enhancing/repressive effects are not observed in the NIH 3T3 cells (Figure 4.4.D) therefore clearly demonstrating cardiac specificity of these modules which is likely mediated by cardiac specific TF binding and associated regulatory processes. In conclusion these findings suggest that, at-least in H9c2 based cardiogenic environment, the UP2 and UP3 distal domains represent bona-fide repressor and enhancer regulatory modules respectively.

*4.2.4 Cardiogenic differentiation-dependent activation of *msl* expression: potential role for the proximal promoter domain.*

Previous work coupled to the current quantitative expression analysis (Figure 4.1.C) has clearly demonstrated that *msl* expression is differentially regulated during cardiac development and maturation with a robust induction during the neonatal to adult transition. This represents a transition from a proliferating neonatal myoblast-like cell to a terminally differentiated mature myocyte which is accompanied by hypertrophy and myofibrillogenesis. In this respect the H9c2 myoblast cells are intrinsically suited for the study of this maturation specific transcriptional induction of *msl* and the regulatory mechanisms governing this. As previously described H9c2 is a clonal myoblast cardiac cell line derived from the embryonic rat heart (Kimes & Brandt, 1976). When confluent, these myoblasts have the ability to trans-differentiate into a more mature cardiac muscle phenotype, similar to the neonatal to adult transition (Menard et al, 1999). This retinoic acid (RA) induced trans-differentiation is accompanied by the expression of cardiac isoforms of L-type voltage-operated Ca^{2+} channels, *Gai3*, *Gai3/0*, myosin light chain 2V and sarcolemmal ATPase splice variants characteristic of the normal adult heart.

In order to evaluate the suitability of H9c2 as a model for maturation dependent cardiac specific *msl* up-regulation, quantitative real time RT-PCR of *msl* and TBP (internal control) using Taq Man® probes was conducted using 1µg of total RNA isolated from proliferating H9c2 myoblasts (maintained in growth media: GM.) and terminally differentiating H9c2 myotubes (upon confluence maintained in cardiogenic

differentiation media: CDM)(Figure 4.5.A) as described in Chapter 2. There was a significant four fold ($p<0.05$) increase in *msl* transcript in differentiating H9c2 compared to proliferating H9c2 myoblasts, thus validating the use of H9c2 as an appropriate cardiogenic differentiation model. To confirm the fidelity of this cardiogenic differentiation, quantitative and semi-quantitative RT-PCR was also used to measure Arc and α_1 cardiac gene expression, respectively (Figure 4.5.B.C.D). Both of these transcripts, which have been demonstrated to be transcriptionally induced during H9c2 RA-induced cardiac differentiation (Menard et al, 1999), were significantly (4.5-fold for Arc and 3-fold; $p<0.05$) more abundant in the differentiated H9c2, thus validating the current differentiation protocol. In conclusion this finding suggests that the H9c2 cell line represents a good model for *msl* cardiogenic differentiation dependent up-regulation. We suspect the mechanisms governing *msl* expression in this model may correlate with the mechanisms mediating maturation dependent *msl* up-regulation *in vivo* during the neonatal to adult transition.

In order to gain an insight into the potential regulatory role of our cardiac specific regulatory domains in this process, we proceeded to test the effect of RA induced H9c2 cardiogenic differentiation on the relative promoter and fragment activity of the isolated promoter reporter constructs described previously (section 4.2.2 and 4.2.3). As a first step we transiently transfected the -1585/+60 and -365/+60 reporters encompassing the UP1 and PP domains respectively into proliferating (white bar, Figure 4.6.A) and differentiated H9c2 (black bar, Figure 4.6.A) cells. Using these two fragments allowed us to probe whether the UP1 domain mediates a differentiation dependent regulatory role not active in the non differentiated H9c2 cells and therefore not detected in the early experiment (Figure 4.2). The relative luciferase activity of the -1585/+60 reporter in non-differentiated H9c2 was normalised to one. In support of the proposed hypothesis outlined in Chapter 3, both the -1585/+60 and -365/+60 reporters were sensitive to the differentiation status of the cell with significant 25% ($p<0.05$) increase in relative luciferase activity with both reporters observed in the differentiated H9c2. This data also suggests that, at least in this model, the UP1 domain does not contribute towards the differentiation dependent increase in relative promoter activity. It is likely that this effect is mediated by the binding and activity of differentiation dependent

cardiac transcription factors within the PP domain. Somewhat surprisingly however, RA induced cardiac differentiation did not have any effect on the relative activity of the UP2 and UP3 domains (Figure.6.B), with both domains maintaining their enhancing and repressive effects on the SV40 basal promoter in the differentiated and non-differentiated cells. In conclusion these findings suggest that the differentiation dependent up-regulation in *ms1* expression observed during H9c2 cardiogenic differentiation is mediated, in part, by the PP regulatory domain. These findings do not rule out roles for UP1, UP2 or UP3 in cardiac differentiation *in vivo*, or alternatively in other cardiac regulatory contexts (e.g. stress).

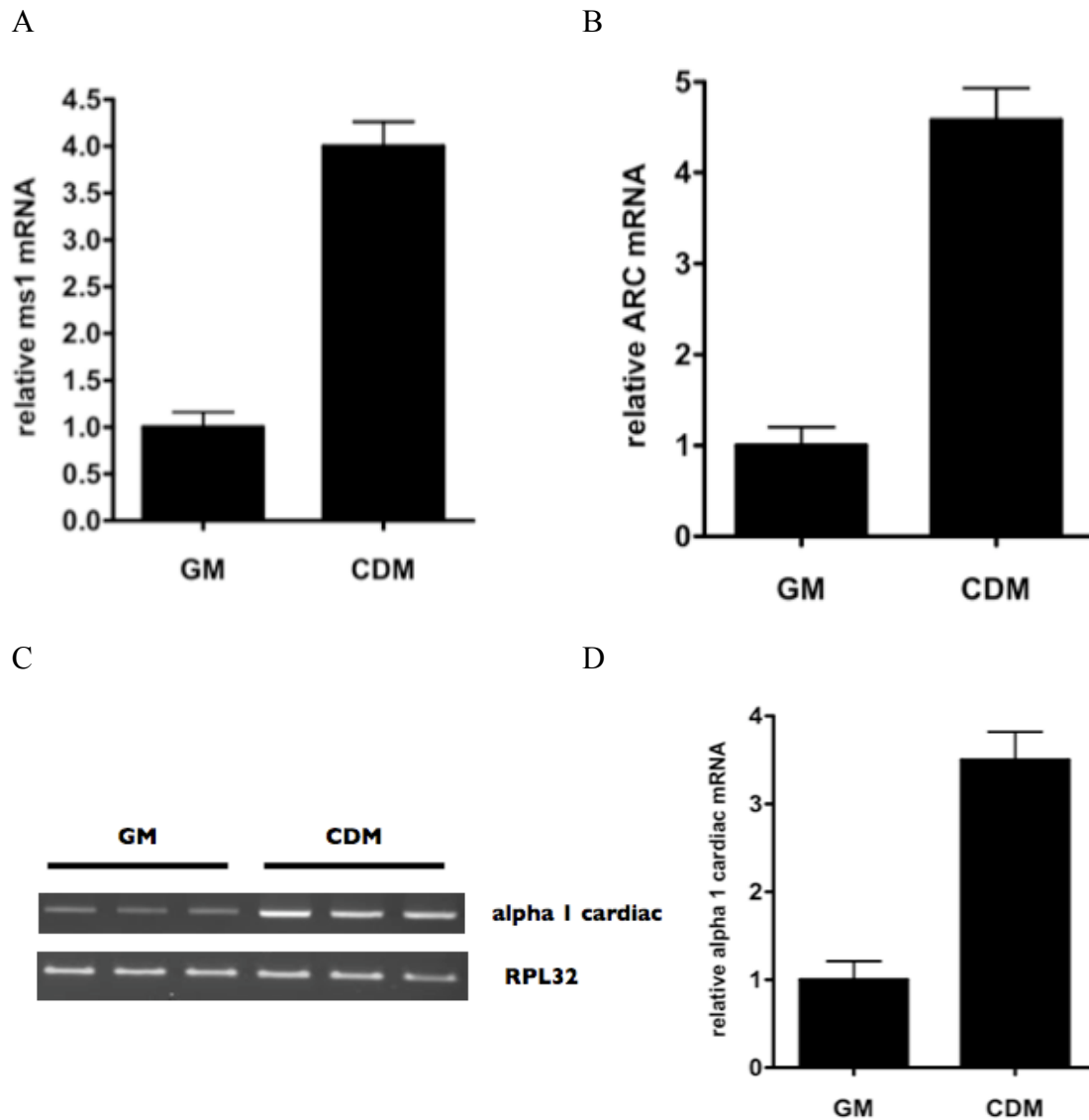


Figure 4.5. Quantitative real-time PCR analysis of ms1 gene expression during cardiogenic differentiation in vitro. Total RNA was isolated from H9c2 myoblasts (in growth media, GM) and myotubes (maintained in differentiation media upon confluence). The RNA was subjected to reverse transcription followed by quantitative PCR with rat ms1 (A), Arc (B) and TBP specific primers. Semi-quantitative PCR was executed using primers specific for alpha 1 cardiac calcium channel and RPL32 (C, D). Densitometric analysis of agarose gel (C) was quantified and relative expression plotted (D). Expression level of all transcripts (ms1, Arc, $\alpha 1c$) in growth media (GM) was arbitrarily set at 1. Values are represented as means \pm SEM of at least three separate experiments. Differences in relative expression of each transcript in CDM versus GM are statistically significant, $p < 0.05$.

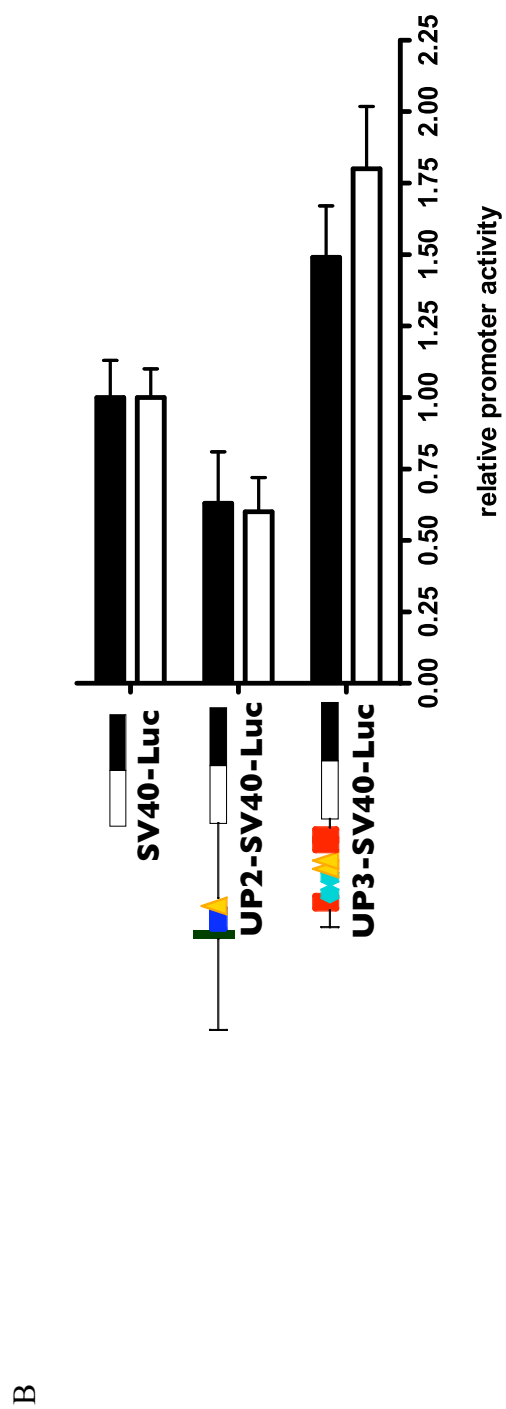
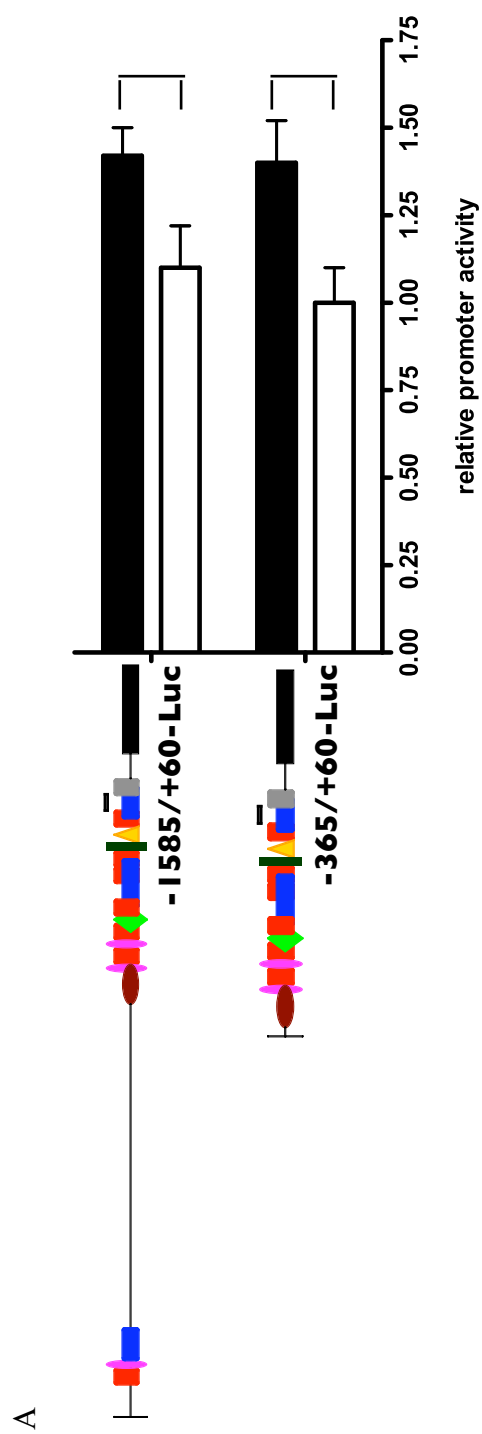


Figure 4.6. Cardiogenic differentiation dependant activity of the ms1 promoter and distal domains. H9c2 myoblasts (solid white bars) and differentiated myotubes (solid black bars) were transiently transfected with the -1585/+60 (A), -365/+60 (A), UP2 (B), UP3 (B) and SV40 (B) luciferase reporter constructs. Promoter activity relative to pGL3-B (A) or pGL3-P (B) alone was then determined in the myoblasts and myotubes with relative activity of the -365/+60 (A) and SV40 reporters in the myoblasts arbitrarily set at 1-fold activation. The results are expressed as mean \pm SEM of at-least three separate transfections in triplicate for each reporter construct. Statistically significant differences are indicated by tailed bars, $p < 0.05$.

4.2.5 *Epigenetic regulatory processes contribute to ms1 cardiac specific expression.*

The data presented so far in this chapter clearly demonstrates that cardiac specific and differentiation dependent *ms1* expression is modulated, at least in part, by activating and/or repressing the proximal flanking interval, specifically the PP domain, and the distal UP2 and UP3 domains. Such cardiac specific enhancing and repressive regulatory activity is typically associated with and coupled to specific epigenetic regulatory phenomena. We therefore suspect epigenetic processes, specifically those with well characterised cardiac specific functionality (specifically histone acetylation and methylation), to represent major determinants of *ms1* cardiac specific expression.

Histone acetylation/deacetylation represents the best characterised chromatin modification modulating cardiac specific gene expression, primarily through the Class I, Class II HDACs and the p300/CBP HAT. In order to investigate a role for histone acetylation in *ms1* transcription regulation, RT-PCR of *ms1* and TBP (internal control) using Taq Man® probes was conducted using 1µg of total RNA isolated from H9c2 treated with the general Class I and Class II HDAC inhibitor trichostatin A (TSA) (Figure 4.7.A). Incubating H9c2 myoblasts for 24 hours with both 100nM and 300nM TSA resulted in a significant 3-fold increase in relative *ms1* transcript abundance compared to relative expression in the vehicle control (*p<0.05). As a positive control we also demonstrated Arc and Per1 TSA dependent transcriptional induction (Figure 4.7.B.C.D), in agreement with published results (Kaneda et al, 2005). These findings support the notion that active HDAC dependent repressive mechanisms are acting on the *ms1* transcriptional unit, presumably through identified CRMs. In order to gain a greater insight into these potential acetylation dependent regulatory events, we proceeded, using chromatin immunoprecipitation (ChIP), to analyse the endogenous proximal *ms1* promoter for histone acetylation. In addition to this proximal domain we were also interested in the distal putative cardiac enhancer, UP3. In our reporter luciferase assays (Figure 4.4) UP3 demonstrated ability to enhance transcriptional activity of the SV40 basal promoter. Enhancer modules are typically associated with

histone hyper-acetylation, thus one would expect increased relative acetylation at the UP3 module in H9c2 myoblasts, *in vivo*.

Using primers specifically designed to generate amplicons spanning the PP domain (PP), non-conserved 5-flanking sequence (FLK) and the UP3 domain (Figure 4.8.A), quantitative SYBR green based real time PCR was performed on formaldehyde-crosslinked, sheared chromatin isolated from H9c2 myoblasts (by Dr Andrew Bingham, Department of Cardiovascular Sciences, University of Leicester), which was immunoprecipitated with anti-acetyl H4, anti-dimethyl H3K4 and IgG specific antibodies. For detailed methods please refer to chapter 2. As shown in Figure 4.8.B, the ms1 PP and UP3 domains display significant enrichment of AcH4 with respect to acetylation at the non-specific FLK region (tailed bars, $p < 0.05$). This specificity in acetylation suggests that these domains represent active H9c2 cardiac specific regulatory domains. In this respect it is not surprising that the PP domain, the site of cardiac specific transcriptional initiation, is significantly more acetylated than the distal UP3 putative cardiac enhancer. As a positive control we also quantified AcH4 enrichment at the NPPA proximal promoter and RE-1 site (Figure.4.8.A), both of which are regions that are hyperacetylated in H9c2 myoblasts. Encouragingly we detected relative levels of H4 hyperacetylation at the NPPA PP and RE-1 comparable to those previously published (Bingham et al, 2007). It was also of note that the NPPA and ms1 PP domains were equally acetylated.

In addition to histone acetylation, H3K4 methylation has recently been demonstrated to represent a dynamic determinant of cardiac specific gene expression, specifically modulating the expression of NPPA and BNP (Bingham et al, 2007). Interestingly, our analysis (Figure 4.8.C) indicated that the ms1 PP is highly enriched with 2MeH3K4. This mark is approximately 3-fold more enriched at the ms1 PP compared to both the NPPA PP and RE-1 regions. We observed no significant enrichment at the non-specific FLK or UP3 domains: this lack of UP3 2MeH3K4 is interesting. Indeed the UP3 has significant acetylation with no dimethylation (Figure 4.8.D), with this having importance with respect to the functional role of the UP3 domain (see discussion).

In our reporter luciferase assays (Figure 4.4) UP2 demonstrated ability to repress transcriptional activity of the SV40 basal promoter. Cardiac specific repressive mechanisms are typically associated with HDAC recruitment and subsequent hypo-acetylation. We therefore suspect that UP2 is repressing ms1 transcription through a HDAC dependent mechanism, which is supported by the TSA de-repression data. In addition the UP2 and the PP domains are enriched with MEF2 TFBS. MEF2 has been shown to mediate context specific cardiac specific gene repression via the specific recruitment of Class II HDAC to target regulatory regions. In order to test this possibility we transiently co-transfected the -1585/+60 and -9498-XhoI/-8608-HindIII reporters into H9c2 with the Class II HDACs 4/5. This HDAC reporter co-transfection strategy has been successfully utilised in other studies (Han et al, 2006). HDACs 4 and 5 in combination were able to significantly repress (* $p < 0.05$) the relative luciferase activity of the proximal interval (-1585/+60) and the UP2 domain (-9498-XhoI/-8608-HindIII) (Figure 4.8.E). Somewhat surprisingly however neither the wild type p300 nor the dominant negative p300 acetyltransferase had any significant effect on the relative fragment activity. These findings do however support direct targeting of HDAC repressive complexes, specifically Class II HDACs, at the ms1 regulatory domains *in vitro*. We suspect the observed UP2 mediated cardiac specific repression is likely to be mediated through direct HDAC targeting, de-acetylation activity and subsequent repression. It is also likely that the observed TSA driven de-repression is a consequence of ms1-CRM targeted HDAC inhibition.

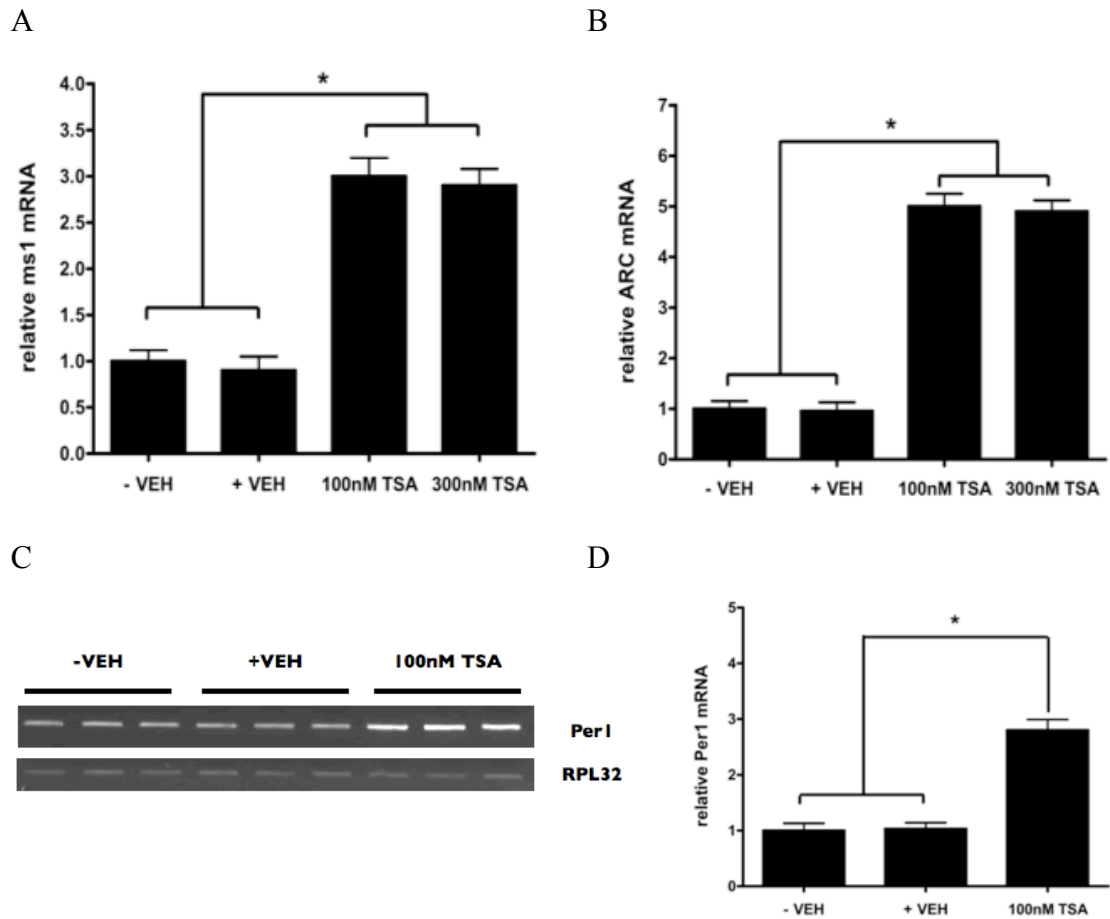
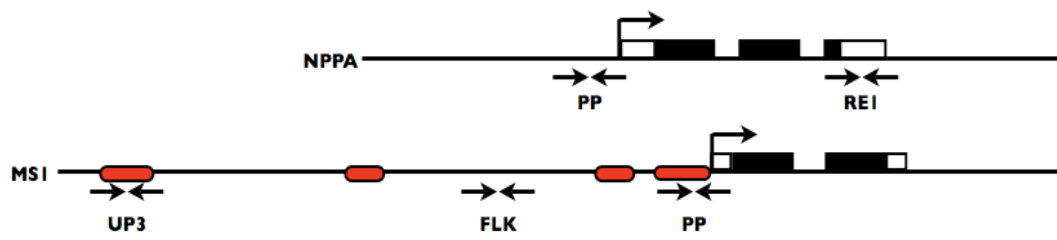
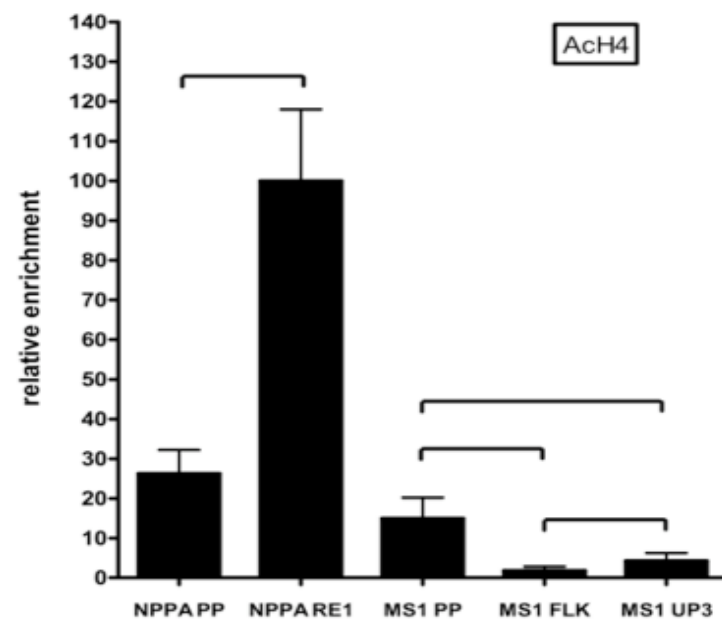


Figure 4.7. Quantitative real-time PCR analysis of ms1 gene expression post Trichostatin A (TSA) treatment *in vitro*. Total RNA was isolated from H9c2 myoblasts treated with 100nM and 300nM TSA and appropriate vehicle control. The RNA was subjected to reverse transcription followed by quantitative PCR with rat ms1 (A), Arc (B) and TBP specific primers. Semi-quantitative PCR was executed using primers specific for Per1 and RPL32 (C, D). Densitometric analysis of agarose gel (C) was quantified and relative expression plotted (D). Expression levels of all transcripts (ms1, Arc, Per1) in untreated cells (-VEH) was arbitrarily set at 1. Values are represented as means \pm SEM of at least three separate experiments. Statistically significant differences are indicated by * $p < 0.05$.

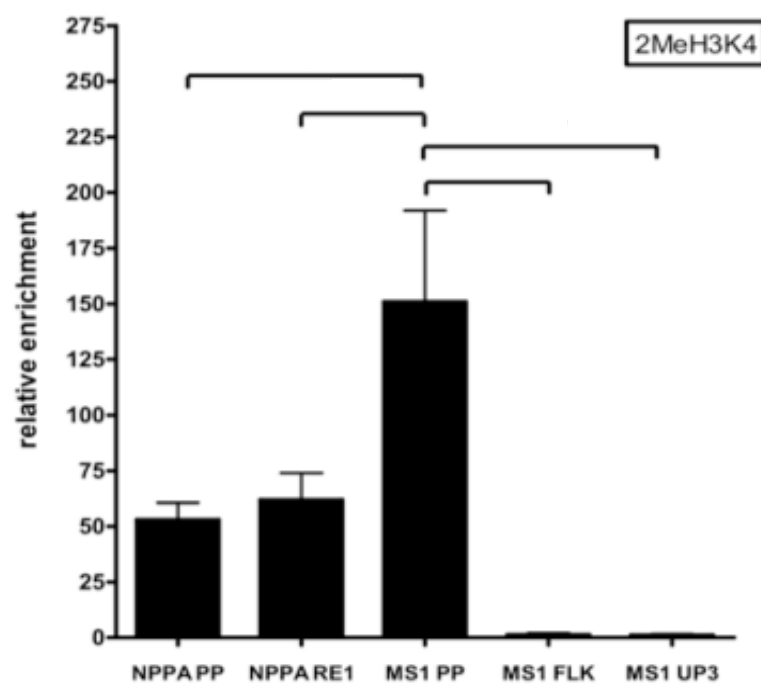
A



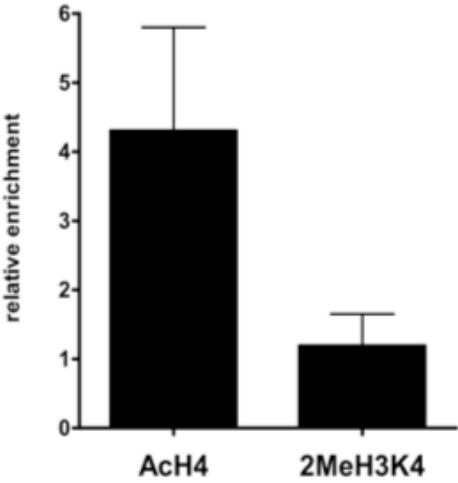
B



C



D



E

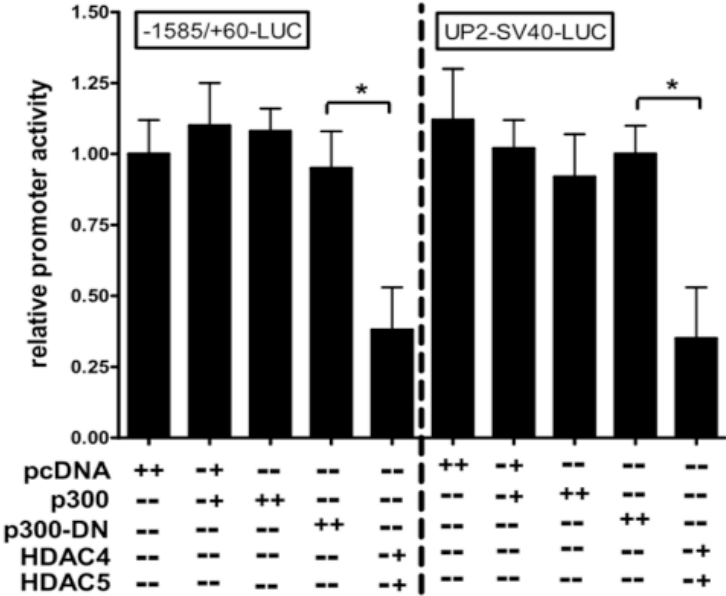


Figure 4.8. Chromatin immunoprecipitation (ChIP) assay performed on H9c2 myoblasts to evaluate *in vivo* histone acetylation and methylation within the ms1 promoter and associated domains. (A) Schematic map of amplified DNA fragments and primer locations encompassing the ms1 and ANP regulatory domains. ms1 PCRM are highlighted by block red circles with the TSS position illustrated by an arrow. Proteins were cross-linked to the DNA (in H9c2 myoblasts) with formaldehyde, DNA was sheared by sonication, and Ab's directed against IgG, AcH4 (B, D) or 2MeH3K4 (C, D) were added to immunoprecipitate protein-DNA complexes. Quantitative real-time PCR was performed on isolated DNA using primers designate in (A). Amplification products were quantified and normalised to the input of each sample. Relative enrichment is displayed as fold enrichment over IgG enrichment at any given domain. (D) Relative enrichment of AcH4 and 2MeH3K4 at the UP3 domain normalised to IgG enrichment. The results are expressed as mean relative enrichment \pm SEM of at least three different ChIPs. Statistically significant differences in relative fold enrichment are indicated by tailed bars, $p < 0.05$. (E) Vectors (+;0.3 μ g, + +0.6 μ g) expressing wild type and dominant negative p300, HDAC4 and HDAC5 were co-transfected in combination with the ms1 promoter reporter construct (P-1585/+60) and UP2 reporter into H9c2 myoblasts. Luciferase activity in cells transfected with pcDNA and reporter was arbitrarily set at 1-fold activation. The results are expressed as mean \pm SEM of at least three different separate transfections in triplicate for each reporter construct. Statistically significant differences are indicated by * $p < 0.05$.

4.2.6 The UP2 distal domain as a potential calcineurin sensitive stress dependent regulatory enhancer.

The data presented so far in this chapter demonstrates that the UP2 PCRM represents a cardiac specific regulatory domain that represses transcription in a basal cardiac state, in addition to not playing an obvious regulatory role during differentiation dependent gene expression. This repressive characteristic also appears to be mediated to some extent by Class II HDACs. This family of HDACs are well known (Chapter 1) to specifically interact with Mef2 family proteins at cardiac CRMs to form Mef2/HDAC repressive complexes. However, under a specific regulatory context, for example post natal adaptation to stress, Mef2 in collaboration with other factors (including the NFATs) is able to toggle between HDAC/HAT association thereby modulating target gene expression.

As highlighted in Chapter 3, our UP2 PCRM is enriched with conserved Mef2 and NFAT motifs (Figure 4.9.A), and we speculated based on this composition that the UP2 domain may represent a calcineurin dependent stress responsive regulatory module. Based on this, and data presented here, we proceeded to test the calcineurin sensitivity of this UP2 module. In order to test this possibility we transiently co-transfected -9498-XhoI/-8608-HindIII reporter into H9c2 and NIH 3T3 fibroblasts with Mef2 and a constitutively active calcineurin protein form (Figure 4.9.B), as described in Chapter 2. Mef2 alone was sufficient to repress the activity of the UP2 domain in a dose dependent fashion only in the H9c2 myoblasts while constitutively active calcineurin transactivated the UP2 domain in dose dependent fashion. Interestingly, at the highest amount transfected plasmid DNA (1.0 μ g) active calcineurin was also able to significantly activate the UP2 domain in the NIH3T3 fibroblasts. In support of the NFAT/Mef2 interaction driving HDAC/HAT toggling, active calcineurin in combination with Mef2 was able to activate UP2 with the negative effect mediated by Mef2 alone completely abolished. In addition, as a positive control to validate the efficacy of the constitutively active calcineurin expression vector used here, the 6XNFAT reporter was also tested for calcineurin sensitivity (Figure 4.9.C). In agreement with published findings, we demonstrate that this reporter is activated by the constitutively active

calcineurin in both H9c2 and NIH 3T3 fibroblasts. However the over-expression of these factors was not sufficient to up-regulate endogenous *msl* expression (Figure 4.9D).

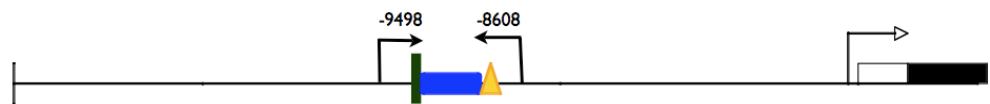
In order to determine whether the endogenous regulatory hardwiring of *msl* (i.e the UP2 domain) is sensitive to calcineurin signalling, *msl* transcript abundance in two models (*in vitro* and *in vivo*) in which the calcineurin pathway is activated was measured. The first model is simulated ischemia-reperfusion, with the calcineurin pathway shown to be activated during reperfusion following sublethal ischemia (Molkentin & Dorn, 2001; Molkentin et al, 1998; Wilkins et al, 2004). *Ms1* expression was measured during a time course of simulated ischemia/reperfusion using a model established by Punnett et al (Punnett et al, 2000). Quantitative real time RT-PCR of *msl* and TBP (internal control) (Figure 4.10.B) using Taq Man® probes was conducted using 1µg of total RNA isolated from H9c2 exposed to simulated ischemia reperfusion as schematised in Figure 4.10.A. A large transient induction in *msl* mRNA ($p < 0.05$) was observed after 1 hour of reperfusion (which followed after one hour of ischemia) with transcript abundance returning to basal levels within two hours. This result suggests that the endogenous regulatory hardwiring is sensitive to stress signals during the reperfusion period, with the calcineurin pathway representing a possible mechanism for action. To confirm the validity of this protocol, quantitative SYBR green based RT-PCR was performed using primers specific to the immediate early genes Jun B and Fra I (positive control genes). Both of these genes have been previously shown to display an acute transient transcriptional profile during reperfusion post ischemia *in vitro*. Encouragingly this profile of expression is also observed in the current *in vitro* model (Figure 4.10.C and -D).

As an *in vivo* model for calcineurin activation we utilised RNA extracted from adult SHR and WKY rat hearts. The SHR rat is a spontaneously hypertensive rat which as an adult displays a robust cardiac hypertrophic phenotype (Clemitson et al, 2007) which is associated with increased calcineurin activity (Molkentin et al, 1998; Wilkins et al,

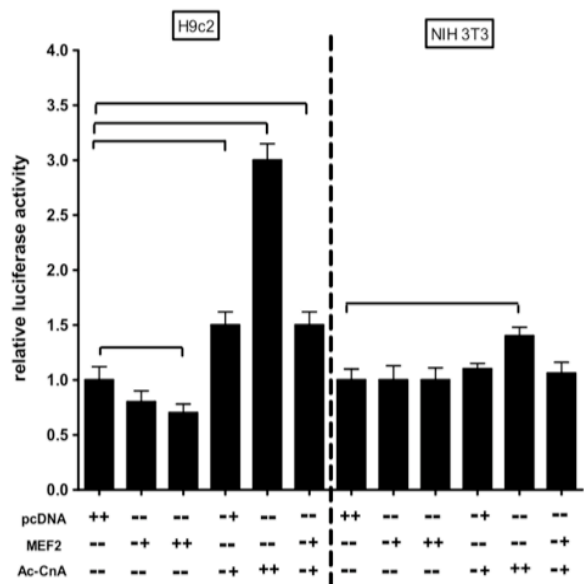
2004). Quantitative RT-PCR indicated a significant increase in *msl* expression in adult SHR heart compared to WKY controls ($p < 0.05$).

In conclusion these data demonstrate that *msl* expression is sensitive to cellular contexts in which the calcineurin signalling pathway is active, and it is proposed that some of this sensitivity is conferred by the distal UP2 calcineurin sensitive regulatory module.

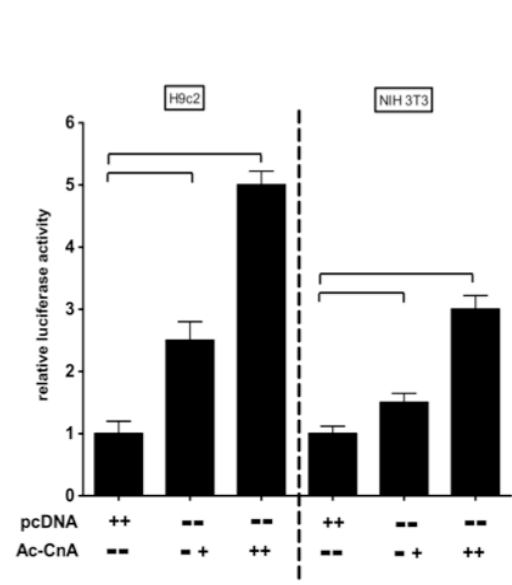
A



B



C



D

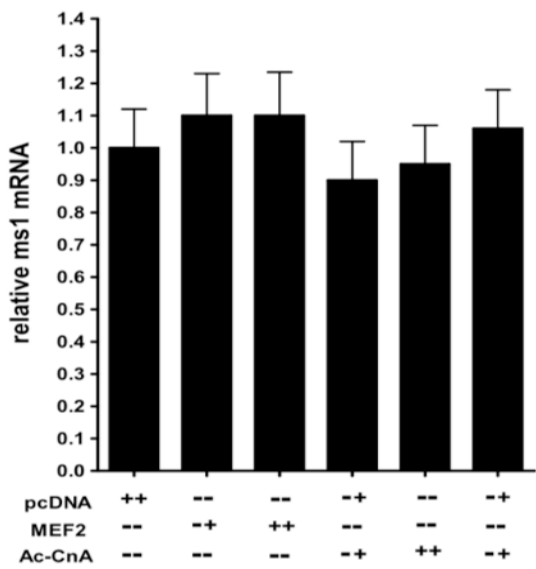
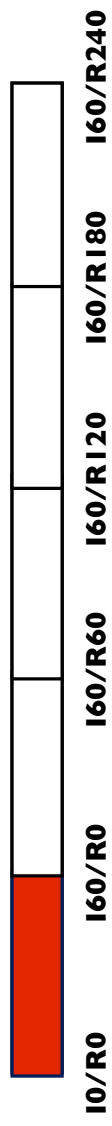
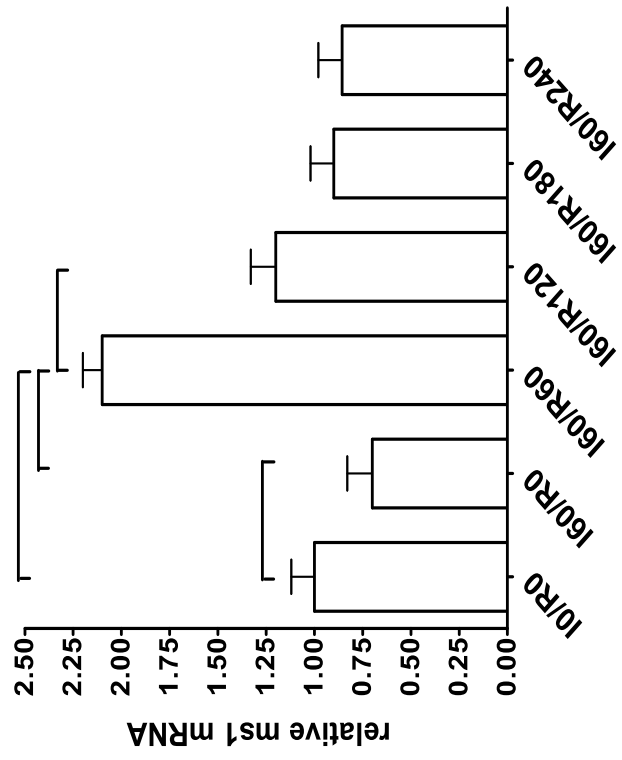


Figure 4.9. The UP2 domain confers calcineurin sensitivity. (A) Schematic representation of the UP2 domain with enriched TFBS as characterised in Chapter 3. Vectors (+;0.3µg, ++0.6µg) expressing constitutively active calcineurin and Mef2 were co-transfected in combination with the ms1 UP2 domain reporter (B) and 6XNFAT (C) reporter into H9c2 myoblasts and NIH 3T3 fibroblasts. Luciferase activity in cells transfected with pcDNA and reporter was arbitrarily set at 1-fold activation. The results are expressed as mean \pm SEM of at least three different separate transfections in triplicate for each reporter construct. Statistically significant differences are indicated by tailed bars, $p < 0.05$. (D) Subconfluent H9c2 myoblasts were transiently transfected with the same combination of expression vectors as before (+;0.5µg, ++1.0µg). 48 hours post transfection total RNA was isolated, reverse transcribed and expression levels of TBP and ms1 were determined by real time PCR.

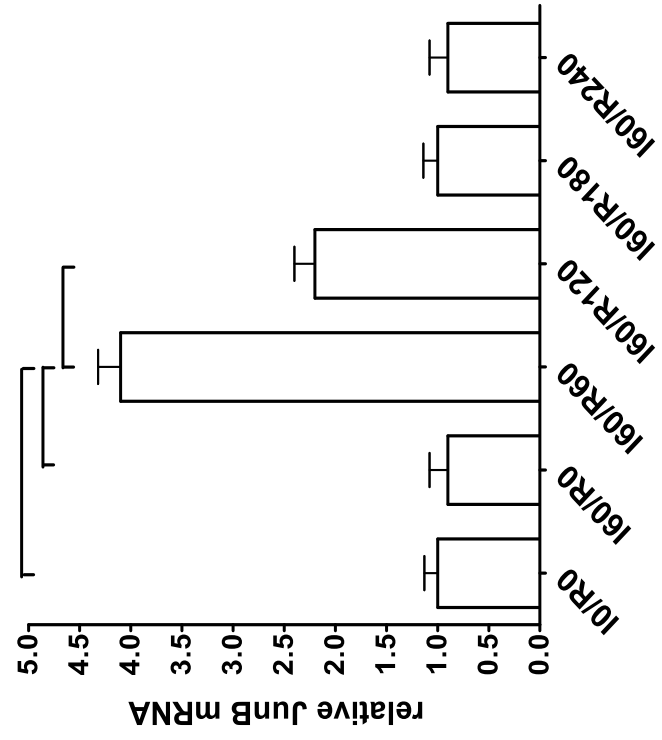
A



B



C



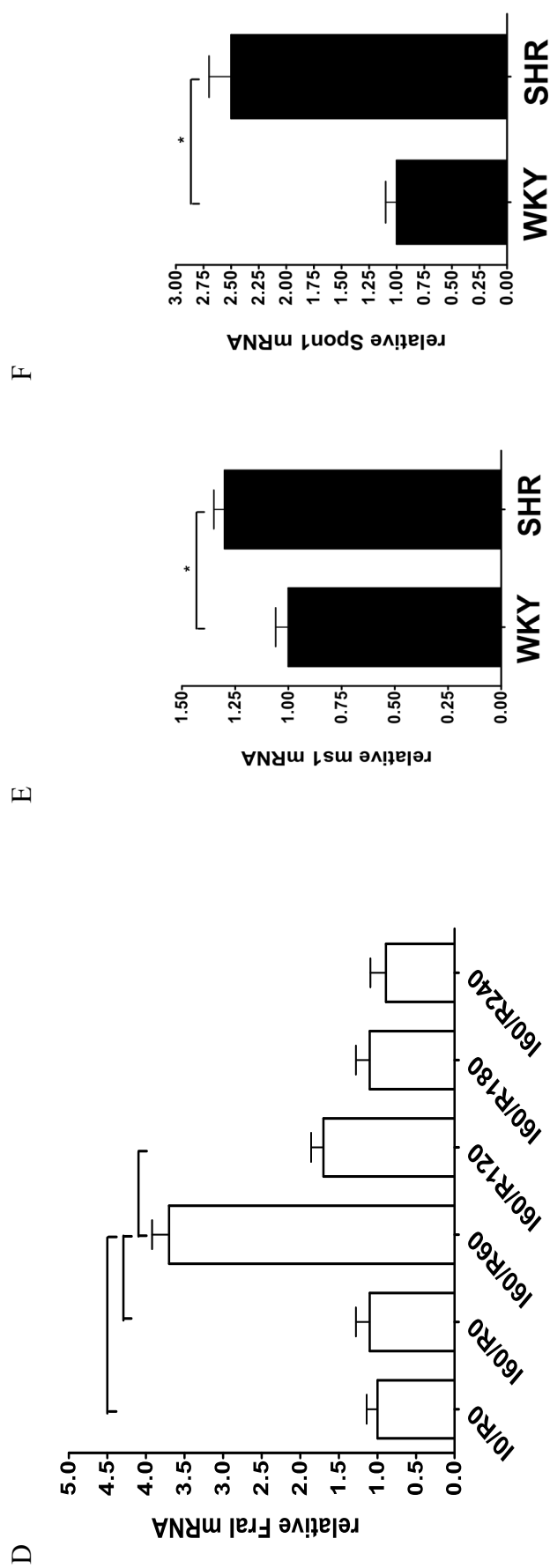


Figure 4.10. Ms1 expression during simulated ischemia and reperfusion in H9c2 cells. H9c2 cells were made ischemic (I) or reperfused (R) for the times indicated in schematic (A). Total RNA was isolated at these time points and subjected to reverse transcription followed by quantitative PCR with rat ms1 (B), JunB (C), Fra1 (D), TBP and RPL32 specific primers. Expression level at I0/R0 was arbitrarily set at 1. Values are represented as means \pm SEM of at least three different experiments. Statistically significant differences are indicated by tailed bars, $p < 0.05$. Total RNA was also isolated from the hearts of adult SHR or WKY rats and subjected to reverse transcription followed by quantitative PCR using primers specific to ms1 and TBP. Relative expression in the WKY heart was arbitrarily set at 1. Values are represented as means \pm SEM from 5 separate heart preparations. Statistically significant differences are indicated by $*p < 0.05$.

4.3 Discussion

Previous work by our group and others (Arai et al, 2002; Kuwahara et al, 2005; Kuwahara et al, 2007; Mahadeva et al, 2002) has clearly demonstrated a unique cardiac restricted temporal-spatial expression profile of *msl* during embryogenesis and post-natal adaptation to stress. Characterising the transcriptional mechanisms governing this will give us an exquisite insight into the regulatory processes active in these cardiac restricted biological processes in addition to illuminating the understanding of the biological role of the MS1 protein.

Our comparative *in silico* analysis (Chapter 3) identified an array of putative CRMs (PCRM) which we suggested modulate *msl* cardiac specific expression, which were functionally examined here. As the first step in this analysis, we quantified relative *msl* expression in the major organs and tissues of the adult rat. In agreement with previously published studies (Arai et al, 2002; Kuwahara et al, 2005; Kuwahara et al, 2007; Mahadeva et al, 2002), we demonstrate using quantitative RT-PCR (Figure 4.1) that *msl* expression is restricted to striated muscle. It was of interest that *msl* transcript, and presumably protein, is more abundant (20-fold) in skeletal versus cardiac muscle. We suspect that this difference in expression relates to the respective demands on SRF signalling coupled to the relative availability of the myocardin/MRTF SRF co-factors in both muscle types (Cen et al, 2004; Chen et al, 2002; Du et al, 2003; Li et al, 2005; Miano, 2003; Selvaraj & Prywes, 2003). The founding member of the MRTF co-factor family, Myocardin, is specifically expressed within cardiac and smooth muscle. Contrary to the MRTFs, Myocardin is constitutively nuclear where it can associate with SRF and transactivate target genes. However, the MRTFs (MRTF-A/-B), which are expressed in skeletal and cardiac muscle, translocate and activate SRF target genes in collaboration with RhoA signalling and *msl* expression. We therefore suspect that increased skeletal expression of *msl* serves to support robust MRTF-SRF dependent signalling, thereby compensating for the lack of Myocardin driven SRF transactivation which occurs only within cardiac and smooth muscle. Of note, recent studies have also demonstrated that Myocardin and the MRTFs (-A/-B) are able to transactivate specific subsets of SRF target genes (Selvaraj & Prywes, 2004). We therefore can not rule out

that high skeletal *msl* expression may drive MRTF dependent transactivation of a subset of SRF target genes, specific to the skeletal muscle lineage.

Quantitative expression analysis of *msl* during embryonic and post-natal cardiac development demonstrated a transcriptional induction occurring primarily during the neonatal to adult period of cardiac maturation (Figure 4.1.C). This phase encompasses a major phenotypic transition, with proliferating neonatal myocytes exiting the cell cycle to terminally differentiate and mature into binucleated adult cardiomyocytes (Olson, 2004). Maturation is accompanied by a down-regulation of fetal gene expression and sarcomeric reorganisation and myofibrillar assembly leading to the development of a mature adult competent contractile apparatus. Numerous studies demonstrate that SRF regulates a distinct programme of gene expression that has direct relevance for myofibrillar assembly and maintenance in addition to physiological control of sarcomeric function (Charvet et al, 2006; Li et al, 2005; Selvaraj & Prywes, 2004; Vandromme et al, 1992; Wei et al, 1998). One would hence expect biological mechanisms exist that couple SRF activity with myocyte maturation. We propose that the observed maturation specific up-regulation of *msl* is responsible for driving SRF transactivation, through the RhoA-MRTF pathway, during this maturation developmental phase.

This maturation specific transcriptional up-regulation also has implications for the regulatory mechanisms governing *msl* expression. Firstly, such a temporal differentiation dependent expression profile is typically mediated by the complex interplay of numerous CRMs and cognate binding factors. It is of interest that predicted TFBS identified within our putative cardiac CRMs include Mef2, GATA4 and SRF, all TFs with described regulatory roles in cardiac differentiation and maturation specific gene expression. We suspect these factors modulate some facet of cardiac maturation specific *msl* expression. Using serial deletion of promoter luciferase reporters, we observe clear cardiac specific activity of the proximal 1.5kbp 5'-flanking interval. It was of interest that our truncated reporters indicated that this cardiac specificity, at least in the H9c2 myoblasts, was mediated through motifs residing within the proximal 300bp, representing the highly conserved proximal promoter domain (Chapter 3). We

also showed that in this *in vitro* context, the UP1 domain is dispensable for reporter activity, although it does not exclude a role for UP1 in cardiac specific expression. It may play a more subtle regulatory role during cardiac development or post-natal homeostasis which was detectable in the present *in vitro* H9c2 model. To this end it is becoming increasingly clear that regulatory elements important for transcription *in vivo*, are typically not readily identified *in vitro* (Ureta-Vidal et al, 2003). In summary we suspect, as hypothesised in chapter 3, ultra conserved cardiogenic motifs within the proximal promoter, including MEF2 and GATA sites, will contribute to cardiac specific regulation of *msl*. In support of this, truncation within the proximal promoter domain (-127/+60 reporter) which ablates highly conserved GATA and MEF2 motifs, resulted in attenuated cardiac specific activity. Therefore motifs within the truncated interval are essential for *in vitro* cardiac specificity, supporting our *in silico* derived hypothesis.

During the course of this thesis Eric Olson's group demonstrated, both *in vivo* using LacZ promoter reporter transgenic mice and *in vitro* utilising luciferase promoter reporter driven assays in NRVM, that the proximal 200bp interval upstream of the STARS TSS (corresponding to our rat PP domain) mediates basal and stress-inducible cardiac specific expression (Kuwahara et al, 2007). This was true both in the embryonic and adult transgenic mouse heart as well as in isolated NRVM. In addition these authors demonstrated a key regulatory role for Mef2C binding at the -135/-124 highly conserved Mef2 site, identified in our *in silico* comparative analysis. These authors suggested that the majority of basal and stress-inducible cardiac specific expression is mediated, in part, through this motif. Interestingly they show that upon truncation of the STARS proximal promoter and thereby ablating this Mef2 motif, they lose approximately 50% of basal cardiac specific activity, both *in vivo* and *in vitro*. This finding complements the current data which demonstrates that, within H9c2 myoblasts, truncation within the ultra conserved proximal domain leads to attenuation of cardiac specific activity, at comparable levels to those observed *in vivo* and *in vitro* by Kuwahara and colleagues (2007). With these findings in mind we suggest our attenuated reporter activity (-127/+60 reporter) may be in part mediated through truncation of the ultra conserved -135/-124 Mef2 motif. It is extremely encouraging that our *in silico* driven *in vitro* assays correlate so well with Olson's *in vivo* and primary myocyte

(NRVM) *in vitro* findings, which suggest the validity for the use of H9c2 myoblasts as a model for cardiac regulatory characterisation (at least with respect to *msl*).

It was of interest that Kuwahara and colleagues (2007) also speculated, like us (Chapter 3), that the proximal ultra-conserved TATA like sequence (TATT_{-30/-25}) represented the bona-fide TATA box and thus core promoter. We have confirmed this suspicion and demonstrated using EMSA and site directed mutagenesis that this proximal sequence does represent the bona-fide TATA box, and hence site of TBP binding and subsequent transcriptional initiation.

In addition to the proximal interval, our *in silico* analysis (Chapter 3) identified two distal CRMs (UP2 and UP3) which we speculated contributed to the unique cardiac restricted temporal-spatial *msl* expression profile. Using the SV40-driven enhancer/repressor luciferase reporter system we demonstrated that both of these distal CRMs modulate (in a cardiac specific manner) enhancing and repressive activity. For example, the UP2 domain efficiently repressed SV40 basal promoter activity. This suggests that UP2 (in this *in vitro* context) represents a cardiac specific repressor. Based on the TFBS composition of the UP2 domain, we speculated (Chapter 3) that this CRM may represent a calcineurin-NFAT/Mef2 dependent stress sensing domain, which through HAT/HDAC toggling may modulate the activity of the *msl* proximal promoter. If true, one would expect this domain to act in a HDAC dependent repressive manner in an unstimulated cardiac context. Our findings here support this hypothesis and raise the intriguing possibility that the transient expression profile observed for *msl* post pressure overload *in vivo* (Mahadeva et al, 2002) may be as a consequence of calcineurin activity and signalling which, via the interaction of NFAT and Mef2 interaction, integrates at the UP2 domain.

Contrary to the UP2 domain, UP3 enhanced the activity of the SV40 basal promoter, which led us to suspect it represents a cardiac specific enhancer. This again correlated with predictions made in chapter 3, in which we speculated this domain may represent a cardiac developmental enhancer. We suggested this enhancer may contribute to early cardiac specific activation in the embryonic linear heart tube, specifically around

E7.5-8. If true, you would expect this enhancer to be active in H9c2 myoblasts, as these are derived from the embryonic ventricle (represents a developmental stage post linear heart tube formation, see Chapter 1). The current findings reported here do support this notion.

Although encouraging, some functionally relevant limitations need to be appreciated when interpreting the present findings. Firstly, these results represent the regulatory effect of the UP2 and UP3 domains on the SV40 promoter. Although bona-fide enhancers and repressors can by definition act non-specifically on basal promoters, one can not rule out the possibility that these modules demonstrate different regulatory activity when acting on the *msl* proximal promoter within the endogenous genomic landscape. Based on our results in this chapter demonstrating that the PP domain is critical for cardiac specific expression, it would be more intuitive to clone the UP2 and UP3 domains upstream of the PP domain and characterise their regulatory activity in this context. In addition, although we believe the H9c2 cells used represent an adequate cardiogenic *in vitro* model, one can not rule out the possibility that in the *in vivo* context, both the UP2 and UP3 have different regulatory activity. We do believe that even when considering these limitations, the current data do give us a cursory insight into the cardiac specific regulatory roles of these distinct CRMs. It was encouraging that not only did these modules mediate differential cardiac specific activity, but they also had activity of which we predicted through our *in silico* analysis (Chapter 3). These findings also question the validity of the conclusions derived by Olson's group because they focused their regulatory analysis on the proximal 1.5kbp 5-flanking interval only (Kuwahara et al, 2007). We suspect, based on this data and findings by others (Kuwahara et al, 2007), that these distal CRMs may contribute to context specific cardiac regulatory activity central to correct *msl* temporal-spatial expression, such as developmental or maturation specific transcriptional induction.

To gain an insight into the roles of these functional cardiac CRMs during maturation specific *msl* up-regulation, RA induced cardiogenic differentiation of H9c2 myoblasts was utilised as an appropriate model system. RA induced differentiation, which is associated with endogenous *msl* up-regulation, specifically increased the activity of the

PP domain reporter which encompasses the highly conserved proximal 300bp upstream of *msl* TSS. No differentiation specific effects were observed with respect to the UP1, UP2 or UP3 domains. We therefore conclude that, at least in our H9c2 model, differentiation dependent transcriptional induction is primarily mediated through the PP domain, and thus TF binding within this domain. This does support our *in silico* derived hypothesis. We suspect that this domain, and associated regulatory mechanisms and pathways converging on this interval, may also play a role in modulating the *in vivo* maturation specific increase in *msl* expression. Characterising the pathways and terminal factors acting within this domain is therefore a priority for subsequent analysis. It is important to emphasise that these findings alone can not exclude an *in vivo* regulatory role for all three distal CRMs during lineage determination, differentiation and maturation cardiac specific gene expression.

We suspected that *msl*, with its complex CRM structure and enrichment of TFBS with known histone modifying binding factors, would be under epigenetic regulatory control. Class I and II HDAC inhibition (using TSA) resulted in a robust de-repression of *msl* expression, which supports an epigenetic regulatory hypothesis. We suspect this de-repression is a product of direct CRM bound HDAC repression, thereby resulting in local hyperacetylation with subsequent relaxed chromatin and transcriptional induction. However one can not rule out the possibility of this effect being secondary to non-histone dependent acetylation. Interestingly, GATA4, which we suspect may bind the *msl* CRMs at conserved GATA motifs (chapter 3), has been shown to undergo HAT/HDAC dependent acetylation and deacetylation, with acetylation increasing DNA binding activity and therefore transactivation potential (Kawamura et al, 2005). It is conceivable that TSA dependent *msl* de-repression could be caused by increased binding of hyperacetylated transcription factors, for example GATA4, thereby resulting in increased transcription at the *msl* promoter.

To further delineate these possibilities we utilised ChIP and demonstrated specific enrichment of acetylated histone H4 at the *msl* proximal promoter. This would support a potential role of dynamic histone acetylation at this *loci*. It was of interest that we also observed specific histone acetylation at the UP3 domain and not at the non-specific

FLK region. In addition, contrary to the proximal promoter, UP3 was weakly enriched for 2MeH3K4. This was a functionally relevant observation in light of recent work by Heintzman and colleagues (2007). This study analysed chromatin modification states in high resolution along 30 Mb of the human genome corresponding to the ENCODE consortium intervals (Heintzman et al, 2007). They found that active promoters are marked by hyperacetylation and specific trimethylation and dimethylation of Lys4 of histone H3 (H3K4), whereas enhancers are marked by histone acetylation and monomethylation, but not tri/dimethylation, of H3K4. Results from our *in vitro* reporter assay and *in vivo* ChIP assay leads us to propose that the UP3 CRM represents a bona-fide cardiac specific enhancer module.

Further evidence that validates a role for HDAC specific regulatory activity, we also demonstrated specific Class II HDAC sensitivity at the PP domain proximal promoter and distal UP2 repressor domains. This suggests Class II HDAC enzymes are directly targeted to these domains. This is not surprising considering the presence of Mef2 binding sites within both the PP and UP2 domains, with MEF2C binding within the PP already been demonstrated (Kuwahara et al, 2007). Prior studies have convincingly demonstrated that MEF2 family proteins selectively associate with Class II HDACs forming a complex on gene regulatory elements resulting in gene repression (Zhang et al, 2002). In response to myocardial stress, for example pressure overload, these HDACs are shuttled from the nucleus to the cytoplasm, which provides a post translational mechanism to override HDAC-mediated repression of gene expression. This redistribution of HDAC frees MEF2 (and other associating transcription factors) to associate with HATs (Zhang et al, 2002), resulting in increased local histone acetylation and activation of downstream genes. It will therefore be of interest to further investigate (via ChIP) if our demonstrated PP/UP2 HDAC sensitivity is mediated via MEF2 factors, and how these MEF2/HDAC complexes contribute to stress dependent *msl* expression, in particular during pressure overload induced cardiac hypertrophy.

In summary, these *in vivo* and *in vitro* findings validate bona-fide cardiac regulatory function for our *in silico* predicted CRMs, thereby validating the utility and specificity of our comparative strategy (Chapter 3). The highly conserved proximal promoter

domain was proposed to be critical for cardiac specific expression, and in our model, cardiogenic differentiation dependent expression. The present findings also suggest that the distal CRMs, UP2 and UP3, represent cardiac specific regulatory modules. The exact regulatory nature of these modules can not be definitively described from these data alone however; we suspect they contribute to cardiac context specific modulation of *msl* expression, presumably during cardiac development and post-natal adaptation to stress. These findings exemplify the complex regulatory hardwiring involved that modulates *msl* expression. Understanding the factors binding to these CRMs will give us a greater understanding into their role during cardiac context specific expression of *msl*. The current analysis has therefore generated specific set of functional regulatory domains for TFBS analysis in subsequent analysis as well as providing an initial illumination into the complexity of *msl* regulatory hardwiring.

Chapter 5

GATA4 modulates cardiac specific expression of *msl* through distinctive mechanisms at the proximal promoter and distal UP3 enhancer domain.

5.1 Introduction

In Chapter 4, a functional characterisation of our *in silico* derived putative *cis* regulatory modules (PCRM) was carried out, which demonstrated these PCRM modulated cardiac specific expression of *msl*. In summary, the PP domain, which encompassed the core promoter, was shown to be essential for basal and differentiation dependent cardiac specific *msl* expression. Evidence was also provided to suggest that cardiac expression was enhanced through the UP3 CRM, which was concluded to represent a bona-fide cardiac specific enhancer. However, to fully delineate the cardiac mechanisms governing context specific *msl* expression, it is imperative to identify the transcription factor binding sites (TFBS) and cognate factors (acting within these characterised CRM) that mediates the associated cardiogenic activity.

The present bioinformatic comparative analysis identified an array of cardiac specific TFBS within the PP sequence and the distal UP3 domain (Chapter 3). The most enriched binding site within these two domains was the GATA motif, which *in vivo* is bound by the GATA family of transcription factor proteins. GATA4, the major GATA binding factor within the embryonic and adult heart is a key transcriptional modulator of numerous cardiac specific genes such as α -myosin heavy chain (α -MHC), b-type natriuretic peptide (BNP) and atrial natriuretic factor (ANF) (Molkentin, 2000; Pikkarainen et al, 2004). A direct role for GATA4 in transcriptional regulation was supported by the observation that antisense GATA4 mRNA expression repressed the basal expression of many cardiac-specific genes in cardiomyocyte cultures (Charron & Nemer, 1999; Charron et al, 1999).

The requirement of GATA4 in regulating cardiac development and myocyte differentiation has been demonstrated through the use of genetically modified mice.

Using a tetraploid rescue strategy, GATA4 $-/-$ embryos exhibited a hypoplastic ventricular myocardium with disrupted looping morphogenesis (Watt et al, 2004). In a similar fashion, conditional disruption of GATA4-loxP-targeted alleles in the heart using *Cre-loxP* approach coupled to a *Nkx2.5-Cre* “knock-in” allele resulted in embryonic lethality (Oka et al, 2006; Pu et al, 2004). This embryonic lethal phenotype was associated with hypoplastic ventricles and generally disrupted cardiac morphogenesis. In light of these transgenic findings it is not surprising that mutations in GATA4 are also associated with congenital heart defects in humans (Garg et al, 2003; Rajagopal et al, 2007). Collectively, these results demonstrate and emphasise an important role for GATA4 as a critical regulator of heart development, a context in which *msl* is differentially regulated.

In addition to its role in maintaining differentiated gene expression in the adult and developing heart, GATA4 also modulates inducible gene expression in the post-natal heart in response to hypertrophic stimuli, specifically endothelin-1, isoproterenol, phenylephrine and pressure overload (Hasegawa et al, 1997; Herzig et al, 1997; Liang et al, 2001b; Morimoto et al, 2000). Over-expressing GATA4 in transgenic mice and cultured myocytes is sufficient to induce cardiac hypertrophy. Conversely, expression of anti-sense GATA4 mRNA or dominant negative GATA4 protein blocked phenylephrine and endothelin-1 induced cardiac hypertrophy in culture (Liang et al, 2001b; Morimoto et al, 2000). A number of these stimuli that induce cardiac hypertrophy and/or heart failure were shown to enhance GATA4 transcriptional activity through post-translation modification, specifically phosphorylation. Numerous studies have confirmed this mechanism using different stimuli such as phorbol esters, angiotensin II, endothelin-1, phenylephrine, isoproterenol and pressure overload (Hasegawa et al, 1997; Hautala et al, 2001; Herzig et al, 1997; Liang et al, 2001b).

On the basis of the context specific regulatory activity of GATA4 described above, coupled to the enrichment of multiple GATA motifs within our cardiac CRMs, we suspect GATA4 may regulate *msl* expression in cardiac specific regulatory contexts (cardiac development and post natal adaptation). In this chapter we have therefore investigated the importance of enriched GATA motifs, and specifically the GATA4

protein, in regulating the cardiac specific activity of the PP and UP3 domains, and hence *msl* gene expression. Given the important functions that GATA4 and SRF have in cardiac development and post-natal function, potential regulatory cross-talk at the level of *msl* gene expression will have fundamental cardiac gene regulatory implications.

5.2 Results

5.2.1 *GATA4 targets the *msl* proximal promoter domain.*

The proximal *msl* 5-flanking interval is highly enriched with GATA motifs, with one in the UP1 domain and six within the proximal promoter (PP) (Figure 5.1.A). The P-1585/+60, -365/+60 and -127/+60 *msl* promoter reporter constructs, spanning the UP1 and PP domains respectively (Figure 5.1.A), were transiently transfected into H9c2 and NIH 3T3 cells in combination with wild type and dominant-negative GATA4 overexpression plasmids. These overexpression plasmids have been demonstrated to drive robust ectopic expression of full length wild type GATA4 protein (Wang et al, 2005) and a dominant negative GATA4 protein form. Both of these plasmids have been used in a number of studies (Majalahti et al, 2007; Wang et al, 2005), with the wild type GATA4 able to transactivate the ANP promoter in a GATA dependent manner. The dominant negative GATA4 (DN-GATA4), which has been shown to inhibit basal and stress inducible GATA4 dependent transcription (Bhalla et al, 2001; Charron et al, 2001) is composed of the DNA-binding domain only. It therefore functions as a dominant negative protein acting as a competitor with endogenous GATA4 for cognate binding sites.

DN-GATA4 robustly repressed the activity of all three promoter reporters similarly in a dose-dependent fashion in H9c2 and NIH 3T3 cells (Figure 5.1.B.C.D), demonstrating that the effect is specific to the proximal 127bp upstream flanking sequence. In H9c2 cells, the -127/+60 reporter was repressed by 50% (0.5µg expression vector, $p<0.05$) and 70% (1.0µg expression vector, $p<0.05$), compared to the control (1.0µg pcDNA only)(Figure 5.1.D). This repression strongly supports the notion that endogenous GATA4 contributes significantly towards cardiac specific expression, presumably through the proximal GATA motifs. However, we also showed that DN-GATA4 represses reporter activity in NIH 3T3 cells. This is surprising considering the non-cardiac nature of these cells. This suggests that DN-GATA4 dependent repression

maybe be mediated through the intrinsic DNA binding capability of this truncated protein. If the mechanism was purely based on competition for binding sites with endogenous GATA4, the repressive effect of DN-GATA4 would be absent in NIH 3T3 (low endogenous GATA4 levels), or at least smaller compared to the effect in H9c2.

In support of this notion, we found that full length wild type GATA4 protein also repressed the activity of all three reporters in a dose-varying fashion (Figure 5.1.B.C.D). In H9c2 cells, the -127/+60 reporter was repressed by 30% (0.5µg expression vector, $p<0.05$) and 50% (1.0µg expression vector, $p<0.05$), compared to the control (Figure 5.1.D) and yet again, repression was not cell specific. This suggests the mechanism of repression is a ubiquitous one, and likely to be a DNA binding effect. The effect of WT-GATA4 was somewhat surprising considering it is typically associated with transcriptional activation, however the level of repression demonstrated with WT-GATA4 was not as dramatic as that of DN-GATA4.

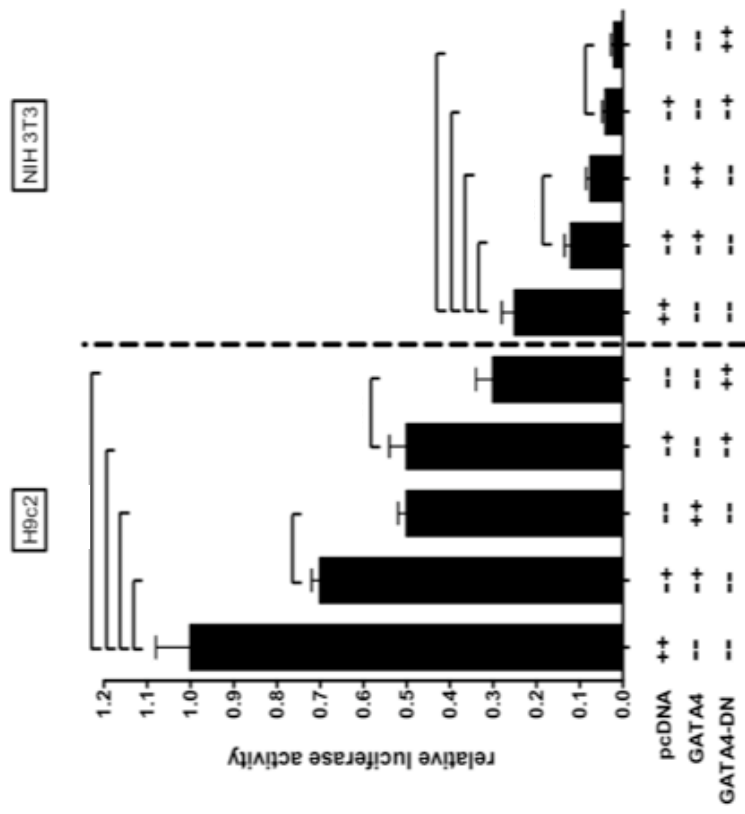
The lack of non-specific activation of the empty luciferase vector, pGL3-B (Figure 5.1.F) suggests that the observed effects described here are specific to the *msl*-promoter reporters. As a positive control to validate the efficacy of the GATA4 expression vectors used, the ANP-luciferase reporter was also tested for GATA4 sensitivity. In agreement with published findings, we demonstrated this reporter is activated by WT-GATA4, with concomitant dose-dependent repression by DN-GATA4 (Figure 5.1.E).

The data presented so far demonstrates that GATA4 represses *msl* promoter activity and that this DNA binding-specific sensitivity is mediated by motifs within the proximal 127bp interval, demonstrated by the exquisite sensitivity of this reporter to both wild type and dominant negative GATA4 expression.

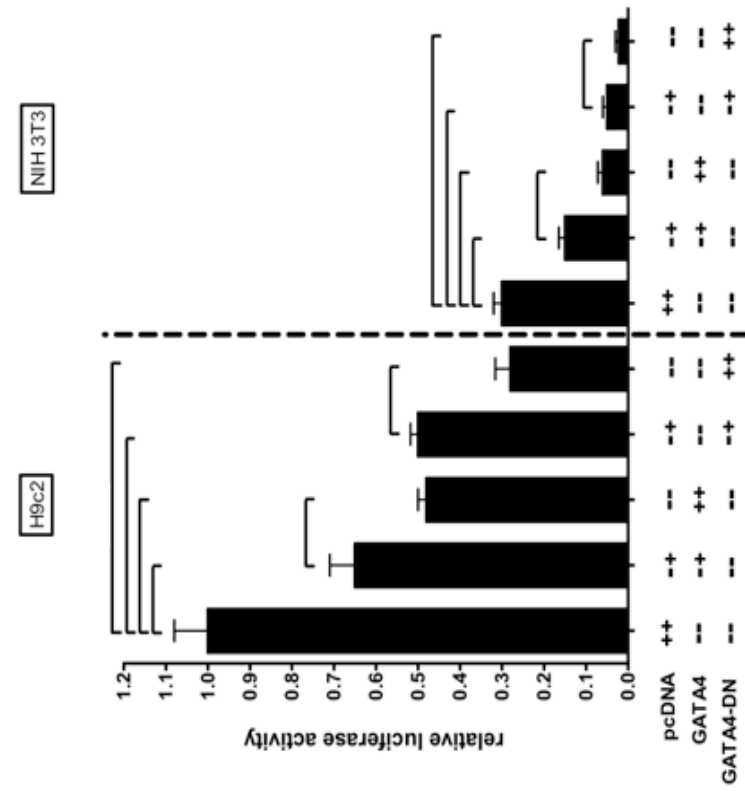
A



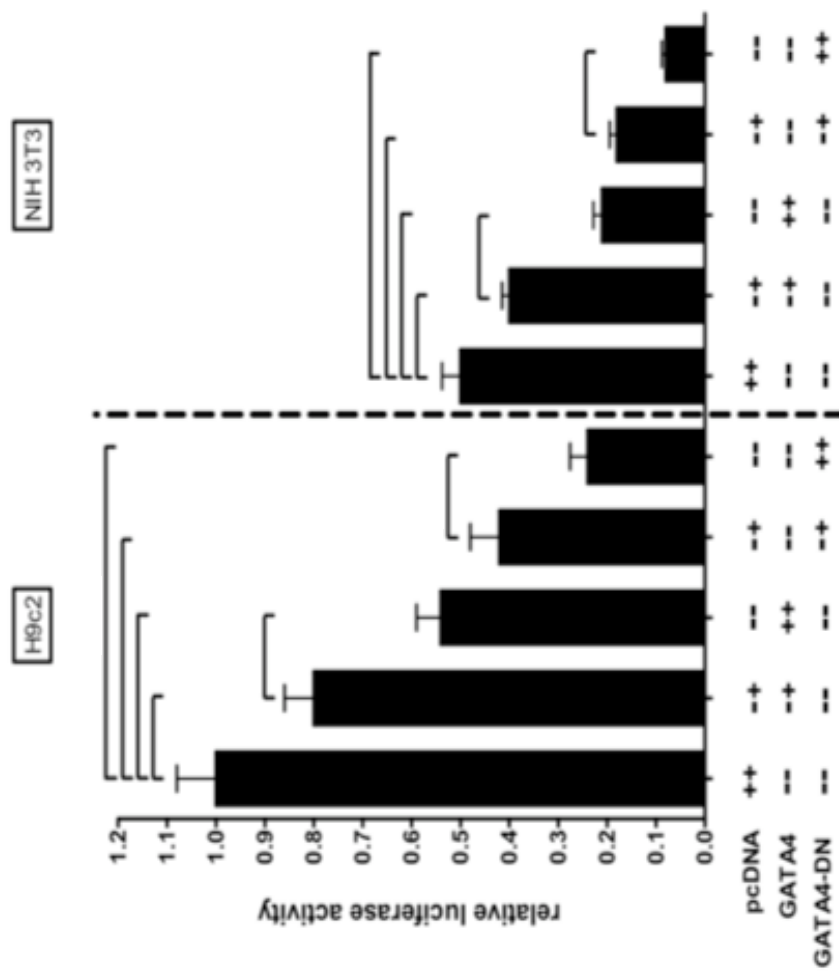
B



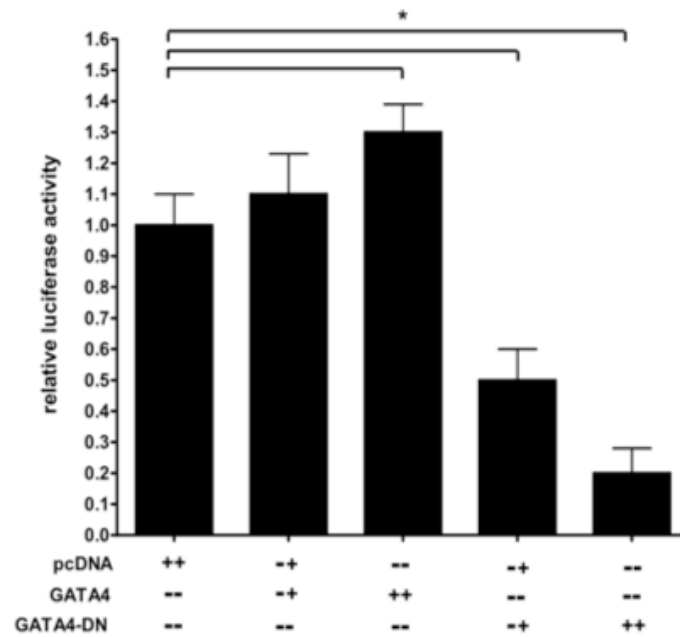
C



D



E



F

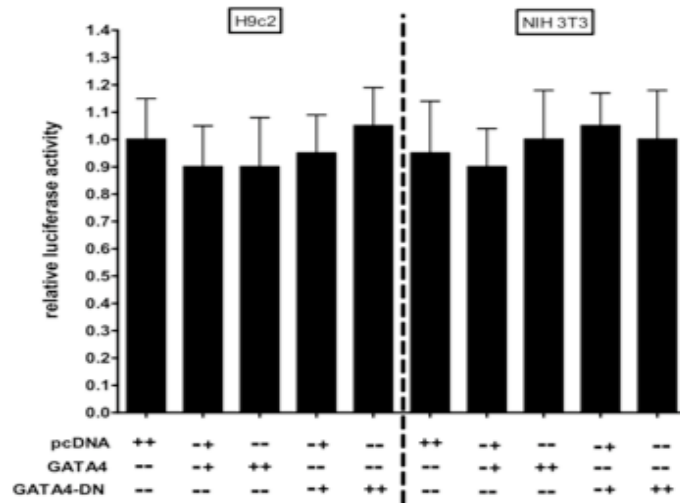


Figure 5.1. The proximal promoter confers GATA4 sensitivity. (A) Schematic representation of the proximal interval domain with annotated luciferase construct primer sequences and enriched GATA motifs represented by solid red boxes. Vectors (+; 0.03 μ g, ++;0.6 μ g) expressing wild type and dominant negative (DN) GATA4 were co-transfected with the -1585/+60 (B), -365/+60 (C), -127/+60 (D), ANP-Luc (E) and pGL3-B (F) reporter vectors into H9c2 myoblasts and NIH 3T3 fibroblasts. The relative activity of the reporter (vs pGL3-B) in H9c2 transfected with pcDNA only was arbitrarily set at 1-fold activation. The results are expressed as mean \pm SEM of at least three different separate transfections in triplicate for each reporter construct. Statistically significant differences are indicated by * and talied bars, $p < 0.05$.

5.2.2 *In vivo binding of GATA4 at the proximal -127/+60 interval*

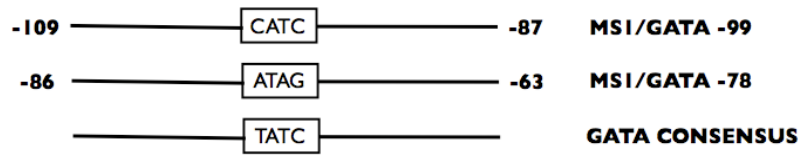
In order to demonstrate the GATA4 binding specificity of the two ultra-conserved GATA motifs within the -127/+60 interval, electro-mobility shift assays (EMSA) were carried out. Briefly, we synthesised specific oligonucleotides containing the GATA elements present in the proximal -127/+60 interval, ms1/GATA-99 and ms1/GATA-78 (Figure 5.2.A). These oligonucleotides were then annealed and end labelled with digoxigen (DIG) generating probes for EMSA. These ms1/GATA probes were then incubated with whole cell protein extracts from H9c2 myoblasts and a molar excess (X200) of complimentary cold unlabelled probe or cold unlabelled GATA4 binding consensus GATA motif. This positive control consensus oligonucleotide is derived from the carnitine palmitoyltransferase 1b promoter and has been shown to bind H9c2 derived GATA4 protein *in vitro* (Krejci et al, 2004).

As shown in Figure 5.2.B and -C, incubation of our MS1/GATA-99 and MS1/GATA-78 labelled probes with H9c2 whole cell protein extracts resulted in the appearance of a single DNA-protein band shift (represented by arrow). However, co-incubating this reaction with a 200-fold molar excess of unlabelled complimentary probe, or GATA consensus probe, did not ablate the appearance of the shifted band, demonstrating this DNA-protein band shift is the product of a non-specific protein/DNA-sequence interaction, rather than a specific interaction. Non specific interactions can be inhibited through increasing the concentration of non-specific inhibitor (poly d(I-C)) in the incubating reaction. We therefore increased the concentration of poly d(I-C) used in our incubations using 1µg and 2µg, respectively (Figure 5.2.B.C, right hand gel images). However, this was not sufficient to inhibit non-specific complex formation on both of our labelled probes.

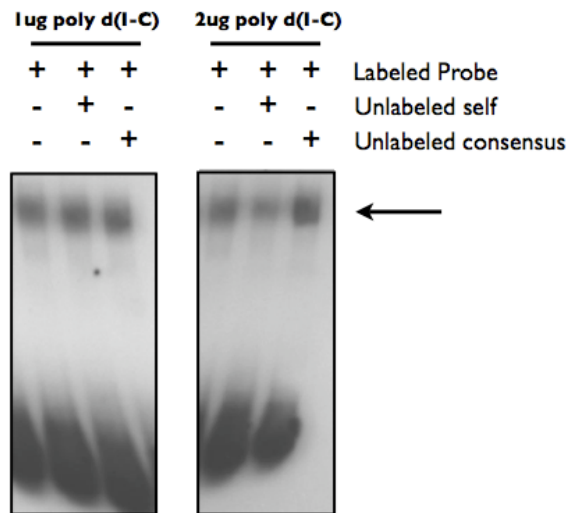
To definitively confirm GATA4 binding at these proximal GATA motifs, the chromatin immunoprecipitation (ChIP) assay was utilised (in collaboration with Prof Donald Menick, Medical University of South Carolina). Using primers specifically designed to generate amplicons spanning the proximal (-127/+60) and distal portion (-300/-127) (Figure 5.3.A) of the PP domain, semi-quantitative PCR was performed on

formaldehyde-crosslinked, sheared chromatin isolated from feline adult cardiomyocytes (isolated by Robert Peterson, MUSC), which was immunoprecipitated with anti-GATA4 antibody (Figure 5.3.C.D). GATA4 antibody specifically pulled down DNA fragments containing the proximal portion of the PP domain corresponding to -127/+60 interval, which contains the -99 and -78 GATA motifs. As a positive control, 5% of the input chromatin (IP) was used for PCR. The specificity of the ChIP was controlled by running negative controls lacking the precipitated antibody (No Ab). To confirm the validity of this pull-down we also demonstrated GATA4 enrichment at the NCX1 GATA motif, a bona-fide GATA4 binding domain (Xu et al, 2006). The ChIP data suggest that GATA4 directly interacts with the proximal promoter *in vivo*. We suspect this is through binding to the proximal -99 and -78 GATA motifs.

A



B



C

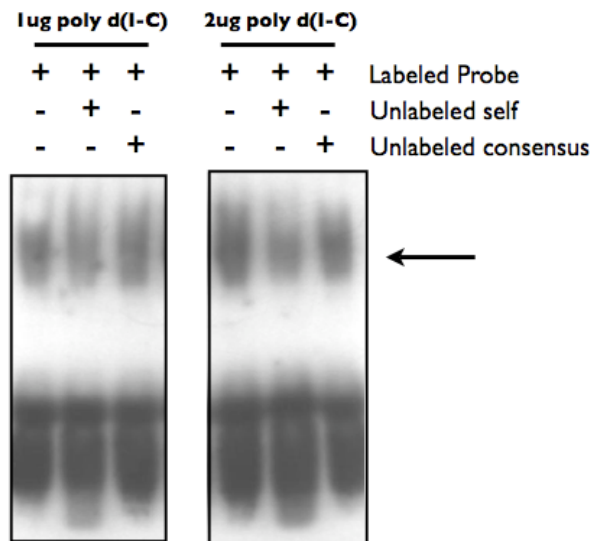


Figure 5.2. EMSA analysis of -99 and -78 GATA motifs within ms1 proximal promoter. DIG-labelled oligonucleotide probes for -99 (B) and -78 (C) GATA motifs (A) within ms1 proximal interval were incubated with whole cell protein extracts made from H9c2 myoblasts (B, C). Competition experiments were performed using a 200-fold molar excess of unlabelled self and GATA consensus probe in addition to different concentrations of polyd(I-C). Arrows indicate resulting band shifts.

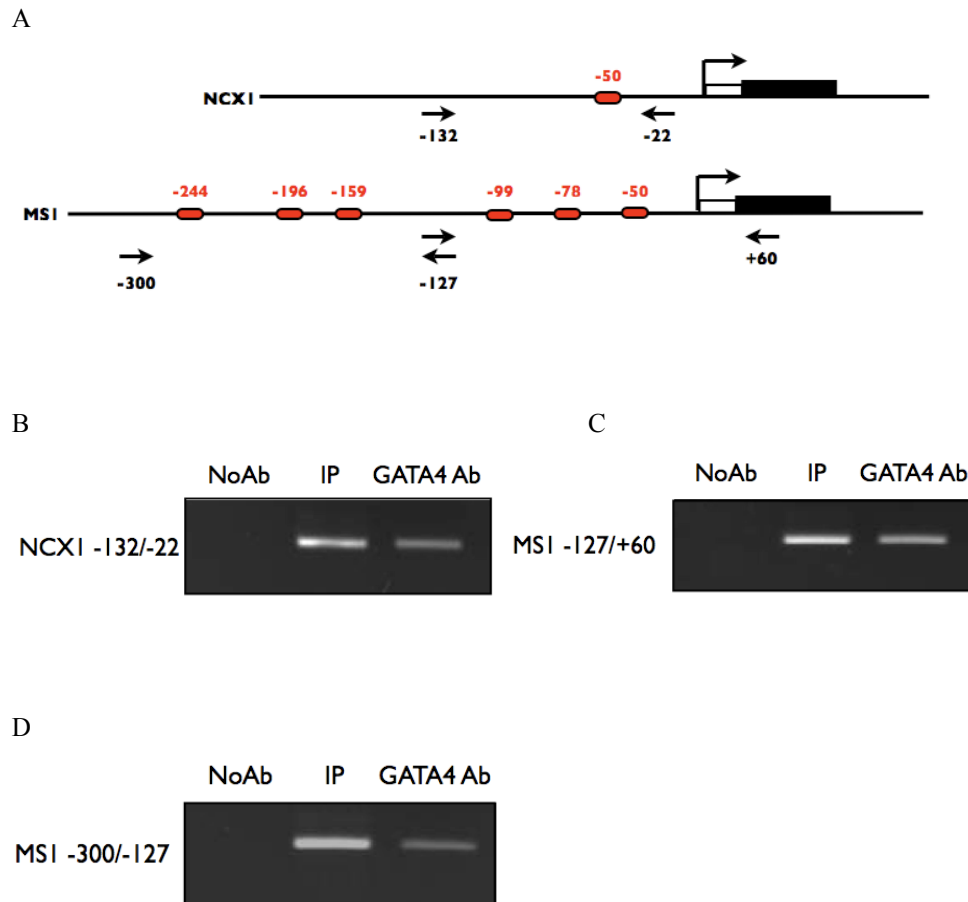


Figure 5.3. Chromatin immunoprecipitation (ChIP) was performed on isolated adult feline cardiomyocytes to evaluate *in vivo* GATA4 occupancy within the ms1 proximal promoter. (A) Schematic representation of amplified DNA fragments and primer locations encompassing the ms1 and NCX1 proximal promoter domains. GATA motifs are highlighted by block red circles with location relative to TSS annotated above. Proteins were crosslinked to the DNA with formaldehyde, DNA was sheared by sonication, and Ab's directed against GATA4 was added to immunoprecipitate protein DNA complexes. Immunoprecipitations were performed without primary antibody (No Ab) as a negative control. Input DNA (IP) is also shown as a positive control. Semi-quantitative PCR was then performed on isolated DNA using primers to amplify the NCX1 GATA4 binding domain (B), the -127/+60 (C) and -300/-127 ms1 promoter intervals Amplification products were electrophoresed on agarose gel for visualisation.

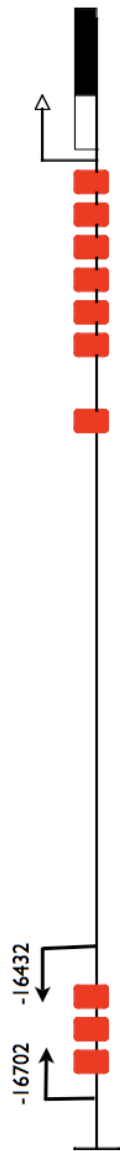
5.2.3 *GATA4 targeting at the UP3 domain is required for enhancer activity.*

The UP3 domain, is also enriched with three GATA motifs including one matching the consensus sequence perfectly (MS1/GATA-16600). To test the sensitivity of this domain to exogenous GATA4 activity, we transiently transfected the UP3-SV40 reporter into H9c2 and NIH 3T3 cells in combination with WT and DN-GATA4 over-expression plasmids (Figure 5.4.B). As one would expect for a bona-fide GATA4 sensitive regulatory module, the UP3-SV40 reporter was activated (in H9c2) in a dose-dependent fashion 1.6-fold (0.5µg expression plasmid, $p<0.05$) and 2-fold (1.0µg expression plasmid, $p<0.05$), compared to the control (1.0µg pcDNA only). Activation occurred in NIH 3T3 cells, although the level of activation was much lower, with only the higher concentration (1.0µg expression plasmid, $p<0.05$) able to activate the reporter 1.4-fold. This indicates cardiac specific mechanisms can augment the transactivation potential of WT GATA4 when acting on the UP3 domain. Interestingly, DN-GATA4 repressed the UP3-SV40 reporter by 20% and 50% ($p<0.05$) in H9c2 cells. This supports the importance of GATA4 targeting to this domain and demonstrates basal GATA4 binding is important for the cardiac specific activity of the UP3 enhancer. All co-transfection combinations were executed with the empty pGL3-B reporter vector, with no observable non-specific effects (Figure 5.4.C).

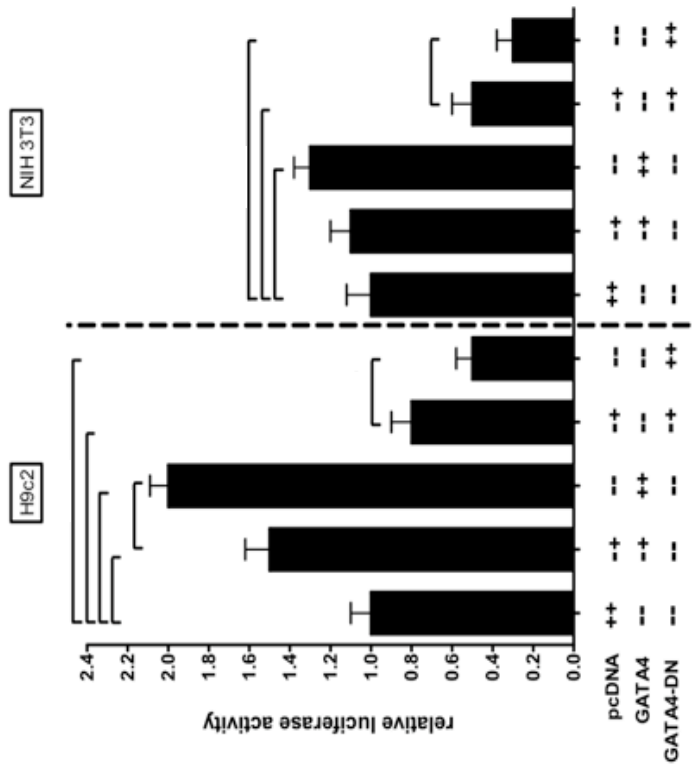
Of the three GATA motifs within the UP3 domain, the -16600 was the most interesting as it represented the perfect GATA4 consensus sequence. We therefore generated a DIG-labelled oligonucleotide probe (Figure 5.5.A) corresponding to this motif and performed EMSA analysis to test the potential *in vitro* binding capacity for H9c2 derived GATA4 protein (Figure 5.5.B). Incubation of this MS1/GATA-16600 labelled probe with H9c2 whole cell protein extracts resulted in the appearance of a single DNA-protein band shift (represented by arrow)(Figure 5.5.A). Co-incubating this reaction with a 200 fold molar excess of unlabelled complimentary probe ablated the appearance of the shifted band, demonstrating this DNA-protein band shift is the product of a protein/DNA-sequence specific interaction. When competing with the cold GATA consensus sequence, we also observed specific band ablation. This suggests that GATA4 is likely to be binding to this GATA motif *in vitro*.

To validate this *in vivo*, semi-quantitative PCR on anti-GATA4 IP DNA was carried out using primers spanning the UP3 domain (Figure 5.6.A). As shown in Figure 5.6.B, GATA4 antibodies specifically pulled down DNA fragments containing the UP3 domain. The “no antibody” negative control did not pull down such sequences. These results indicate that GATA4 is able to bind this domain *in vivo*, however, due to the close proximity of individual GATA motifs, we can not definitively ascertain which of the three motifs are binding GATA4. Nevertheless, these data suggest that GATA4 directly interacts with the UP3 domain *in vivo*, with this binding important for cardiac specific activity of this enhancer.

A



B



C

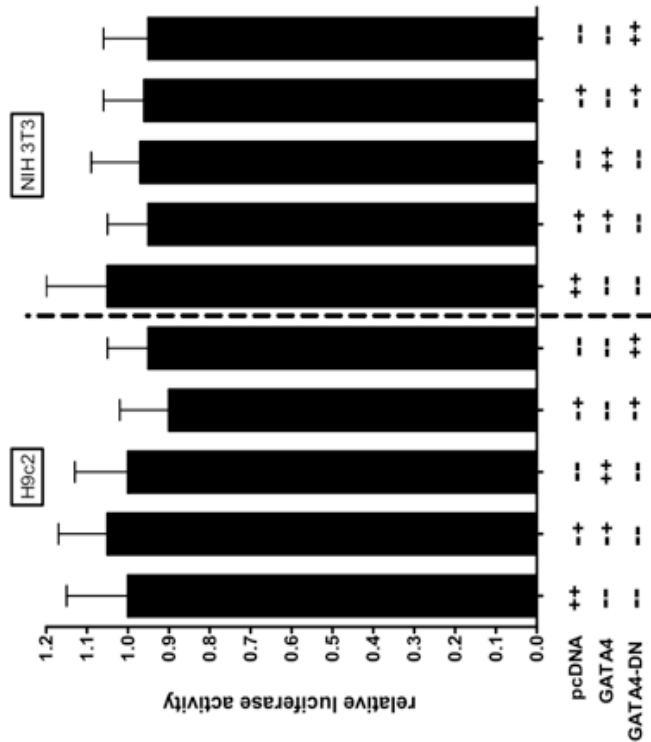
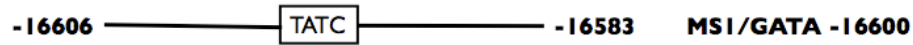


Figure 5.4. The UP3 domain confers GATA4 sensitivity. (A) Schematic representation of the UP3 domain with annotated luciferase construct primer sequences and enriched GATA motifs represented by solid red boxes. Vectors (+;0.03 μ g, ++;0.6 μ g) expressing wild type and dominant negative (DN) GATA4 were co-transfected with the UP3 (B) and pGL3-P (F) reporter vectors into H9c2 myoblasts and NIH 3T3 fibroblasts. The relative activity of the reporter transfected with pcDNA alone was arbitrarily set at 1-fold activation. The results are expressed as mean \pm SEM of at least three different separate transfections in triplicate for each reporter construct. Statistically significant differences are indicated by tailed bars, $p < 0.05$.

A



B

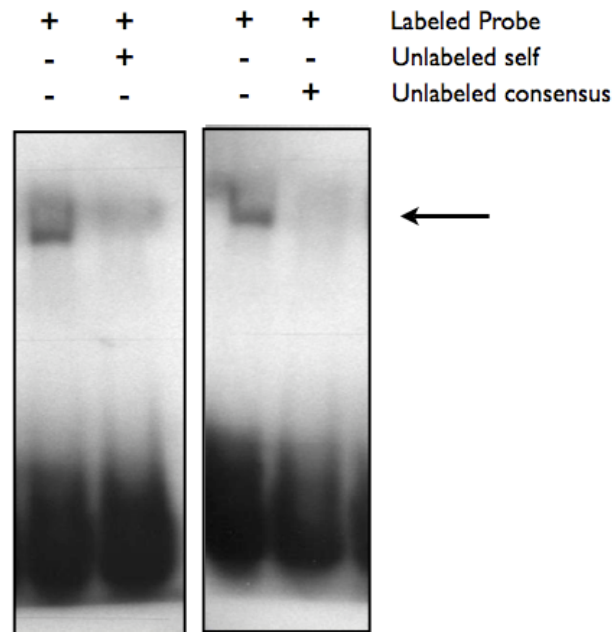


Figure 5.5. EMSA analysis of -16600 GATA motif within the UP3 domain. DIG-labelled oligonucleotide probe for -16000 ms1 GATA motif (A) within the UP3 domain were incubated with whole cell protein extracts made from H9c2 myoblasts (B). Competition experiments were performed using a 200-fold molar excess of unlabelled self and GATA consensus probe (B). Arrow indicates resulting band shift.



Figure 5.6. Chromatin immunoprecipitation (ChIP) was performed on isolated adult feline cardiomyocytes to evaluate *in vivo* GATA4 occupancy within the ms1 UP3 domain. (A) Schematic representation of amplified DNA fragments and primer locations encompassing the ms1 UP3 domain. GATA motifs are highlighted by block red circles. Proteins were crosslinked to the DNA with formaldehyde, DNA was sheared by sonication, and Ab's directed against GATA4 was added to immunoprecipitate protein DNA complexes. Immunoprecipitations were performed without primary antibody (No Ab) as a negative control. Input DNA (IP) is also shown as a positive control. Semi-quantitative PCR was then performed on isolated DNA using primers to amplify the -16702/-16432 ms1 UP3 interval (B). Amplification products were electrophoresed on agarose gel for visualisation.

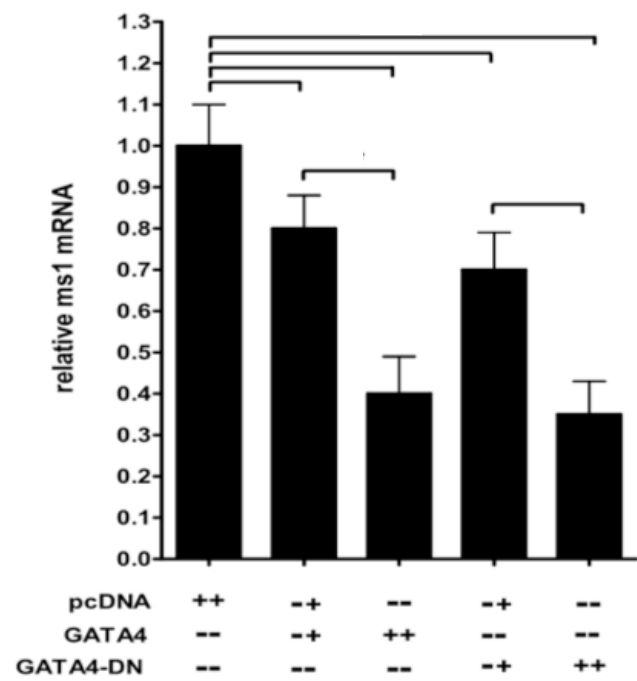
5.2.4 *GATA4 modulates endogenous *msl* expression in vitro.*

In order to reconcile the regulatory differences in activity associated with GATA4 binding at the proximal and distal domains (PP and UP3), the sensitivity of the endogenous *msl* promoter to GATA4 expression was examined in our cell culture model. The WT and DN-GATA4 proteins were overexpressed in H9c2 myoblasts, in varying concentrations as performed in the reporter co-transfection assays. The expression level of endogenous *msl* mRNA in both control (empty pcDNA vector alone) and WT/DN GATA4 transfected cells was determined by quantitative real time RT-PCR using Taq Man probes. Interestingly, WT GATA4 over-expression significantly decreased *msl* mRNA levels in a dose dependent manner: 20% (0.5 μ g wtGATA4, $p < 0.05$) and 70% respectively (1.0 μ g, $p < 0.05$) (Figure 5.7.A). This demonstrates that WT GATA4 represses endogenous *msl* expression suggesting the net effect of GATA4 at the repressive PP domain abolishes UP3-GATA4 dependent activation. This raises interesting questions as to the bona-fide *in vivo* role of the UP3 cardiac enhancer domain. DN-GATA4 over-expression also decreased *msl* mRNA levels in a dose dependent fashion: 20% (0.5 μ g wtGATA4, $p < 0.05$) and 70% respectively (1.0 μ g, $p < 0.05$), supporting the hypothesis that repression is via the DNA binding capacity of the GATA4 protein. These findings corroborated the present luciferase based reporter assays,

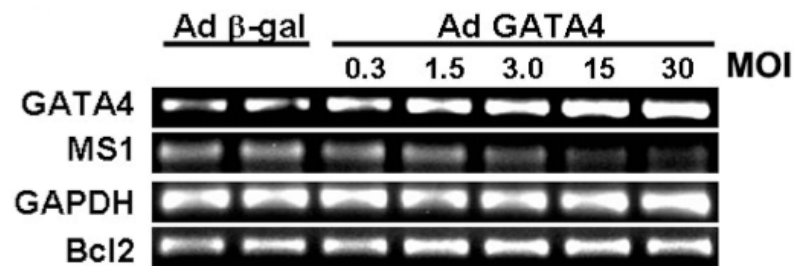
In order to confirm these *in vitro* findings in a more physiological environment, primary NRVM and an adenovirus based over-expression strategy was used (in collaboration with Rong Liang, Sanford School of Medicine, University of South Dakota). To determine whether over-expression of WT GATA4 could decrease endogenous *msl* expression in primary cardiac cells, NRVM were infected with adenovirus encoding β Gal or GATA4 (executed by Satoru Kobayashi, SSM, USD, for method refer to (Kobayashi et al, 2006)). As shown in Figure 5.7.B, over-expression of GATA4 was able to repress endogenous *msl* expression in a dose-dependent manner. This dose-dependent repression correlated with increased GATA4 abundance and Bcl2 expression [a characterised GATA4 target gene (Kobayashi et al, 2006)]. The greatest repression of *msl* mRNA was achieved with a Ad-GATA4 MOI of 30 at 48 h after viral infection,

resulting in a significant 50% down-regulation of *msl* (Figure.4.7.C, n=4 p<0.05). These results demonstrated that GATA4 is able to repress endogenous *msl* expression in primary NRVM *in vitro*, thus confirming the H9c2 data. To further demonstrate the necessity of GATA4 for the baseline modulation of *msl* expression, an adenovirus encoding a short hairpin RNA targeted to GATA4 (AdGATA4i) was used to specifically knock down GATA4 in NRVM (executed by Satoru Kobayashi, SSM, USD). As shown in Figure 4.7.D infection with AdGATAi at MOI of 60 for 48 hours resulted in a significant 1.3-fold increase in endogenous *msl* expression (n=4 p<0.05) when compared to infection with an adenovirus encoding a random scrambled short hairpin RNA (AdCONi). These results demonstrate that endogenous GATA4 is required for cardiac specific repression of *msl* expression in NRVM, *in vitro*. This supports our previous dominant-negative findings and demonstrates this effect is specific to GATA4 rather than a non-specific GATA6 mediated pathway.

A



B



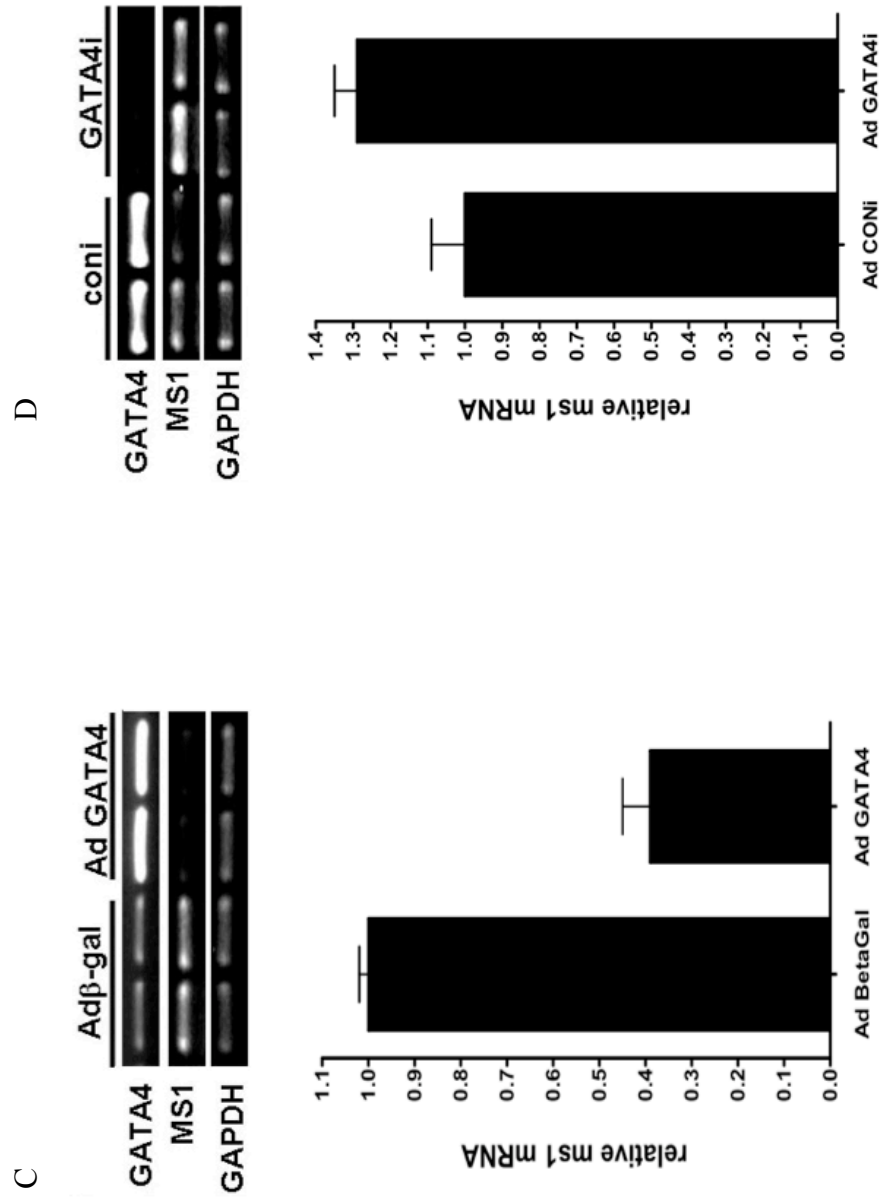


Figure 5.7. GATA4 modulates endogenous *ms1* expression *in vitro*. (A) H9c2 myoblasts were transiently transfected with vectors (+;0.05µg, +;1.0µg) expressing wild type and dominant negative (DN) GATA4. 48 hours post transfection at which point the cells were 80% confluent, total RNA was isolated, reverse transcribed and expression levels of TBP and *ms1* were determined by real time quantitative PCR. Statistically significant differences are indicated tailed bars, $p < 0.05$. (B) Neonatal rat ventricular myocytes (NRVM) were infected with Adgal or AdGATA4 for 48 hours at MOI stated (B) or MOI of 30. Alternatively (C) NRVM were infected with AdConi or AdGATA4i at MOI of 60 for 48 hours. Total RNA was extracted and subjected to reverse transcription followed by semi-quantitative PCR using primers specific for GATA4, GAPDH and *ms1*. Amplified products were visualised on ethidium bromide stained agarose gel and quantified by densitometry. Values are represented as means \pm SEM of at least 4 separate experiments.

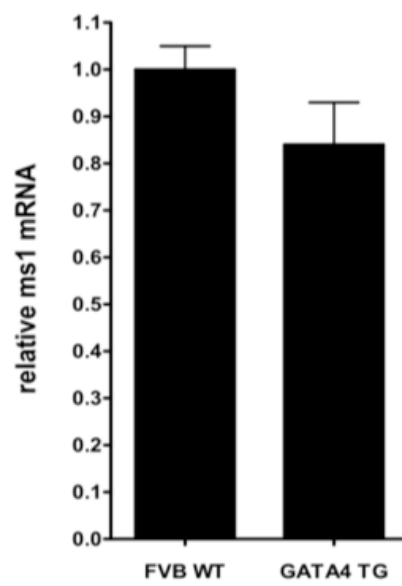
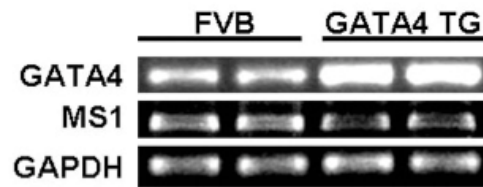
5.2.5 *GATA4 modulates endogenous *msl* expression in embryonic and adult murine hearts in vivo.*

To examine the effect of GATA4 on *msl* gene expression *in vivo*, semi-quantitative RT-PCR was used to determine mRNA levels of GATA4 and *msl* in FVB/N WT and GATA4 transgene (TG) hearts (Figure 5.8.A). These GATA4 TG mice were kindly provided by Professor Jeffrey Molkentin (Cincinnati Childrens Hospital) and have previously been described (Liang et al, 2001a). These mice demonstrate a robust increase in GATA4 mRNA (Figure 5.8.A) which corresponds to ~2 to 4-fold increase in GATA4 protein levels relative to the wild type controls (Liang et al, 2001a). Remarkably, as shown in Figure 5.8.A, *msl* mRNA levels were significantly down-regulated in the GATA4 TG hearts compared with the WT controls (20% decrease in *msl* mRNA levels, n=4, p<0.05). These results therefore suggest that GATA4 modulates *msl* expression with increased levels able to repress *msl* expression *in vivo*.

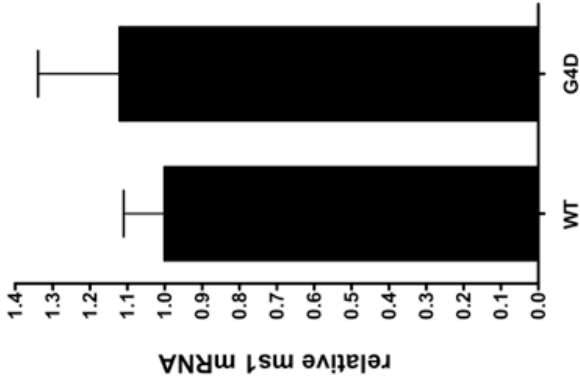
With over-expression of GATA4 repressing endogenous *msl* expression *in vivo*, we sought to confirm that basal GATA4 levels within the whole heart also contribute to basal *msl* expression. We therefore, in collaboration with Dr William Pu (Harvard Medical School), proceeded to measure endogenous *msl* expression in gene targeted mice that express reduced levels of GATA4 in the adult heart (Pu et al, 2004). These mice were on a uniform genetic background that were heterozygous for cardiac specific deletion of the second exon of GATA4 (GATA4 WT/TriEx2, abbreviated G4D), which includes the start codon and N-terminal 46% of the coding region (Pu et al, 2004). These mice have been shown to express ~50% decrease in GATA4 mRNA and protein within the adult heart. We proceeded to execute quantitative real time RT-PCR using *msl* and TBP Taq man probes on 1µg of total RNA obtained from the adult hearts (n=4) of G4D mice and wild-type littermate controls. As shown in Figure 5.8.B, *msl* expression was not significantly altered in these mice, suggesting that within the adult whole heart, a 50% decrease in GATA4 is not sufficient to alter endogenous *msl* expression.

To further investigate this we utilised a different GATA4 gene-targeted knock out model. This model (Zeisberg et al, 2005) utilises *Cre/Lox* technology to conditionally ablate GATA4 (floxed GATA4 allele as used in the G4D mice) at embryonic day 9.5 (E 9.5) specifically within the cardiomyocytes of the developing myocardium. This early cardiac restricted deletion of GATA4 was achieved through the utilisation of Nkx2.5Cre, in which *Cre* recombinase expression is driven by the endogenous Nkx2.5 locus (Moses et al, 2001), resulting in robust *Cre*-mediated recombination by E 9.5. A strategy of heterozygous and homozygous floxed GATA4 allele in combination with Nkx2.5Cre was used to dose-dependently ablate GATA4. The homozygous floxed GATA/Nkx2.5*Cre* mouse hearts expressed 90% less GATA4 mRNA compared to the wild type controls as previously demonstrated through quantitative RT-PCR (Pu et al, 2004). We proceeded, using previously extracted RNA (Zeisberg et al, 2005), to execute quantitative RT-PCR on total RNA obtained from the hetero and homozygous floxed GATA-4/Nkx2.5-*Cre* murine hearts in addition to Nkx2.5-wild type hearts (RNA extracted by Bill Pu). Encouragingly, as shown in Figure 5.8.C, we demonstrate a dose-dependent increase in *msl* expression in the hetero (1.75-fold, $p < 0.05$) and homozygous (2.25-fold, $P < 0.05$) floxed GATA-4/Nkx2.5-*Cre* murine hearts compared to the non-*Cre* Nkx2.5 wild type hearts.

A



B



C

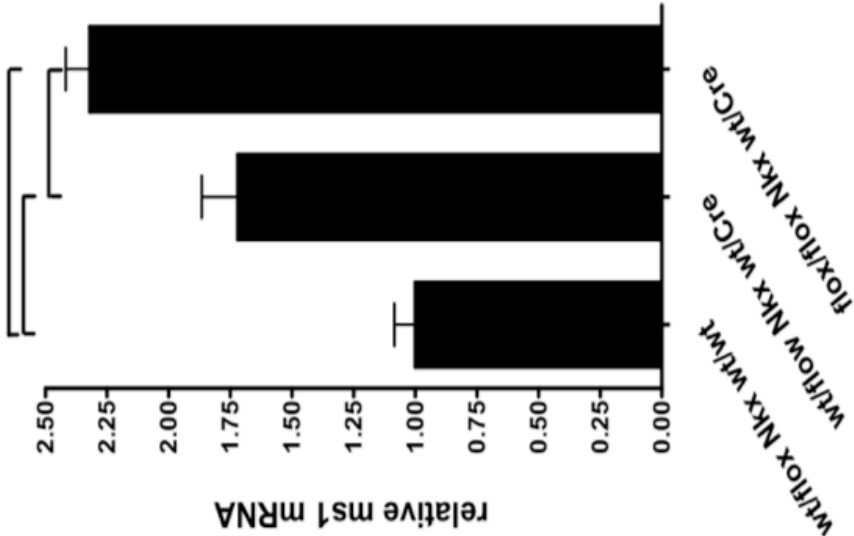


Figure 5.8. Total RNA was prepared from the hearts of FVB (A), GATA4 TG (A), wild type (B), G4D (B), wt/flox Nkx wt/wt (C), wt/flox Nkx wt/Cre (C) and flox/flox Nkx wt/Cre (C) mice and subjected to reverse transcription followed by semi-quantitative (SQ) (A) and quantitative real time PCR with mouse GATA4 (A), ms1 (A, B, C), GAPDH and TBP specific primers. For SQ-PCR (A), amplified products were visualised on ethidium bromide stained agarose gel and quantified by densitometry. Values are represented as means \pm SEM of at least 4 separate heart preparations. Statistically significant differences are indicated by tailed bars, $p < 0.05$.

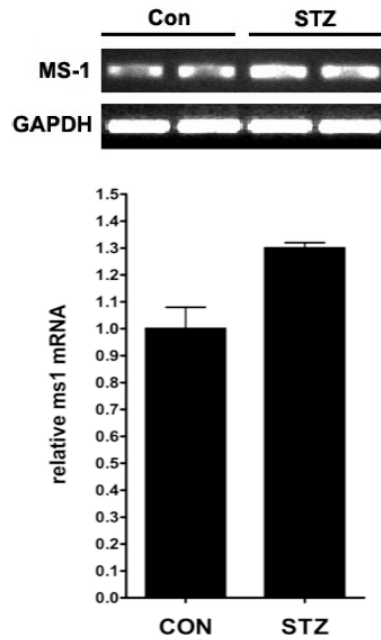
5.2.6 *GATA4 associated pathological phenotypes correlate with *msl* transcriptional dysregulation.*

In addition to its role as a crucial transcription factor essential for developmental, physiological and adaptive responses of cardiomyocytes (Aries et al, 2004; Kobayashi et al, 2006), recent studies have also implicated GATA4 in hyperglycemia (HG) induced cardiomyopathy (Kobayashi et al, 2006). Kobayashi and colleagues (2005) demonstrated that diabetes associated HG induces GATA4 depletion in cardiomyocytes. This was due to increased degradation of GATA4 protein via the ubiquitin proteasome system (UPS). These authors suggest a mechanism for this whereby HG induces the expression of the E3-ubiquitin ligase, CHIP, which specifically ubiquitinates GATA4 targeting it for UPS-mediated degradation. This study reported that within mouse models for type 1 and type 2 diabetes, CHIP levels are high with concomitant low levels of GATA4 protein. This decrease in GATA4 leads to increased cardiomyocyte apoptosis, presumably as a result of down-regulated expression of anti-apoptotic genes downstream of GATA4, which results in cardiomyopathy.

Data presented in this chapter demonstrates that endogenous GATA4 levels within cardiomyocytes is essential in *msl* expression. We therefore hypothesised that GATA4 depletion, as observed in the above pathological phenotypes, may result in dysregulated *msl* expression, which was measured in the diabetic mouse heart (in collaboration with Rong Liang). Type 1 diabetes was induced in two-month old FVB mice with streptozotocin (STZ, i.p. 150 mg/kg body weight in 10mmol/L sodium citrate, pH4.5, executed by Kobayashi) which is a well established agent that destroys pancreatic β cells (Rossini et al, 1977a; Rossini et al, 1977b). At 4 weeks post STZ treatment total RNA was extracted from these diabetic hearts and vehicle controls. At this time point, these STZ treated hearts have been shown to express 44% lower levels of GATA4 protein compared to VEH controls. In these mice, we observe a 1.3-fold increase in relative *msl* mRNA ($p < 0.05$) (Figure 5.9.A) suggesting GATA4 depletion induces *msl* transcription. We also measured *msl* expression in 4-month old db/db mice, a well characterised Type 2 insulin-resistant diabetic model with early onset cardiomyopathy (Aasum et al, 2003). Similarly, *msl* mRNA was increased in db/db diabetic hearts

compared with db/+ non-diabetic controls (Figure 5.9.B, 1-3 fold increase in *msl* mRNA, $p < 0.05$), with the db/db diabetic hearts previously shown to express 30% less GATA4 protein. In summary, diabetes (type 1 and type 2) diminishes cardiac GATA4 protein levels, which is associated with significant increase in *msl* mRNA abundance. This correlates with our findings in NRVM, and again supports the hypothesis that endogenous GATA4 represses *msl* expression in the adult heart.

A



B

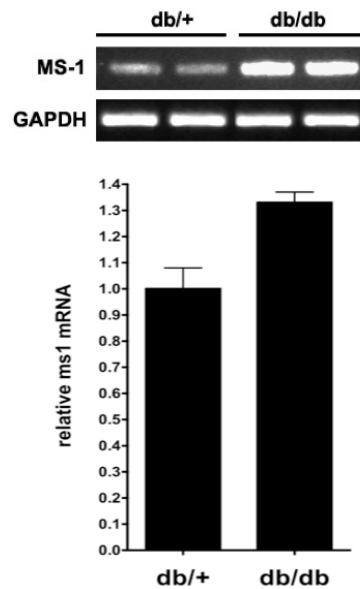


Figure 5.9. Total RNA was prepared from the hearts CON (A), STZ (A) (Type 1), db/+(B) and db/db (B) mice and subjected to reverse transcription followed by semi-quantitative (SQ) PCR with mouse ms1, and GAPDH specific primers. Amplified products were visualised on ethidium bromide stained agarose gel and quantified by densitometry. Values are represented as means \pm SEM of at least 4 separate heart preparations. Differences between db/+ and db/db are statistically significant, $p < 0.05$.

5.3 Discussion

In previous chapters (Chapters 3 and 4) we identified and functionally characterised two CRMs, PP and UP3, with cardiac specific regulatory activity. However, in order to elucidate the exact cardiac specific regulatory nature of these domains one needs to identify and characterise the TFBS and cognate factors acting on them. The comparative *in silico* analysis (Chapter 3) demonstrated that GATA motifs were the most highly enriched TFBS within the PP and UP3 domains. The predominant GATA binding protein within the heart, GATA4, has been well characterised in numerous studies as a critical mediator of cardiac specific gene expression. In particular, GATA4 driven gene expression contributes to many cardiac regulatory contexts including cardiac development, differentiation, maturation and post-natal adaptation, all contexts in which *msl* is differentially expressed. We therefore hypothesised that GATA4 may play an important role in regulating *msl* expression in these cardiac specific regulatory contexts.

Using both gain- and loss-of-function approaches, the current chapter provides strong evidence to suggest that *msl* is a bona-fide target of GATA4 *in vivo*. However the exact mechanisms are complex, with the present data showing that GATA4 acts in both a negative and positive manner at the proximal (PP) and distal (UP3) cardiac regulatory domains respectively. As the first step in the analysis we experimentally determined the regulatory effect mediated by GATA4 (wild type and dominant negative protein forms) on the PP and UP3 domains in isolation. Our findings demonstrated that GATA4 repressed reporter activity via the proximal 127bp upstream of the *msl* TSS. This repression was as a consequence of the DNA binding activity of GATA4 rather than a cardiac specific repressive mechanism (for example, HDAC recruitment), exemplified by the observation that wild type GATA4 repressed *msl* in the non-cardiac (low GATA4 expressing) NIH 3T3 fibroblast cells. ChIP analysis confirmed binding of endogenous GATA4 to the proximal -127 loci in adult cardiomyocytes. This data do not exclude a positive regulatory role for GATA4. Indeed, our UP3 domain was activated by GATA4 in a cardiac and non-cardiac environment. GATA4 activated the UP3 domain more

potently in the cardiac environment, suggesting the utilisation of cardiac specific factors which synergise and co-operate with GATA4 to enhance its activity. This is not an uncommon observation with many studies demonstrating combinatorial activation of gene expression through GATA4 interactions with other cardiac restricted transcription factors (Kuo et al, 1999; McBride et al, 2003; Morin et al, 2000; Morin et al, 2001; Sepulveda et al, 1998). Endogenous GATA4 was shown to bind this domain *in vivo* with some of this binding occurring through the -16600 consensus GATA motif. Data presented here also suggested that the cardiac specific activity associated with UP3 (demonstrated in Chapter 4) is mediated, at least in part, through endogenous GATA4 binding. This is inferred from the effect of DN-GATA4 which repressed basal activity of the enhancer.

The current findings begin to highlight the complex hardwiring of distinct CRM in mediating *msl* transcriptional control and demonstrate the ability of GATA4 to regulate *msl* activity at distinct *cis* regulatory domains in both a positive and negative manner. However, one caveat with this data is that they are based on the activity of the *msl* CRMs in isolation and thereby give limited insight into the regulatory effect exerted by GATA4 on the endogenous *msl* transcriptional unit. This endogenous regulatory circuit will comprise of the PP and UP3 domains in their native genomic environment where they are interacting with each other and potentially other domains (Firulli & Olson, 1997), with all of these native interactions modulating the overall transcriptional effect of GATA4. It was therefore important to reconcile the regulatory effects of GATA4 on the UP3 and PP domains *in vivo* and determine the “net” effect of GATA4 on the endogenous *msl* transcriptional unit. Through utilising an adeno-viral based strategy to over-express and knock down (via shRNA) GATA4 in NRVM, we demonstrated a clear effect on endogenous *msl* expression *in vitro*. We showed that *msl* mRNA was depleted by ectopic GATA4 in NRVM, suggesting that the “net” effect of GATA4 on the endogenous *msl* transcriptional unit is repression. In support of this, siRNA-mediated GATA4 knockdown markedly increased *msl* mRNA, indicating that endogenous GATA4 is also able to repress *msl* transcription. This finding is of particular importance as it demonstrates a specific role for GATA4 rather than GATA6 in *msl* control.

In support of the *in vitro* findings, at the whole animal level, *msl* mRNA was significantly down-regulated in GATA4 expressing transgenic hearts compared with WT, whereas in embryonic heart specific GATA4 targeted mice *msl* mRNA was significantly up-regulated. It was interesting to note that within the adult GATA4 gene targeted mouse, *msl* mRNA was not significantly increased. This may be a consequence of functional redundancy, or suggest that alternative GATA4-dependent regulatory mechanisms are acting within the adult, embryonic and neonatal hearts respectively. To conclude, the data presented here demonstrates that within embryonic, neo-natal and adult cardiomyocytes, GATA4 serves as a bona-fide transcriptional regulator of *msl*. Although our data suggest that GATA4 is able to positively regulate *msl* through the UP3 domain in isolation, within the native genomic context, the “net” effect of GATA4 is to repress *msl* transcription. In light of our promoter reporter assays, we suggest that this observed “net” repression is mediated via GATA4 binding within the PP domain, and specifically to our enriched GATA motifs within the proximal -127/+60 region.

The present findings raises important questions with respect to the exact nature and mechanism of this GATA4-dependent repressive process. Two important characteristics emerge from our data that potentially give an insight: repression appears to be dependent on the DNA-binding activity of GATA4 protein, and this repressive DNA binding is occurring within the proximal -127/+60 region. These proximal GATA motifs are within close proximity to our bona-fide TATA box, the site of transcriptional initiation. We therefore propose that the location of the GATA4 binding GATA motifs, with respect to the TATA box, may have direct functional implications on *msl* transcription and specifically GATA4-dependent repression. It is therefore of great interest that Murakami and colleagues (2002) have reported that the Fgf-3 promoter activity is dependent on positive and negative regulatory interactions with GATA4. They found that GATA4 bound two domains within the promoter, designated PS4A and PS13, which function as positive and negative regulatory elements respectively. These authors demonstrated that the negative regulatory activity of the PS13 element, located only 106bp upstream of the TSS, was dependent on its close proximity to the TSS (Murakami et al, 2002). These authors suggested that the negative effect on transcription could be due to the GATA4 complex on PS13 sterically interfering with

assembly of a common set of basal factors on the core promoter. This would then interfere with TBP recruitment and subsequent transcriptional initiation and elongation thereby blocking and repressing transcription. The three “repressive” GATA motifs presented here are located 99, 78 and 33 bp upstream of the *msl* TSS, a region which encompasses the core promoter, raising the possibility that a similar mechanism may exist for *msl* gene regulation. We therefore propose that the GATA4 dependent repression of the *msl* promoter is occurring through a steric hindrance mechanism that interrupts the binding of the basal transcriptional apparatus to the TATA box. This mode of action serves to “lock” the *msl* promoter, thereby providing an exquisite mechanism for tight control of *msl* transcription. Endogenous GATA4 levels therefore act as a “rheostat” for this transcriptional locking with this having fundamental implications for *msl* transcription during GATA4-associated cardiac regulatory contexts.

To add, what is the regulatory purpose of GATA4 activation at the UP3 module when in our models, the “net” GATA4 effect is to repress *msl* transcription? We suspect that UP3 modulates a cardiac GATA4-dependent regulatory process in a context that were not able to experimentally mimic, one example being early cardiac lineage determination. As mentioned in Chapter 4, the UP3 module is enriched with Smad binding elements in addition to GATA motifs. We speculated that the UP3 module may represent an early cardiac developmental enhancer, analagous to the Nkx2.5 enhancer (Lien et al, 2002), which through BMP-Smad signalling in collaboration with GATA4 initiates *msl* expression at around E8.5. The data presented here (and chapter 4) provide some support for this possibility as GATA4 is involved in regulating UP3 activity. It would be of great interest to further examine the Smad binding motifs within this module and determine if, like for the Nkx2.5 enhancer, Smad proteins and GATA4 collaborate to induce early developmental cardiac specific gene expression.

Through functional regulatory analysis of numerous promoters, best exemplified by the ANP and BNP promoters (Temsah et al, 2005), GATA4 has been shown to collaborate with numerous cardiac regulatory factors in addition to the Smad proteins including SRF, Nkx2.5, Hand1, Mef2 and the Klf s (Kuo et al, 1999; Lavalley et al, 2006; McBride et al, 2003; Morin et al, 2000; Morin et al, 2001; Sepulveda et al, 1998). When one

considers the enrichment of binding sites for these regulatory factors in our PCRM, it is interesting to consider possible interactions occurring at the *msl* promoter, and how such regulatory interactions contribute to *msl* transcriptional modulation. Indeed the regulatory intervention of these factors *in vivo* may help us gain an insight into the differences in *msl* sensitivity to GATA4 perturbation in the embryonic and adult models used here. Potentially these factors may represent embryonic/adult regulatory proteins that act as genetic modifiers of *msl* transcription in the developmental and mature cardiac context.

In conclusion, the data presented in this chapter demonstrates that *msl* is a bona-fide target of GATA4 *in vivo*. Supplementary to giving us an insight into the composition and regulatory functionality of our identified PCRM, this work further enlightens our fundamental understanding of GATA4 function. From this perspective, it is clear that through regulation of *msl*, the GATA4 cardiac gene regulatory network (GRN) is able to integrate with the SRF GRN. This would be of functional significance for processes including cardiac development and differentiation which are critically dependent on temporal coordination between the GATA4 and SRF GRNs. These findings also have implications for disease states in which GATA4 abundance and/or activity is compromised. Clearly, the exquisite interaction between GATA4 and *msl* expression would suggest any changes in GATA4 abundance and/or activity would reflect on *msl* abundance. We demonstrate here that Type1/2 diabetes associated GATA4 depletion results in concomitant increase in *msl*. Spurious *msl* expression and SRF activity has been shown to lead to cardiomyopathy (Kuwahara et al, 2007) and it would be tempting to speculate that hyperglycemia induced diabetic cardiomyopathy [linked to downregulation of GATA4 via increased targeting to the UPS (Kobayashi et al, 2006)], may be in some part mediated by sustained *msl* expression and SRF activity. Taking this further, one can not rule out the possibility that dysregulated *msl* expression and subsequent effects on SRF activity may also be a primary driving force in many of the adverse phenotypes associated with GATA4 dysregulation such as congenital heart defects, cardiac hypertrophy and cardiomyopathy (Pikkarainen et al, 2004).

Chapter 6

Regulatory characterisation of the transcriptional mechanisms governing skeletal muscle myocyte stress 1 expression.

6.1 Introduction

During mammalian embryogenesis, the development of skeletal muscle is mediated by a co-ordinated series of events that begins with the commitment of mesodermal precursor cells to the skeletal muscle lineage, followed by myoblast fusion and the subsequent progression of a programme of muscle specific gene expression (Buckingham et al, 2003; Christ & Ordahl, 1995; McKinsey et al, 2002). A specialised group of transcription factors control this process of myogenic specification and differentiation. These factors, designated the myogenic regulatory factors (MRFs), include four basic helix-loop-helix (bHLH) E-Box binding proteins: MyoD, Myf5, Myogenin and MRF4 (Molkentin et al, 1995). During development MyoD and Myf5 dictate myoblast specification while Myogenin and MRF4 regulate terminal differentiation (Andres & Walsh, 1996; Dedieu et al, 2002). In collaboration with the MRFs, the MADS-box myocyte enhancer factor (MEF) family of proteins also contribute to the programme of muscle specific gene expression (Black & Olson, 1998; Molkentin & Olson, 1996).

Serum response factor (SRF), the DNA binding transcription factor component of the MS1-MRTF-SRF transcriptional axis, is also important for skeletal muscle specific gene expression. In addition to binding and regulating a wide array of skeletal muscle specific promoters, the targeted perturbation of SRF activity *in vitro* has been shown to severely impair myoblast fusion and differentiation (Carson et al, 1996; Catala et al, 1995; Croissant et al, 1996; Vandromme et al, 1992; Wei et al, 1998). To confirm this important role for myogenic SRF activity in an *in vivo* context, a conditional skeletal muscle specific murine SRF knockout was generated. Mice derived from this knockout

demonstrated impaired myoblast fusion causing severe skeletal muscle myopathy that ultimately resulted in perinatal lethality (Li et al, 2005).

A requisite role for the SRF co-factor MRTF-A (a myocardin related transcription factor) in skeletal muscle development has also been inferred from experiments in cultured muscle cells. Within this *in vitro* context knockdown of MRTF-A, via RNAi, repressed SRF-dependent gene expression resulting in impaired myoblast fusion and attenuated formation of terminally differentiated multinucleated myotubes (Selvaraj & Prywes, 2003). *In vivo*, transgenic mice expressing a dominant negative form of MRTF-A displayed a phenotype reminiscent of the skeletal muscle SRF knock out mice, thus supporting an important role for the MRTFs in the control of muscle specification and differentiation (Li et al, 2005).

As a critical regulator of the MRTF-SRF transcriptional axis it is unsurprising to find that *msl* has a functional role in skeletal muscle differentiation and maturation. In the C2C12 myoblast cell line *msl* perturbation, gene knockdown via RNAi resulted in a significant attenuation of muscle specific SRF activity caused by decreased MRTF nuclear translocation (Kuwahara et al, 2005). We have recently shown that morpholino knockdown of zebrafish *msl* (*ZMS1*) *in vivo* results in severe musculoskeletal deformities with curvature and shortening of the longitudinal axis (Mahadeva, H unpublished findings). These findings, in addition to the published work demonstrating the important myogenic role of the MRTF-SRF pathway, lead us to speculate that mechanisms governing *msl* expression will be implicated in skeletal muscle determination, differentiation and maturation.

At the commencement of the current project little was known about the transcriptional mechanisms governing skeletal muscle expression of *msl*. However it is of interest that Eric Olson's group have recently demonstrated that the proximal 1.5kbp 5'-flanking sequence was able to direct LacZ expression in adult cardiac and skeletal muscle (Kuwahara et al, 2007). Two Mef2 responsive motifs within this region were shown to

be critical for the observed cardiac specificity although the factors, motifs and regulatory mechanisms governing skeletal muscle specific expression were not examined. However this study did demonstrate that the proximal 1.5kbp 5'-flanking sequence contains sufficient cis information to drive skeletal muscle specific expression within the *in vivo* context.

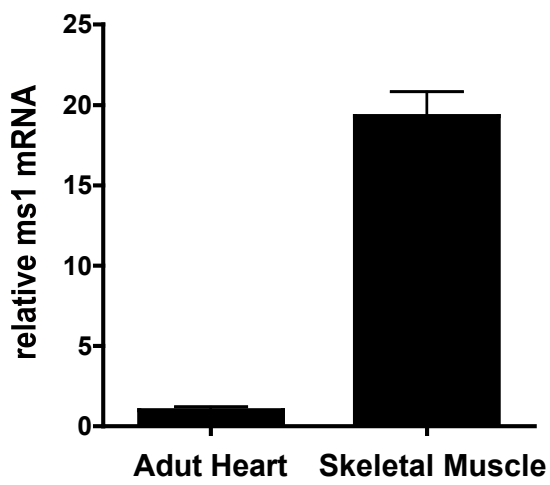
The aim of this chapter was to investigate the transcriptional regulation of *msl* during myogenic differentiation using the C2C12 and H9c2 myoblast cells lines as established model systems. Both of these cell lines can differentiate in a co-ordinated manner from a proliferating myoblast to a differentiated multinucleated myotube, when cultured in the appropriate media. Upon differentiation the myotubes derived from both these cell lines exhibit skeletal muscle characteristics similar to those observed *in vivo*, and as such provide excellent *in vitro* model systems for the investigation of skeletal muscle differentiation and maturation.

6.2 Results

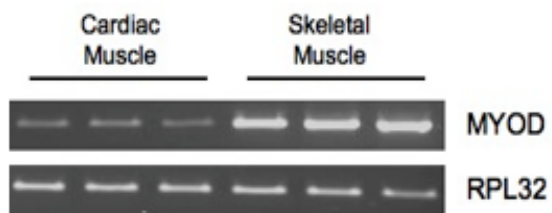
6.2.1 Quantitative analysis of relative *msl* mRNA in adult cardiac and skeletal muscle.

During its initial identification and characterisation, *msl* expression was measured, using semi-quantitative RT-PCR (Mahadeva et al, 2002) and Northern blot analysis (Kuwahara et al, 2007), to be restricted to cardiac and skeletal muscle. Due to the qualitative nature of these techniques we decided to examine the relative transcript abundance of *msl* in these tissues using a quantitative method. Using total RNA isolated from adult cardiac and skeletal muscle (isolated by Harin Mahadeva, samples checked for quality and stored at -80°C), quantitative real time RT-PCR of *msl* and TBP (internal control) using Taq Man® probes was performed as described in the methods (Figure 6.1A). Consistent with the previous reports, *msl* mRNA is significantly more abundant in adult skeletal muscle compared with cardiac muscle (Mahadeva et al, 2002). The present quantitative analysis demonstrates a significant 20-fold increase in the relative expression in skeletal muscle versus cardiac muscle (Figure 6.1A $p < 0.05$). To validate the myogenic status of the RNA samples examined, semi-quantitative RT-PCR of MyoD and RPL-32 (internal control) was performed as described in the methods (Figure 6.1C). MyoD, the classical MRF, is expressed at high levels in skeletal muscle with lower levels in cardiac muscle (Davis et al, 1987). The high abundance of MyoD transcript (12-fold increase relative to cardiac expression, $p < 0.05$) in our skeletal muscle RNA samples relative to cardiac muscle RNA samples validates the myogenic nature of these samples and is in agreement with published MyoD expression in these two adult tissue types.

A



B



C

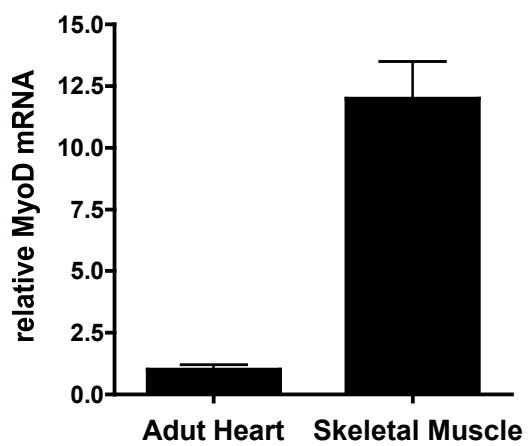


Figure 6.1. Quantitative real-time and semi-quantitative PCR analysis of *msl* and MyoD mRNA in adult striated muscle tissue. Total RNA was isolated from adult skeletal and cardiac tissue. Total RNA was subjected to reverse transcription followed by quantitative (A) and semi-quantitative (B, C) PCR with *msl* (A), MyoD (B, C), TBP and RPL32 specific primers. Relative expression in the adult heart was arbitrarily set at 1. Values are represented as means \pm SEM of at least three different experimental samples.

6.2.2 Analysis of *msl* mRNA during myogenic differentiation

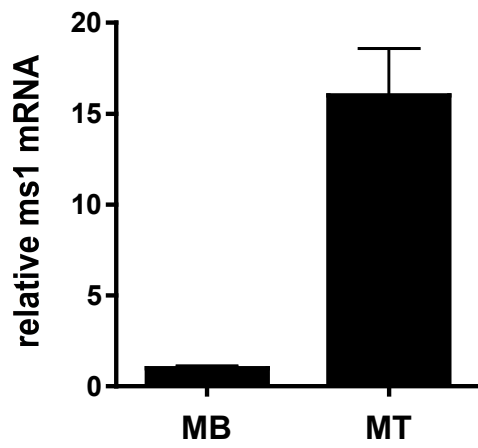
Previous work by others and us has demonstrated that *msl* mRNA is up regulated during myogenesis in both vertebrate (Kuwahara et al, 2005; Kuwahara et al, 2007) and invertebrate model organisms (Mahadeva et al, Unpublished findings). In order to examine *msl* expression during myogenesis *in vitro*, cultured C2C12 and H9C2 myoblasts were utilised. C2C12 is a myoblast cell line established from the leg muscle of the C3H mouse (Yaffe & Saxel, 1977). These myoblasts remain proliferative in the presence of high serum concentrations, with serum depletion stimulating the myoblasts to differentiate and fuse to form multinucleated myotubes (Blau et al, 1985). In order to evaluate the suitability of C2C12 as a model for the myogenic *msl* up-regulation, quantitative SYBR® Green based real time RT-PCR of *msl* and E1alpha (internal control) was conducted using 1µg of total RNA isolated from proliferating C2C12 (Myoblast; MB) and terminally differentiated C2C12 cells (Myotube; MT, represent 3 days in differentiation media), as described in the methods. There was a significant increase in *msl* transcript (approximately 15-fold) in terminally differentiated C2C12 myotubes compared to proliferating C2C12 myoblasts (Figure 6.2A) thus validating the use of C2C12 as a model.

This data demonstrated a differentiation dependent mechanism for *msl* transcriptional up-regulation during myogenesis *in vitro*. However, the temporal expression profile of a specific gene during myogenic differentiation can also give an exquisite insight into both the function of the gene and the regulatory processes governing its expression (Blais & Dynlacht, 2005; Delgado et al, 2003). I therefore proceeded to measure the temporal expression profile of *msl* during the graded differentiation of C2C12 myoblasts over a three-day period. This time period has been empirically shown to be sufficient for the complete terminal differentiation of C2C12 myoblasts, which is associated with the temporal up-regulation of the myogenic regulatory factors including MyoD and Myogenin. Quantitative SYBR® Green based real time RT-PCR of *msl* and E1α (internal control) using RNA isolated on consecutive days during differentiation demonstrated that *msl* transcript is significantly induced within the first day of differentiation (Figure 6.2C) (10-fold, $P < 0.05$). We then observed a maintained increase

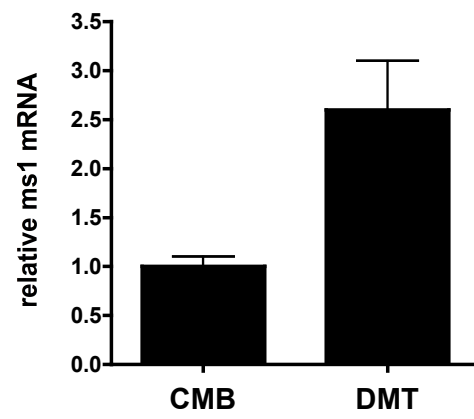
in expression over the subsequent two days (Figure 6.2C). This finding demonstrates that *msl* is an early wave myogenic transcript (Tomczak et al, 2004) in C2C12 differentiation, with this having implications for its role and its regulation during myogenesis.

To demonstrate this up-regulation of *msl* is ubiquitous to myogenesis and not a specific effect observed only in C2C12, the myogenic expression of *msl* was examined in an alternative cell line, H9c2. H9c2 is a clonal cardiac cell line derived from the embryonic rat heart (Menard et al, 1999). Although cardiac in origin, confluent H9c2 myoblasts have the ability to trans-differentiate into skeletal muscle myotubes. When confluent, cultivating the myoblasts in low serum leads to the formation of multinucleated myotubes that express the myogenic markers MyoD, Myogenin and Troponin T (Menard et al, 1999). These myotubes also display a very slow activating and inactivating L-type Ca^{2+} current which resembles those recorded in skeletal muscle myotubes, and other differentiated skeletal muscle cell lines and primary cultures. We proceeded to extract total RNA from confluent H9c2 myoblasts and differentiated myotubes as described in the methods. Semi-quantitative RT-PCR of *α Is* and RPL-32 (internal control) was performed with a significant up-regulation (20-fold $p < 0.05$) of *α Is* mRNA observed in the differentiated myotubes (Figure 6.2D). This level of up-regulation is in agreement with other studies in which *α Is* expression has been used as the phenotypic marker for H9c2 myogenic differentiation (Menard et al, 1999). We then proceeded to execute quantitative real time RT-PCR of *msl* and TBP (internal control) using Taq Man® probes on total RNA derived from these H9c2 differentiation samples. Myogenic differentiation was associated with a significant increase in *msl* expression (2.6 fold, $p < 0.05$). Although this up-regulation was not as dramatic compared to the C2C12 model, these findings demonstrate that *msl* is sensitive to myogenic differentiation with up-regulation observed in two independent models of *in vitro* myogenic differentiation.

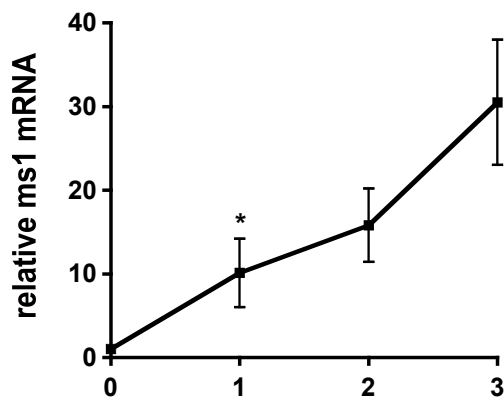
A



B



C



D

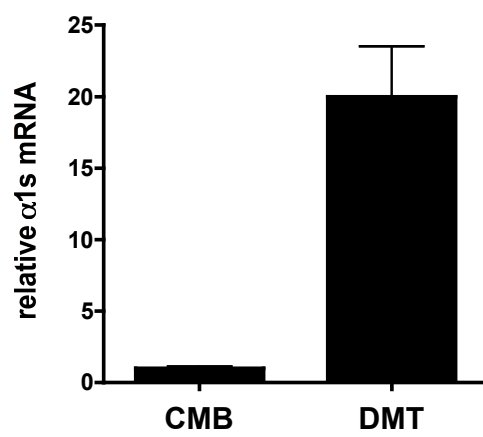
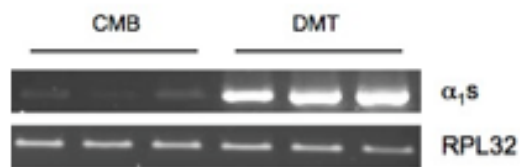


Figure 6.2. Quantitative real-time and semi-quantitative PCR analysis of *msl* mRNA expression during myogenic differentiation *in vitro*. Total RNA was isolated from C2C12 sub-confluent myoblasts (MB) and myotubes (MT: 3 days post differentiation)(A). Total RNA was also isolated from C2C12 cells during myogenic differentiation (Day 0 to day 3) (C). RNA was subjected to reverse transcription followed by quantitative PCR with mouse *msl* and EF1 α specific primers. Expression level at day 0 (MB) was arbitrarily set at 1. Total RNA was also isolated from confluent (CMB) and differentiated (DMT) H9c cells and subjected to quantitative real-time (B) and semi-quantitative (D) PCR with rat *msl* (B), α 1s (D), TBP and RPL32 specific primers. Expression level in H9c2 myoblasts (CMB) was arbitrarily set at 1. Values are represented as means \pm SEM of at least three different experiments.

6.2.3 Identification and myogenic specificity of the *msl* promoter.

The consistent up-regulation of *msl* during myogenic differentiation both *in vitro* (present C2C12 and H9c2 data) and *in vivo* (Kuwahara et al, 2005; Kuwahara et al, 2007) suggests its expression might be targeted by differentiation promoting transcription factors. During myogenesis, differentiation is under the control of E-Box binding myogenic basic helix-loop-helix (MyoD, Myogenin, Mrf4, Myf5) and MADS domain Mef2 proteins (Brand-Saberi, 2005), with most myogenic genes containing binding motifs for these factors in their promoters and associated regulatory domains (Lassar et al, 1989; Li & Capetanaki, 1994). Such motifs within the *msl* promoter would represent ideal candidates as putative differentiation dependent myogenic regulatory sites.

With this in mind, it was interesting to observe that our comparative *in silico* regulatory sequence analysis (Chapter 3) identified two putative regulatory domains, UP1 and PP, with potential myogenic activity. These two domains, which are both located within the proximal 1.6kbp 5' flanking sequence, are enriched with conserved myogenic binding sites including three E-Box and two Mef2 motifs (Chapter 3). Data presented in Chapter 4 demonstrated that the UP1 domain, which contains E-Box -1556/-1550 and Mef2 -1489/-1477 has minimal cardiogenic activity. One could thus speculate they serve to function in another cellular context and it was hypothesised they may contribute to the myogenic sensitivity of the *msl* promoter (Chapter 4). In support of this, Eric Olson's group have demonstrated that the proximal 1.6kbp 5' flanking sequence is able to direct expression of a lacZ reporter gene in the skeletal muscle of adult transgenic reporter mice (Kuwahara et al, 2007). On the basis of this finding, and the data presented in Chapters 3 and 4, it is reasonable to hypothesise that the proximal 1.6kbp 5' flanking sequence contained sufficient *cis* information to drive myogenic differentiation dependent up-regulation of *msl*.

To validate this hypothesis, the intrinsic myogenic responsiveness of the P-1585/+60 luciferase reporter construct was examined, as previously described in Chapter 4. This reporter, which encompasses the UP1 and PP domains, was transiently transfected into

the myogenic C2C12 and H9c2 cell lines, and the non myogenic NIH 3T3 and COS-7 cell lines, as described in the methods. In support of our hypothesis, the reporter was approximately four times ($p < 0.05$) more active in the myogenic C2C12 and H9c2 cells than in the non-myogenic NIH 3T3 mouse fibroblasts and COS-7 monkey kidney cells (Figure 6.3). This data supports our initial hypothesis and suggests that there is sufficient myogenic *cis* information encompassed within this promoter reporter to drive myogenic cell specific activity, at least in the context of our *in vitro* myogenic systems.

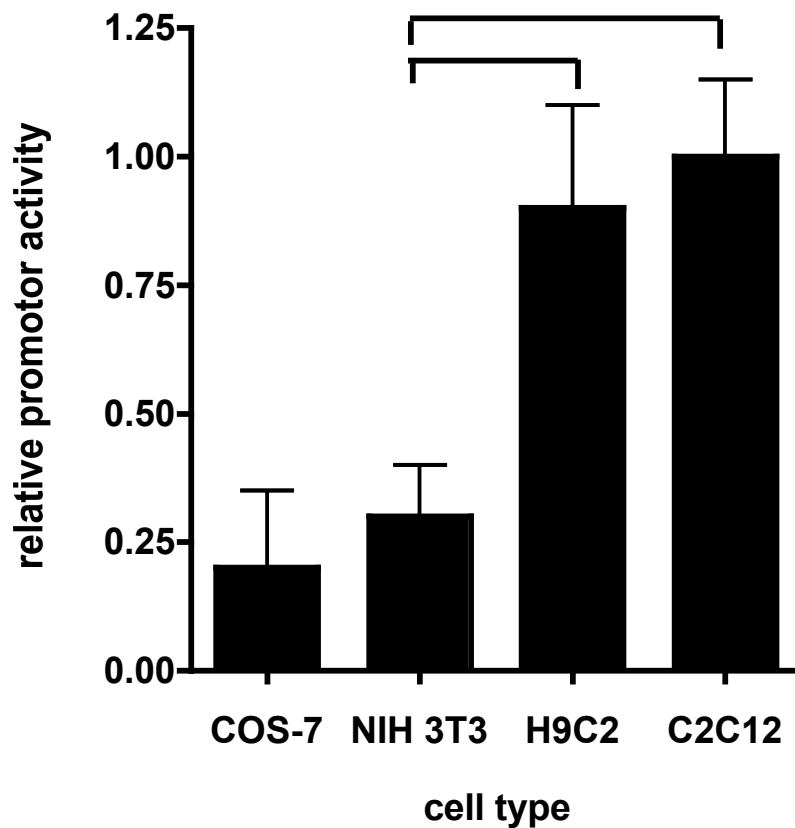
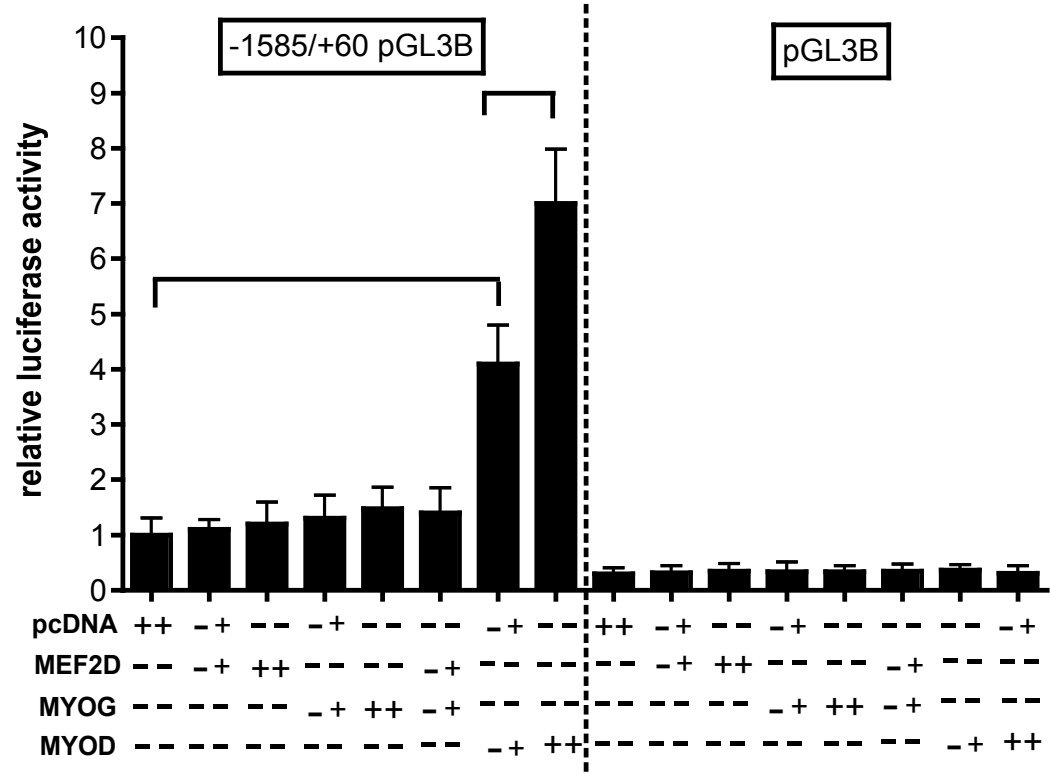


Figure 6.3 Analysis of *msl* promoter activity in myogenic and non myogenic cell lines. Subconfluent COS-7, NIH3T3, H9c2 and C2C12 myoblasts were transiently transfected with the -1585/+60 luciferase promoter reporter construct. Promoter activity relative to pGL3-B alone was determined in each cell line 48 hours after transfection. The relative activity of the promoter (vs pGL3-B) in each cell line was then determined with relative activity in C2C12 assigned arbitrary value of 1. The results are expressed as mean \pm SEM of at least three separate transfections in triplicate for the reporter construct. Statistically significant differences compared to relative activity in NIH3T3 are indicated by tailed bars, $p < 0.05$.

6.2.4 Identification of myogenic factors that can modulate the *msl* promoter

Since conserved E-Box and Mef2 binding motifs are located within the UP1 and PP domains, we decided to test the intrinsic sensitivity of the promoter to the ectopic expression of the cognate myogenic binding proteins. The promoter reporter construct, P-1585/+60, was transiently transfected into C2C12 myoblasts in the presence or absence of the specific MRFs (MyoD and Myogenin) and Mef2D, the prominent skeletal muscle MEF2 isoform (Ohkawa et al, 2006), as described in the methods. MyoD and Myogenin both bind myogenic E-Boxes and mediate early and late myogenic specific gene expression respectively. The pEMSV-MyoD and pEMSV-Myogenin over-expression plasmids have been demonstrated to drive robust ectopic expression of MyoD and Myogenin in myoblast cell lines (Ohkawa et al, 2006). MyoD was found to be a strong activator of *msl* promoter activity, activating P-1585/+60 in a dose-dependent manner: ~4-fold (0.5ug expression vector, $p<0.05$) and ~7-fold (1.0ug expression vector $p<0.05$) to that of control (1.0ug pcDNA only) (Figure 6.4A). In contrast, Myogenin alone or in combination with Mef2D, an interaction previously reported to synergistically activate other myogenic promoters, did not significantly activate the *msl* promoter. Mef2D alone was also insufficient in this regard. To validate the observed effects were specific to the *msl* promoter, all of the co-transfections were done using the empty pGL3-B reporter vector (Figure 6.4A). The lack of non-specific activation of the empty pGL3-B reporter suggests that the observed effects described here are specific to the inserted -1585/+60 genomic promoter sequence.

A



B

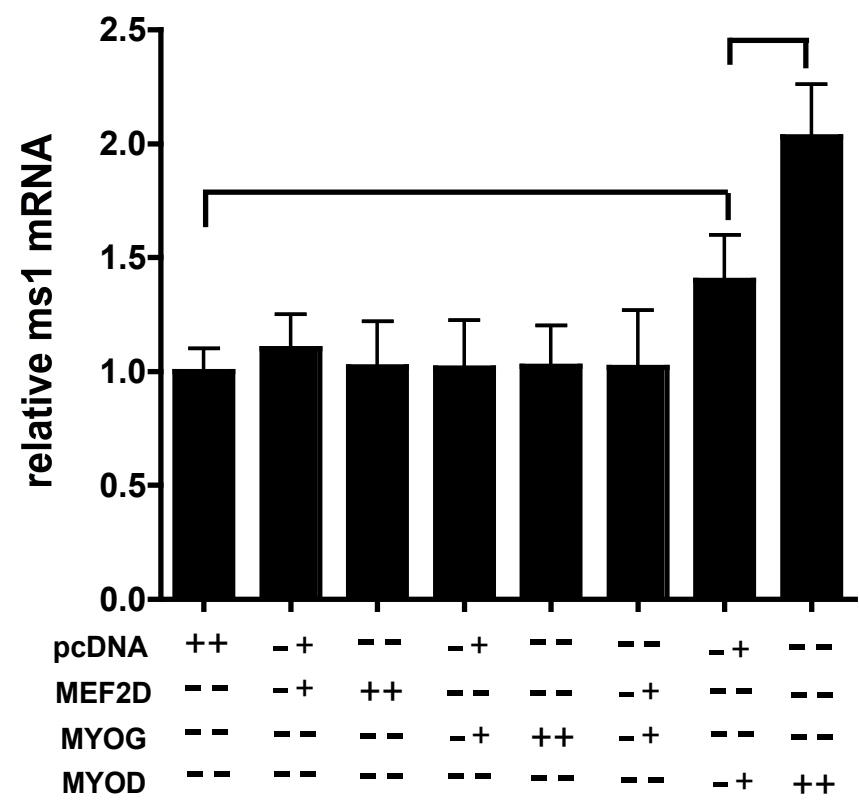


Figure 6.4. Plasmids (+; 0.3µg, ++0.6µg) expressing Mef2D, myogenin and MyoD were co-transfected in combination with the *msl* promoter reporter construct (P-1585/+60) into C2C12 myoblasts. Luciferase activity in cells transfected with pcDNA and *msl* promoter reporter was arbitrarily set at 1 (A). Subconfluent H9c2 myoblasts were transiently transfected with the MRF overexpression plasmids (B) (+; 0.5µg, ++1.0µg). 48 hours post transfection (cells were 80% confluent), total RNA was isolated, reverse transcribed and expression levels of TBP and *msl* were determined by real time PCR. Statistically significant differences are indicated by tailed bars, $p < 0.05$.

To determine whether this sensitivity would manifest at the endogenous *msl* promoter, the MRF and MEF2 proteins were ectopically expressed in H9c2 myoblasts, in dose dependent combinations as executed in the reporter co-transfection assay. The expression level of endogenous *msl* mRNA in both control (empty pcDNA vector alone) and MRF/Mef2D transfected cells was determined by quantitative real time RT-PCR using Taq Man® probes as described in the methods. MyoD over-expression significantly increased *msl* mRNA levels in a dose dependent manner 1.5 and 2-fold respectively (Figure 6.4B, $p < 0.05$). In agreement with the reporter co-transfection data, Myogenin and Mef2D alone, or in combination, had no effect on endogenous *msl* mRNA levels.

Collectively, this data suggests that MyoD is able to target and activate the endogenous *msl* promoter in agreement with the observed sensitivity of the exogenous promoter reporter. This data demonstrates that (at least in our *in vitro* myogenic context) MyoD is the primary myogenic activator of the endogenous *msl* promoter. However, these data do not exclude a possible role for Mef2D and Myogenin in regulating *msl* expression, and we suspect these factors may act in another myogenic environment, possibly during the later stages of myogenesis.

6.2.5 Site directed mutagenesis of the *msl* promoter

MyoD dose-dependently target the *msl* promoter (*in vitro* and *in vivo*), which suggests that MyoD might directly bind to the specific sites on the promoter to mediate its activation. MyoD activates target promoters via heterodimerisation with ubiquitous E2A proteins (E12, E47, E25)(Murre et al, 1989), which allows a stable DNA binding complex to bind the E-Box sequence (consensus sequence, CANNTG). In order to examine the contribution of the three conserved E-Box's (in UP1 and PP) in mediating MyoD sensitivity, these sites were deleted using site directed mutagenesis (as described in chapter 4 and methods).

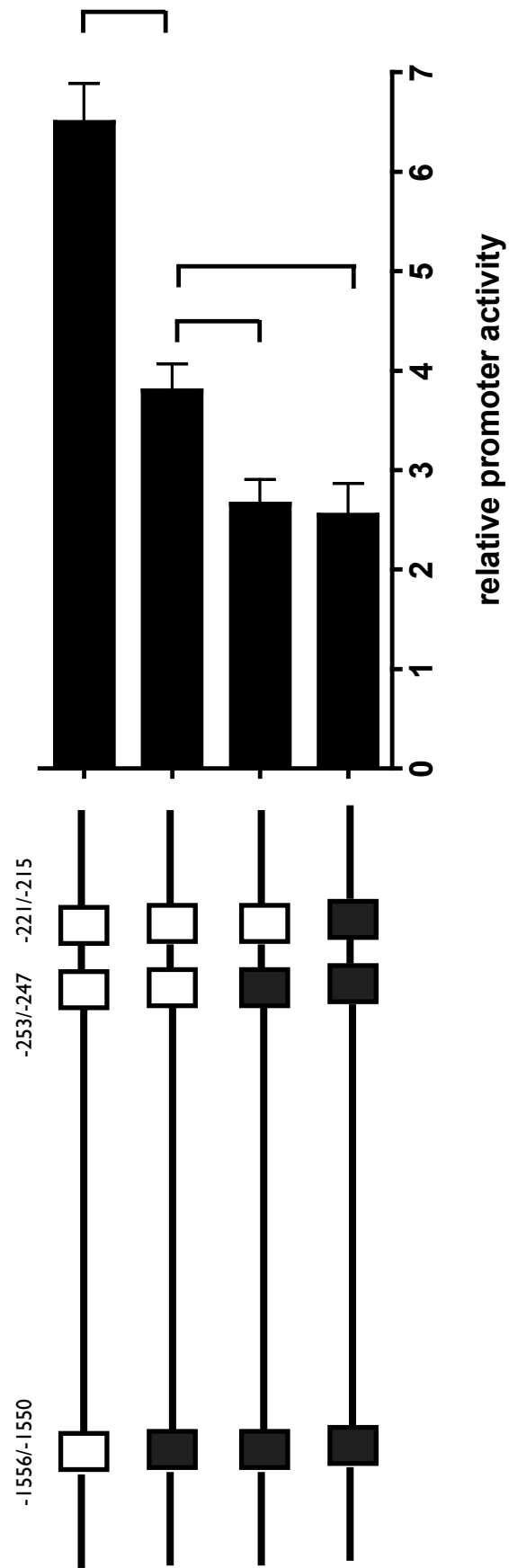


Figure.6.4. *MsI* promoter sensitivity to exogenous MyoD with targeted mutations of putative MyoD binding E-Box sequences. The E-Box sequences 1, 2 and 3 (-1556/-1550, -253/-247 and -221/-215) contained within the wild type P-1585/+60 construct were subjected to site directed mutagenesis. The subsequent single, double and triple E-Box mutant constructs were transiently co-transfected with MyoD into subconfluent C2C12 myoblasts for luciferase assays that were harvested 48 hours later, when the cells were 80% confluent. Luciferase activity, representing promoter sensitivity to ectopic MyoD expression (fold activation vs pcDNA), was inhibited by approximately 50% in the single E1 mutant with the double E1/E2 mutant resulting in a 67% reduction in activity with respect to wild type promoter sensitivity. The results are expressed as mean \pm SEM of at least three different experiments, in triplicate for each construct. Statistically significant differences compared to the wild type construct are indicated by tailed bars, $p < 0.05$.

The generated reporter constructs containing the mutated sequences for the three E-Boxes [in singular and in combination (Figure 6.5)], were co-transfected with the MyoD expression plasmid (1µg) into C2C12 myoblasts, as described in the methods. The wild type promoter was activated 7-fold by MyoD compared to control co-transfection (pcDNA only). Mutating E1 (-1556/-1550) reduced promoter sensitivity to ectopic MyoD expression by 50% (from 7-fold activation to 3.5-fold, $p<0.05$, Figure 6.5), demonstrating this E-Box is important for promoter sensitivity to MyoD. The additional mutation of E2 (-253/-247) further attenuated promoter to 2.5 fold activation compared to pcDNA control (approximate 67% decrease in sensitivity $p<0.05$). The combined triple E-Box mutant (mutating E3, -221/-215) did not result in further decrease in promoter sensitivity to MyoD with respect to the double mutant. This data suggests that in the *in vitro* reporter context both E1 and E2 (but possibly not E3) are sufficient for MyoD activation of the *msl* promoter. It also suggests the presence of other unidentified motifs within this reporter that are capable of mediating sensitivity to MyoD, possibly via direct or indirect mechanisms, which will be discussed.

5.2.6 *In vitro* binding of myogenic proteins to E1 and E2

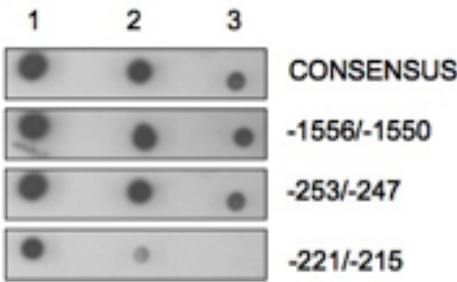
To further characterise the biological importance of the E1, E2 and E3 sequences, we performed electro-mobility shift assays (EMSA) to demonstrate the *in vitro* myogenic binding specificity of these motifs. Specific oligonucleotides (Table 10, Chapter 2) containing the E-Box elements present in UP1 and PP, E1, E2 and E3 were synthesised, annealed and end labelled with digoxigen (DIG) to generate the specific probes for the EMSA analysis. The efficiency of the labelling reaction was determined via a spot blot method as specified in the methods. On the basis of the relative labelling of each probe (Figure 6.6A), equally labelled amounts of the E-Box probes were incubated with whole cell protein extracts from C2C12 myoblasts and a molar excess (X200) of complimentary cold unlabelled probe. As shown in Figure 6.6B, incubation of E1, E2 and E3 labelled probes with myogenic whole cell extracts resulted in the appearance of single DNA-protein band shift. Co-incubating this reaction with a 200 fold molar excess of unlabelled complimentary probe ablated the appearance of the shifted band, demonstrating the DNA-protein band shift is the product of specific protein/DNA-

sequence specific interaction. As a positive control the consensus MyoD binding site from the ACC β promoter was also incubated with the myogenic protein cell extract. This sequence has previously been shown to generate a specific single shifted band in EMSA analysis using myogenic protein extracts. This is as a consequence of specific MyoD binding to the consensus sequence within this probe. In our experimental context we also observed a single specific band shift in agreement with the published finding (Figure 6.6B).

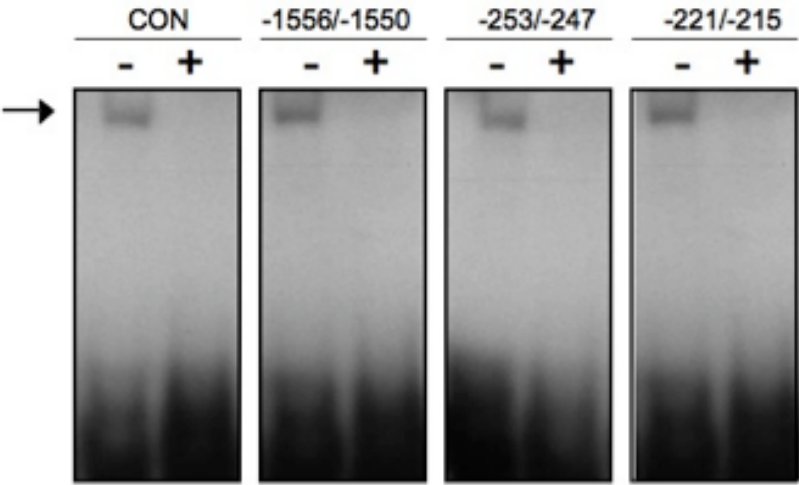
It was of particular interest to observe that all three E-Box probes generated a robust single band shift (represented by the arrow Figure 6.6B), all of which exhibited similar migration to each other and to the characterised consensus MyoD binding E-Box probe. This is suggestive of a common protein complex associating with all of our probes, and based on the myogenic MyoD binding capacity of the consensus E-Box we speculated this to be a MyoD containing complex.

To support this hypothesis, the E1, E2 and E3 labelled probes were incubated with C2C12 myoblast whole cell protein extracts and a 200 fold molar excess of unlabelled consensus MyoD binding E-Box probe. Interestingly the band shifts observed with the E1 and E2 probes was completely ablated by the cold consensus probe (Figure 6.6C). This suggests that MyoD is capable of binding to E1 and E2 *in vitro*. In contrast, E3 could not be competed with unlabelled consensus suggesting other E-Box (or non-EBox) binding proteins expressed in C2C12 myoblasts are binding E3 *in vitro*. These findings are in agreement with our mutagenesis analysis and support the hypothesis that MyoD targeting and subsequent activation of the *msl* promoter occurs at E1 and E2, but not E3, at least in the present *in vitro* experimental system.

A



B



C

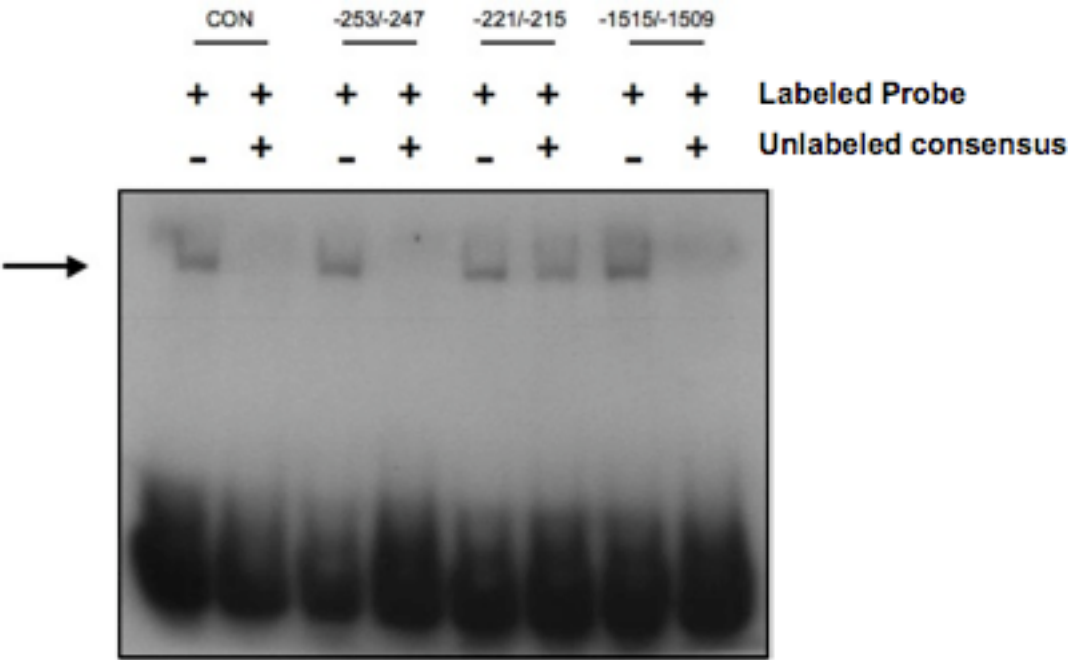


Figure 6.6. EMSA analysis of E1, E2 and E3 in the *msl* promoter. DIG-labelled oligonucleotide probes for the MyoD E-Box binding consensus, E1, E2 and E3 binding sites were blotted and visualised for the relative determination of labelling efficiency. Equally labelled amounts of probe were then incubated with whole cell protein extracts made from sub-confluent C2C12 myoblasts (B, C). Competition experiments were performed using a 200-fold excess of unlabelled self (B) to test for specificity and a 200-fold excess of unlabelled MyoD E-Box consensus probe (C). Arrows indicate the resulting band-shifts.

6.2.7 Direct binding of MyoD to the endogenous UP1 and PP domains during myogenic differentiation.

The data presented so far suggests that within the current *in vitro* myogenic context, MyoD is the primary myogenic activator of the *msl* promoter, with the majority of this sensitivity mediated by the UP1 and PP domain E-Boxes, E1 and E2 respectively. To definitively confirm this [in collaboration with Prof Tony Imbalzano (UMASS)], chromatin immunoprecipitation (ChIP) was used to determine whether MyoD is physically recruited to the UP1 and PP domains *in vivo*, and to examine the dynamics of this recruitment during C2C12 myogenic differentiation.

Using primers specifically designed to amplify E1, E2 and E3 within the UP1 and PP domains (Table 11, Chapter 2, Figure 6.7A), quantitative SYBR® Green based real time PCR was performed on formaldehyde-crosslinked, sheared chromatin (isolated by Dr Caroline Dacwag, UMASS) during C2C12 differentiation, which was immunoprecipitated with MyoD and IgG specific antibodies. The relative enrichment of MyoD at these two domains was quantified as described in the methods.

As shown in Figure 6.7B, MyoD appears to be constitutively bound at the E2/E3 genomic domain during the whole differentiation process, with no significant difference in relative recruitment occurring upon *msl* up-regulation. This was interesting, firstly because it confirmed the binding of MyoD to the E2/E3 domain *in vivo*, and secondly because this binding appears to precede the transcriptional induction of the *msl* transcript. Due to the resolution of the ChIP derived PCR (sonicated DNA fragments are approximately 300-500bp in length) it was not possible to conclusively demonstrate that MyoD binding within this E2/E3 domain is specific to E-Box E2, or E3. However based on the *in vitro* data derived from the mutagenesis and EMSA analysis, it is likely that the MyoD binding within this domain is occurring at E-Box E2.

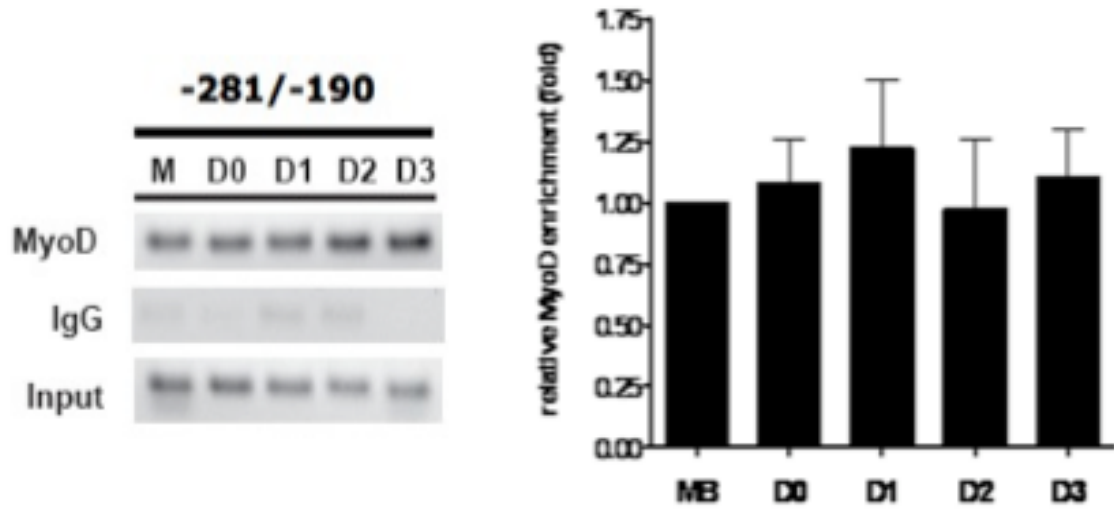
In contrast to MyoD recruitment at the E2/E3 domain, MyoD is not recruited to the E1 domain until day 1 of the differentiation process (Figure 6.7C). A 5-fold enrichment in relative binding of MyoD occurs at day 1 compared to day 0 (Figure 6.7C, $p < 0.05$) and

this can be qualitatively visualised in the SYBR Green stained agarose gel. This binding is then maintained over the subsequent two days and interestingly coincides with the transcriptional up-regulation in *msl* (Figure 6.2C). This observation suggests that binding of MyoD at E1 may be required for the differentiation dependent up-regulation in *msl* transcript, and we therefore speculate this E-Box and the encompassing UP1 domain may represent a differentiation dependent myogenic enhancer. In conclusion the temporal dynamics of MyoD binding during myogenic differentiation clearly demonstrates the differential roles played by E-Boxes E1 and E2 in mediating myogenic differentiation dependent activation of *msl*. This insight could not be gained using only *in vitro* techniques and emphasises the utility of ChIP in this context.

A



B



C

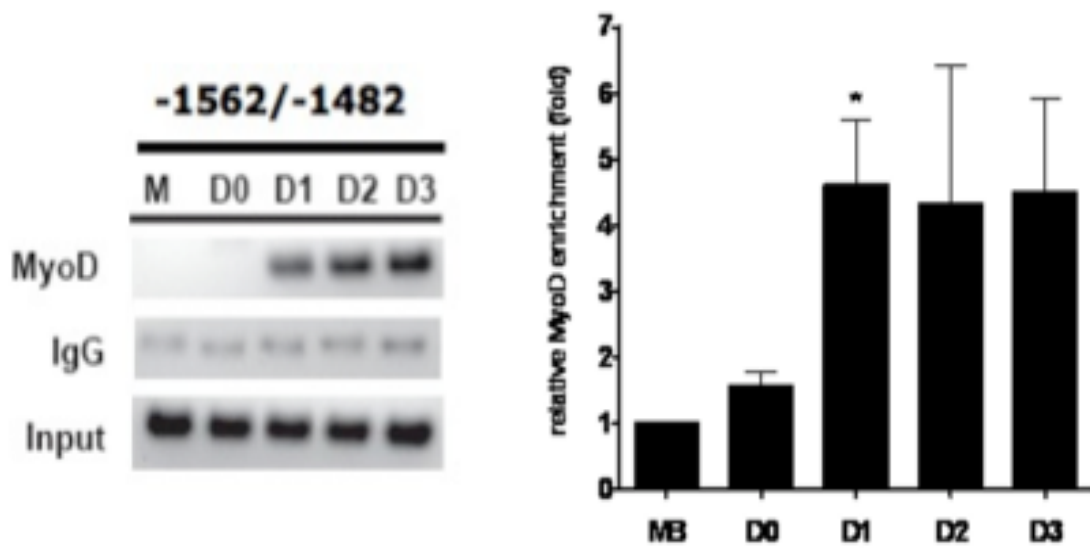


Figure 6.7. Chromatin immunoprecipitation (ChIP) assay performed on differentiating C2C12 cells to evaluate *in vivo* MyoD binding within the *msl* promoter. (A) Schematic map of amplified DNA fragments and primer locations encompassing E1, E2 and E3. TSS position is also illustrated by an arrow. Samples analysed included proliferating sub-confluent myoblasts (MB), confluent myoblasts harvested prior to the induction of differentiation (D0), myoblasts subjected to differentiation conditions for 24 hours (D1), and differentiating myotubes at 48 (D2) and 72 (D3) hours post differentiation. Quantitative real time PCR was performed on isolated DNA using the primers designated in (A) amplifying the proximal (B) and distal (C) E-Box sequences (E2/E3 and E1 respectively). Amplification was quantified and normalised to the input of each sample. The results are expressed as mean \pm SEM of at least three different ChIPs. Statistically significant differences in fold enrichment are indicated by * $P < 0.05$. Representative PCR reactions were stopped in the linear amplification range and run on agarose gel for visualisation.

6.2.8 Temporal dynamics of Histone Deacetylase (HDAC) enzyme recruitment to the endogenous UP1 and PP domains during myogenic differentiation.

The binding of MyoD at E2/E3 prior to gene activation suggested that this MyoD is recruiting repressive complexes to the *msl* promoter, where it inhibits transcriptional induction. This functional mechanism of MyoD action is common and indeed numerous studies have demonstrated that MyoD is able to target promoters for repression via the specific recruitment of histone deacetylase enzymes, in particular the Class I family member, HDAC1 and HDAC2. The present data therefore suggests such enzymes may be recruited to the *msl* promoter and repress transcriptional induction prior to the differentiation stimulus. It is of interest that data presented in Chapter 4 clearly demonstrated that in H9c2 myoblasts, HDAC inhibition via Trichostatin A (TSA), resulted in transcriptional induction in *msl*. This supports the hypothesis by demonstrating that the *msl* promoter is repressed by HDAC enzyme activity in a myoblast cellular environment.

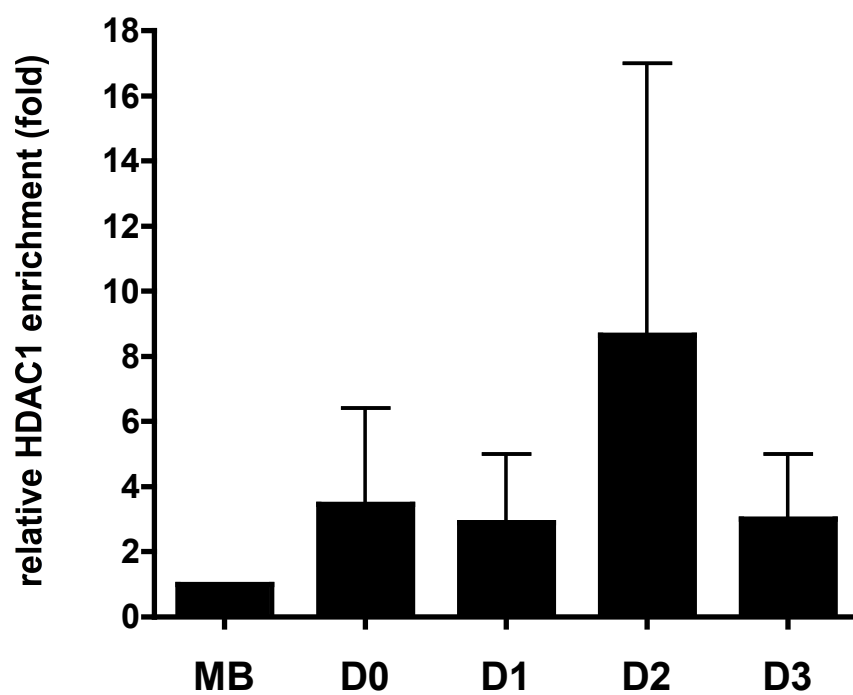
To validate this, quantitative SYBR® Green based real time PCR was performed on formaldehyde-crosslinked, sheared chromatin isolated during C2C12 differentiation, which was immunoprecipitated with HDAC1, HDAC2 and IgG specific antibodies. The relative enrichment of HDAC1 and HDAC2 at these two domains was quantified as described in the methods. ChIP analysis demonstrated the presence of HDAC1 and HDAC2 at the E2/E3 domains but there was no significant difference in temporal recruitment of HDAC1 (Figure 8A) or HDAC2 (Figure 8C) to the proximal E2/E3 domain during myogenic differentiation.

One general principle that has emerged over recent years is that transcription factor mediated activation of gene expression correlates with HDAC removal from the genomic promoter domains of interest and concomitant recruitment of histone acetyltransferase (HAT) enzymes. We therefore predicted that the temporal MyoD recruitment at the UP1 domain and associated transcriptional up-regulation in *msl* might correlate with relative enrichment of HDAC1 and HDAC2 at the UP1 domain. However somewhat surprisingly, considering recent studies that demonstrate such

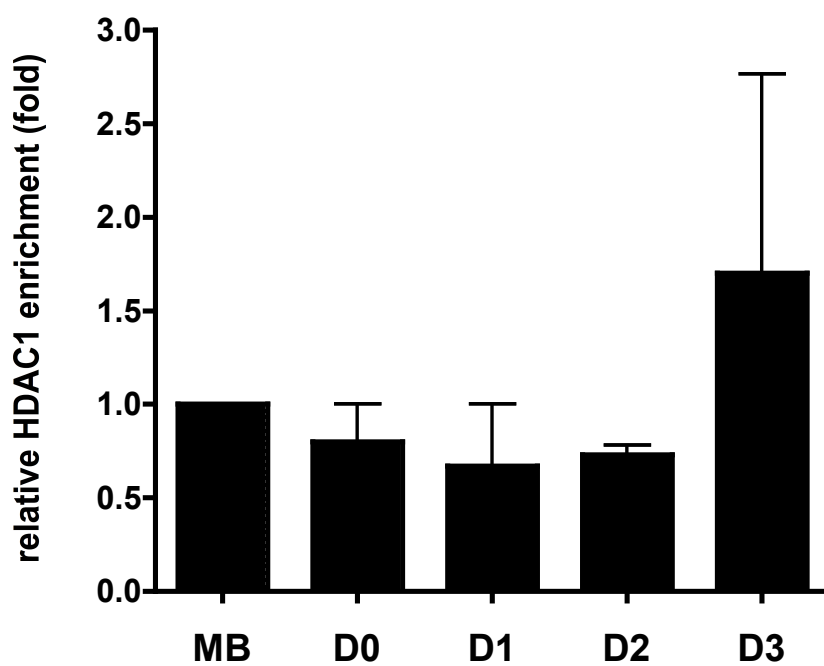
MyoD/HDAC1/2 interactions, we did not observe any significant difference in HDAC1/2 recruitment to the UP1 domain during myogenic differentiation (Figure 8B and 8C). Interestingly we do observe a significant decrease in relative HDAC2 recruitment at the UP1 domain in confluent myoblasts (D0) compared with proliferating myoblasts (MB)(Figure 8D, $p < 0.05$). This decreased enrichment in HDAC2 may have implications with respect to the subsequent recruitment of MyoD and the associated transcriptional up-regulation at Day 1.

In summary we have demonstrated that Class I family members HDAC1 and HDAC2 are actively associated with the *msl* proximal and distal E-Box containing regulatory domains (PP and UP1). However during myogenic differentiation we did not observe any ablation in HDAC recruitment neither at the proximal or distal (UP1) domain. Interestingly we have demonstrated a significant difference in HDAC2 recruitment at the UP1 domain in proliferating versus confluent myoblasts, with this potentially having major implications for *msl* regulation as a whole.

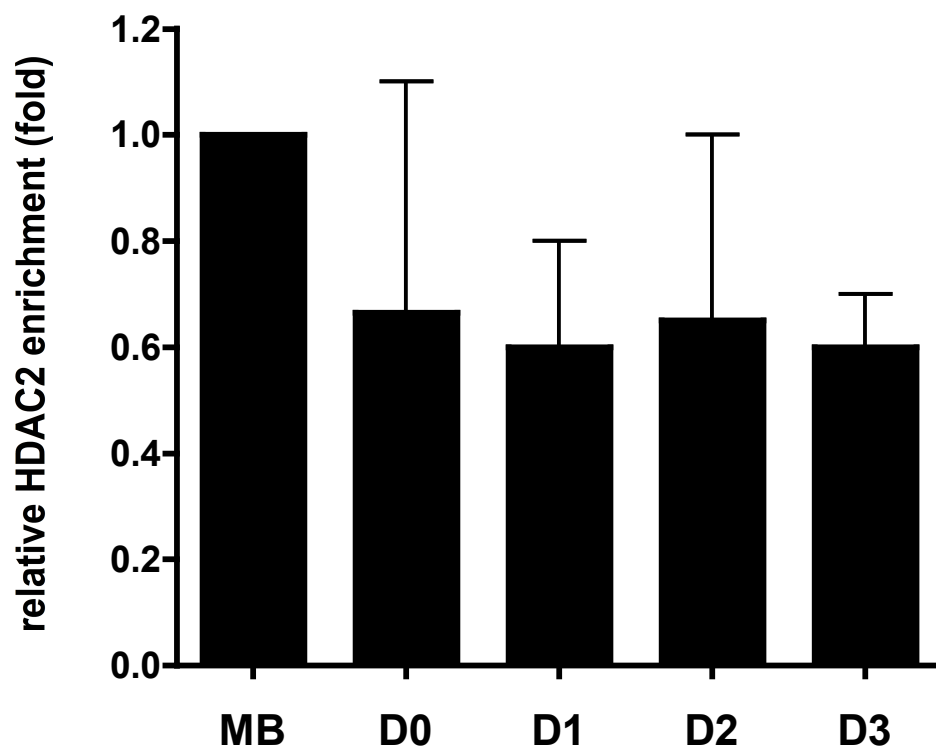
A



B



C



D

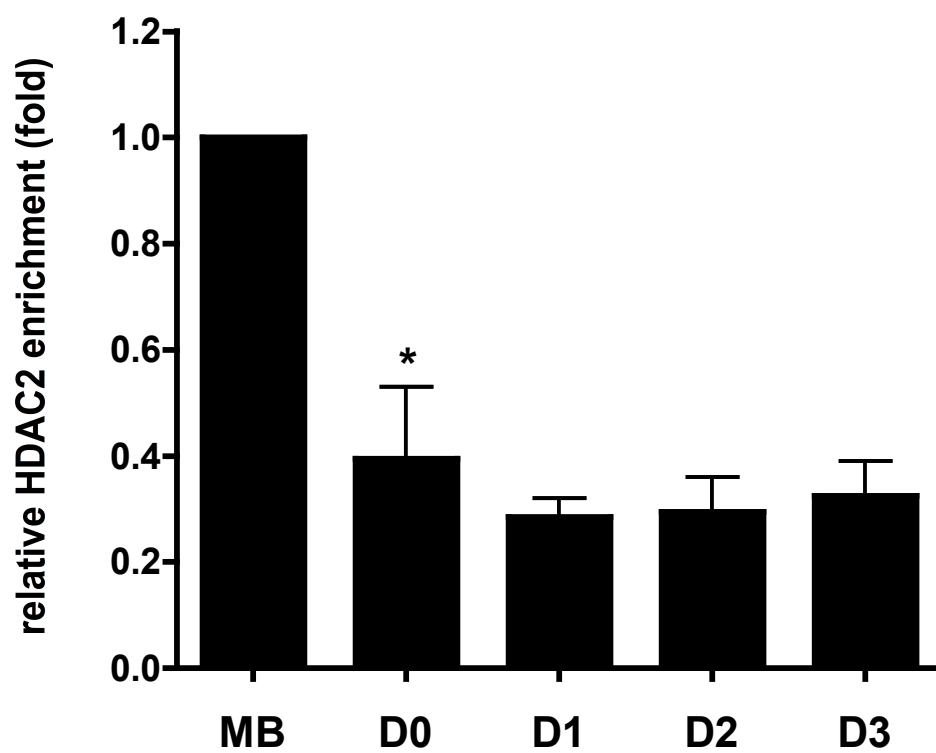


Figure 6.8. Chromatin immunoprecipitation (ChIP) assay performed on differentiating C2C12 cells to evaluate *in vivo* histone de-acetylase (HDAC) binding within the *msl* promoter. Quantitative real time PCR was performed on isolated DNA using the primers designated in (Figure 6.7A) amplifying the proximal (A,C) and distal (B, D) E-Box sequences (E2/E3 and E1 respectively) from HDAC1 (A, B) and -2 (C, D) immunoprecipitated chromatin. Amplification was quantified and normalised to the input of each sample. The results are expressed as mean \pm SEM of at least three different ChIPs. Statistically significant differences in fold enrichment are indicated by *P<0.05.

6.3 Discussion

The MS1-MRTF-SRF signalling axis has been implicated as a critical pathway in myogenesis, regulating myogenic differentiation, maturation and post-natal regeneration (Chapter 1). It is therefore reasonable to suspect that the mechanisms regulating the myogenic specific expression of *msl* will impact upon the MRTF-SRF transcriptional axis, and thus on myogenesis as a whole. In this chapter, we investigated the molecular mechanisms governing myogenic expression of *msl* using two established *in vitro* systems for myogenesis, specifically the C2C12 and H9c2 myogenic cell lines.

Prior to analysing *msl* expression *in vitro* we quantified the relative expression level of *msl* in adult cardiac and skeletal muscle, *in vivo*. In agreement with the previous studies we show that *msl* is expressed at higher levels in the adult skeletal muscle with respect to the heart. The exact reasons for this are not clear however this is suggestive that the MRTF-SRF axis is (in functional terms) more in demand in skeletal versus heart muscle. This is a good possibility because the MRTF-SRF pathway has been implicated in cardiogenesis, with its downstream genes central to the assembly of the actin cytoskeleton, sarcomere and contractile apparatus (Carson et al, 1996; Catala et al, 1995; Croissant et al, 1996; Vandromme et al, 1992; Wei et al, 1998). However within cardiac muscle, and not within skeletal muscle, Myocardin (the founding member of the myocardin related transcription factors) is also highly expressed. Myocardin is a potent co-activator of SRF, like the MRTFs. It is therefore likely that the MRTFs and Myocardin play redundant roles in regulating SRF activity in cardiac muscle. Increased *msl* expression in skeletal muscle would therefore support robust MRTF-SRF signalling and thus offer a compensatory mechanism for the lack of Myocardin protein.

As a first step towards understanding the transcriptional mechanisms governing *msl* expression in muscle differentiation, *msl* expression in differentiating C2C12 and H9c2 cells were examined. A robust transcriptional induction in *msl* expression was observed during the first day of differentiation indicating *msl* to be an early wave myogenic transcript (Delgado et al, 2003). This temporal expression profile has implications for the mechanisms governing its regulation and also gives insight into the potential

functionality of this early myogenic induction. It has previously been demonstrated that SRF can regulate myoblast fusion and differentiation as a consequence of regulating MyoD expression (Carnac et al, 1998). In addition, MyoD and SRF have been shown to physically interact and synergistically activate target myogenic promoters, with consensus SREs enriched in bona fide MyoD target promoters (Blais & Dynlacht, 2005). It is therefore imperative that SRF activity coincides with MyoD expression during myogenic differentiation and it is proposed here that this early myogenic expression of *msl* drives muscle specific activity of the MRTF-SRF axis, thereby coupling this pathway with MyoD expression and activity. It would be interesting to investigate the effect of perturbing *msl* expression during myogenic differentiation however this was not within the scope of the current PhD project.

The proximal 1.5kbp 5' flanking sequence, which is capable of driving *in vivo* skeletal muscle expression of a LacZ reporter (Kuwahara et al, 2007), was enriched with conserved MRF binding sites (Chapter 3) specifically clustered within two independent regulatory domains, UP1 and PP. It is of interest that multiple E Boxes (binding motifs for MRFs) forming independent clusters (like within the UP1 and PP domains) are often found on the myogenic regulatory regions, either promoters or enhancers, of muscle specific genes. The promoter reporter used here, which encompassed the UP1 and PP myogenic domains, showed specific activity within a myogenic cellular environment. Furthermore, the exogenous reporter and endogenous promoter showed specific sensitivity to the MRF, MyoD, and not Myogenin or MEF2D. This specific pattern of sensitivity suggests that at least within a myoblast cellular environment; MyoD represents the major myogenic activator of the *msl* promoter.

This above finding complements the temporal expression profile observed for *msl* during myogenic differentiation. This is because MyoD, and not Myogenin or Mef2D, is responsible for myogenic gene activation during the early stages of differentiation (Blais & Dynlacht, 2005; Buckingham et al, 2003), the time at which *msl* transcription is up regulated. Myogenin and Mef2D are themselves subsequently induced by MyoD. Imbalzano and colleagues (Ohkawa et al, 2006; Ohkawa et al, 2007) have recently demonstrated that late gene expression correlates with replacement of MyoD with

Myogenin at MyoD target promoters. This exchange is accompanied by the appearance of a muscle specific Mef2D isoform (Mef2D.1B), which along with the Brg1 ATPase collaborate to drive and maintain promoter specific transcription. These complexes also appear to replace MyoD binding at early myogenic genes (like *msl*) consolidating and maintaining their expression during the late stages of differentiation. When one considers the maintained high levels of *msl* expression during late myogenesis and in adult skeletal muscle, it would be interesting to determine whether myogenin or Mef2d can replace MyoD on the *msl* promoter and maintain constitutive expression. It is of interest that Mef2C binding to our MEF2 consensus identified here has been shown to mediate basal and stress inducible cardiac specific promoter activity both *in vitro* and *in vivo* (Kuwahara et al, 2007). It is plausible that within the differentiated myogenic environment this motif could also be functional. These findings pose interesting questions for further analysis..

It was speculated that the observed MyoD sensitivity was via specific targeting of MyoD to the conserved E-Boxes identified within the UP1 and PP regulatory domains. Indeed, using site directed mutagenesis and EMSA, it was demonstrated that E1 and E2, but not E3 appeared to be targeted by MyoD. However, it was somewhat surprising that in our mutagenesis reporter assays, the triple E-Box reporter mutant retained a significant sensitivity to ectopic MyoD expression. Firstly this may suggest the presence of other non-conserved E-Boxes within the reporter being targeted by MyoD or potentially MyoD is being recruited to the promoter in an E-Box independent fashion, possible via interaction with other DNA binding sequence specific transcription factors. Alternatively this effect may be as a result of a MyoD directed up-regulation of other myogenic transcription factors, which subsequently target and activate the *msl* promoter at other motifs. From this viewpoint it is interesting to consider the presence of a conserved SRE (Figure 3A) within the PP domain. We have already demonstrated that SRE binds this motif *in vivo* within a cardiogenic context (Chapter 4). Ectopic MyoD is able to up-regulate endogenous *msl* expression (Figure 4B), which would subsequently activate the MRTF-SRF signalling axis. It is conceivable that ectopic MyoD could, via endogenous *msl* expression, activate the SRF signalling axis which then (independent of MyoD binding to E1, E2 and E3 targets) activates the exogenous

promoter reporter via the conserved SRE. A similar principle may be true with respect to Mef2D which could also be induced by ectopic MyoD and potentially activate the reporter at the conserved MEF2 motif. It would be interesting to incorporate mutations within the SRE and MEF2 sites and determine if the level of sensitivity is still maintained in triple E-Box mutant.

Many muscle specific genes are activated at different times during the myogenic differentiation process. Numerous studies suggest that this differential expression of each target gene is a product of specific temporal binding of MyoD at distinct E-Boxes within the cis regulatory loci of the gene, which itself is coupled to chromatin modification and remodelling (McKinsey et al, 2002). We therefore used ChIP to measure *in vivo* binding of MyoD and the histone deacetylase enzymes HDAC1 and 2 at the UP1 and PP E-Box domains during C2C12 differentiation. This analysis demonstrated that MyoD is constitutively bound to the E2/E3 domain during differentiation, with binding present in myoblast at day 0, which precedes *msl* transcriptional induction. *Ms1* transcriptional induction at day 1 coincided with MyoD binding at E1, located within UP1. This data suggests that MyoD targeting at E1 may be required for transcriptional activation of *msl* and it is proposed here that UP1 with its constituent E1 represents a potential differentiation-dependent skeletal muscle specific enhancer. Future *in vivo msl* promoter driven LacZ reporter studies would allow us to definitively confirm this proposition.

What is the purpose of MyoD binding at the proximal E2/E3 domain and how does this contribute to the subsequent transcriptional induction? Interestingly myogenin, which has a remarkably similar temporal expression profile to *msl* during myogenesis, also has a similar promoter structure and associated temporal recruitment of MyoD (Salma et al, 2004), suggesting potential commonality in regulatory mechanisms. MyoD is targeted to the myogenin promoter prior to gene activation by a Pbx/Meis complex (Salma et al, 2004). Chromatin remodelling complexes are then targeted to the MyoD/Pbx/Meis DNA bound complex resulting in local chromatin acetylation and remodelling. This promotes the further association with the ATPase dependent chromatin remodelling complex (SWI/SNF) which then drives the opening of canonical

E-Boxes and Mef2 sites for factor binding, resulting in a stably bound activator complex that drives transcription. The presence of a fully conserved Pbx motif (Chapter 3) overlapping the transcriptional start site within the *msl* PP domain suggests that such a mechanism of induction may also occur at the *msl* promoter during differentiation.

Prior to activating transcription, MyoD associated with HDACs serves to mark myogenic genes for subsequent differentiation cues and thus activation (Mal & Harter, 2003; Mal et al, 2001). Therefore it is proposed here that MyoD binding to the *msl* promoter prior to differentiation primes the promoter in a ‘poised’ myogenic state. In this poised state MyoD-HDAC complexes compact the local chromatin environment thus preventing the potential association of MyoD and other activating factors with other motifs within the domain. However appropriate differentiation cues can stimulate MyoD to toggle between HDAC and HAT recruitment in addition to association with differentiation-specific myogenic factors (McKinsey et al, 2002). This dissociation of HDACs and associated recruitment of HATs stimulates local chromatin remodelling allowing the formation of active transcriptional complexes to target the specific promoter.

ChIP was therefore used to investigate potential differential HDAC recruitment at the *msl* UP1 and PP domains, although we did not observe any differential HDAC1/2 recruitment. This suggests that these HDACs may not be specifically modulated at the *msl* promoter during differentiation and raises the possibility of other HDACs being present. We did however demonstrate the temporal loss in HDAC2 recruitment at the UP1 domain during the proliferating myoblast to confluent myoblast transition. This is interesting because although myoblastic, confluent myoblasts (represent Day 0 in differentiation process) have already entered the differentiation process, where proliferating myoblasts have not. This specific loss of HDAC2 at UP1 is therefore likely to be of functional importance. We suspect this attenuation of HDAC2 activity at UP1 will result in increased acetylation at this domain thereby rendering the local chromatin more accessible to activating factors, such as MyoD. Indeed one could speculate this step is essential in allowing MyoD binding at Day 1 and therefore subsequent transcriptional induction.

Taking the present data together, a model is proposed whereby MyoD binding at the *msl* PP domain (possibly via E2 only) in myoblasts represses transcription via the specific recruitment of HDAC enzyme complexes (such as HDAC1 and 2). This PP domain binding is essentially priming the *msl* promoter, placing it in a poised state for sensing appropriate differentiation cues. Upon differentiation MyoD associates with HATs and SWI/SNFs, which subsequently causes remodelling of the local chromatin environment. This coupled to the decreased HDAC enrichment at UP1 allows MyoD to bind E1 within this domain (Day 1). Further targeting of HATs and specifically SWI/SNF complexes (at E1 and PP) can facilitate the binding of TBP and other factors involved in polymerase II pre-initiation complex formation and promote transcriptional elongation (Brown et al, 1996; Corey et al, 2003; Imbalzano et al, 1994; Lomvardas & Thanos, 2001; Salma et al, 2004; Soutoglou & Talianidis, 2002). We propose that temporal targeting of MyoD at E1 is required in order to establish the optimum environment for Pol II action and robust transcription.

In conclusion, the work highlighted in this chapter demonstrated that temporal recruitment of MyoD (and HDACs) at specific myogenic motifs (E1 and E2 located within the PP and UP1 domains) is important in mediating the correct temporal expression profile of *msl* during myogenic differentiation. This data also implicates *msl* as a potentially key component of a MyoD generated feed-forward regulatory circuit, where factors induced by MyoD (like *msl*) feed-forward to regulate late MyoD activity (via SRF) at subsequent target genes, therefore acting to temporally coordinate the timing of gene expression during skeletal myogenesis. This MyoD-MS1-SRF feed-forward network would serve to consolidate and amplify the myogenic cascade. We believe this is the first piece of data to demonstrate a potential mechanism linking MyoD activity and SRF transcriptional signalling, with *msl* serving as the nodal point to integrate these two central myogenic regulatory networks.

The findings presented here also have implications for myogenic disease phenotypes. Studies have recently identified IGF-1 and IL-4, both central to post-natal muscular regenerative processes, as SRF downstream genes in the post-natal regenerative context

(Charvet et al, 2006). Therefore understanding the molecular mechanisms governing *msl* expression may allow us to identify and develop therapeutic strategies for the up-regulation of *msl* gene expression in the disease phenotype, which would facilitate regeneration via stimulation of SRF activity and resulting upregulation of IL-4 and IGF-1.

Chapter 7

General Discussion

Myocyte Stress 1 (Ms1/STARS) is a striated muscle restricted actin binding protein that critically modulates the MRTF/SRF signalling axis in a RhoA dependent fashion (Arai et al, 2002; Kuwahara et al, 2005; Kuwahara et al, 2007; Mahadeva et al, 2002). Considering the diverse regulatory role of this axis, *ms1* has also been implicated in many facets of striated muscle biology including cardiac and skeletal muscle development, differentiation, maturation and post-natal homeostatic function (Arai et al, 2002; Kuwahara et al, 2005; Kuwahara et al, 2007; Mahadeva et al, 2002). Most importantly to cardiac pathology, alteration in *ms1* expression has been shown to contribute to cardiac hypertrophy and dysfunction in both rodent and human models (Kuwahara et al, 2007). Despite these important physiological and pathological findings, nothing was known of the molecular mechanisms governing the transcriptional regulation of *ms1* at the onset of this thesis. Such an understanding would provide an important insight into *ms1* function and the regulatory nature of its associated normal and pathological states. Therefore the overall aim of this thesis was to identify, experimentally interrogate and functionally characterise these transcriptional regulatory processes.

As the first step in this process a computational comparative sequence analysis of the *ms1 Rattus norvegicus* genomic interval was performed (Chapter 3). The premise for this was the identification of highly conserved genomic domains and transcription factor binding sites (TFBS) as potential regulatory elements for experimental investigation. This analysis identified four highly conserved putative *cis* regulatory modules (PCRM) within the *ms1* 5'-flanking interval, designated PP, UP1, UP2 and UP3. Within each of these domains multiple transcription factor binding sites (TFBS), which we hypothesised could represent functional motifs, were identified. The clustering of specific TFBS identified within each PCRM were of particular interest due to the inherently large number of striated muscle associated TFBS enriched and in particular

the co-localisation of characterised interacting motifs (for example, GATA, NFAT, Mef2). As an indication of the utility of this analysis, we believe, based purely on this comparative regulatory analysis, it was feasible to retrospectively predict aspects of *msl* context specific expression and function. For example, based on TFBS composition, one would predict the potential utilisation of the PP, UP2 and UP3 domains in striated muscle regulatory contexts including differentiation, development and Ca²⁺ dependant stress signalling, all contexts in which *msl* has been demonstrated to be differentially expressed. In conclusion, the *in silico* analysis described in Chapter 3 served to decrease the genomic search space for functional analysis therefore focusing our experimental strategy.

Subsequent to this comparative analysis we experimentally interrogated our identified PCRM and TFBS through the utilisation of a diverse array of *in vitro* and *in vivo* experimental models which were coupled with gain- and loss-of function studies. In Chapter 4, it was demonstrated that the cardiac specific regulatory activity of *msl* was associated with the PP, UP2 and UP3 domains but not the UP1 domain. It was proposed that the positively acting PP domain represented the striated muscle proximal promoter which contained the core promoter and TATA box. This proximal domain is essential for all cardiac specific promoter activity and serves to integrate context specific signals from the distal UP2 and UP3 domains, analogous to the ANP, BNP and NCX proximal promoters (Temsah & Nemer, 2005; Xu et al, 2006). It is encouraging that our *in vitro* findings support and extends Olson's work in which they demonstrate the proximal 127bp was sufficient to drive cardiac specific expression of a LacZ reporter in TG-reporter mice *in vivo* (Kawahara et al, 2007). This confirms the utility of the current *in vitro* models used and suggests findings from this approach will translate to the *in vivo* situation

The current literature (Kawahara et al, 2007) concludes that context specific regulatory activity of the *msl* gene is encompassed purely within the -1.5kbp 5'-flanking region which (with the knowledge gained from the current thesis) is premature and somewhat naive. Contrary to their work we provide strong evidence, both *in vitro* and *in vivo*, for important cardiac specific regulatory roles for the UP2 and UP3 domains. For example,

the present study provides evidence to suggest that the UP2 domain represents a Ca^{2+} dependent Calcineurin/NFAT sensitive module. We suspect this module, in collaboration with Olson's proximal Mef2 motif, mediates stress signal dependent activation. The observation that *msl* mRNA is up-regulated in two cardiac pathological models (sub-lethal I/R, SHR rat) in which calcineurin signalling is implicated would support this. If these findings translate *in vivo*, our UP2 module may therefore represent an interesting therapeutic regulatory target. For example, therapeutic perturbation of Calcineurin signalling at UP2 may normalise *msl* expression and thereby attenuate the subsequent cardiac phenotype. From a broader viewpoint, these data suggest that via the *msl* promoter (UP2 domain), the Ca^{2+} dependent calcineurin pathway can potentiate muscle specific SRF activity with this having a significant impact on striated muscle biology and disease, in particular cardiac hypertrophy where calcineurin has been strongly implicated.

To take this work further and expand upon our current findings and deductions it would be of interest to explore the functionality of the UP2 model *in vivo*, preferably via ChIP and the generation of LacZ reporter TG mice. In this transgenic setting one can mutate the NFAT residues within the UP2 module and observe how they manifest on LacZ activity in appropriate regulatory contexts. Quantifying the recruitment of endogenous NFAT at this module, and correlating this recruitment with *msl* expression during the appropriate contexts would validate the current findings. For example, how does binding at UP2 change during simulated sub-lethal I/R or post pressure overload? Ultimately determining how important calcineurin activity is for this activation would be of central importance. It is noteworthy however that in the calcineurin transgenic mouse model for heart failure, *msl* expression is very high, thereby supporting our hypothesis (Kuwahara et al, 2007).

In Chapter 5 we proceeded to interrogate the PP and UP3 domains for the transcription factors governing their cardiac specific activity. Through the utilisation of a wide array of *in vitro* and *in vivo* techniques including TG murine models and adenoviral based over-expression and knockdown strategies we demonstrated that both the PP and UP3 domains are targeted by GATA4. The PP domain was repressed in a DNA-binding

dependent fashion by GATA4 which we suspected was as a result of a steric hindrance mechanism while conversely the UP3 distal CRM was activated by GATA4, with its basal enhancing activity (see Chapter 4) was dependent on endogenous GATA4 binding. However, the overall “net” effect of GATA4 on the endogenous *msl* transcriptional unit was repression, with this having implications for the function of Msl/MRTF/SRF and GATA4 in cardiac muscle development and disease. Firstly, our findings demonstrate that *msl* is under exquisite “locking” via potential GATA4 steric hindrance mechanism at the proximal promoter. Although initially an unexpected finding however, when one considers that *msl* is the master muscle specific regulator of SRF activity and poor regulatory control of the *msl* promoter would result in attenuated or precocious expression and subsequently affect SRF activity. Numerous studies have demonstrated that increased or decreased SRF activity is implicated with detrimental cardiac function (Figure 7.1) including the development of cardiac hypertrophy and cardiomyopathy associated with sarcomeric and myofibrillar disarray (Miano, 2003; Niu et al, 2005; Sahai et al, 1998; Sepulveda et al, 2002; Shin et al, 2002; Sotiropoulos et al, 1999; Wang et al, 2001; Wang et al, 2002; Xing et al, 2006; Zhang et al, 2001a; Zhang et al, 2001b). Therefore, poor control of *msl*, and by association SRF, can have detrimental effects on cardiac development, differentiation, maturation and post natal homeostatic function.

GATA4 is also a well characterised master regulator of these cardiac processes and hence it is not surprising the cardiac regulatory system would require integration between the GATA4 and SRF gene regulatory networks (GRNs). We therefore propose that our exquisite “locking” mechanism provides a means by which GATA4 can robustly regulate *msl* expression and thus SRF activity. We have therefore established a new link between these two GRNs and discovered an auto-regulatory negative feedback network motif because SRF can also regulate GATA4 expression. Thus, precocious *msl* expression would increase GATA4 expression, when uncontrolled, can lead to the development of cardiac phenotypes and malfunction. Through binding and repressing *msl* at the PP domain, GATA4 has established a negative feedback loop, which serves to prevent precocious expression of *msl* and therefore the development of associated detrimental phenotype (Kuwahara et al, 2007)(see Figure 7.1 for summary).

The newly identified GATA4-*msl*-SRF GRN proposed here has implications for disease phenotypes. For example, any changes in GATA4 levels will have profound effects on *msl* and therefore SRF activity. Numerous studies have implicated GATA4 deregulation in developmental (congenital heart defects) and post natal (cardiac hypertrophy, cardiomyopathy) cardiac pathologies (Aries et al, 2004; Charron & Nemer, 1999; Charron et al, 1999; Garg et al, 2003; Hautala et al, 2001; Liang et al, 2001; Majalahti et al, 2007; Morin et al, 2000; Pikkariainen et al, 2004; Pu et al, 2004; Rajagopal et al, 2007). However, although directly associated, the exact mechanism via which GATA4 causes these phenotypes remains an area of uncertainty (so called “black box”). We suggest that our findings place *msl* within this black box and propose a mechanism via which dysregulated GATA4 expression causes subsequent changes in *msl* expression and SRF activity which ultimately contribute to cardiac pathological phenotypes. Data presented in chapter 5 begins to support this hypothesis. Hyperglycemia (HG) induced cardiomyopathy has been demonstrated to be mediated by depletion of cardiac GATA4 levels (Kobayashi et al, 2006). We demonstrated in two models of HG (Type 1 and Type 2) induced diabetic cardiomyopathy (which is accompanied by GATA4 depletion) that *msl* mRNA is robustly induced. In light of the role played by *msl* in cardiomyopathy, we suspect *msl* up-regulation may contribute to the diabetic cardiomyopathic phenotypes observed in these Type 1 and Type 2 diabetic models. It is also of interest that our findings here may help resolve some paradoxes in the literature. For example, work from the group of Dr William Pu demonstrated that despite gene targeted deletion of GATA4, many hypertrophic associated gene transcripts were still being up-regulated in response to pressure overload (Bisping et al, 2006). Many of these genes are undoubtedly regulated by SRF and we suspect gene targeting of GATA4 would lead to precocious *msl* expression and therefore SRF activity, thus potentially up-regulating many of these hypertrophy associated transcripts and reconciling the paradoxical observations made by Bill Pu’s group.

In the future one of the main objectives must be to convincingly demonstrate that many of the described GATA4 phenotypes are as a consequence of *msl* dysregulation. If this is the case, then whether therapeutic manipulation of *msl* can serve to palliate some of

the GATA4 associated pathological phenotypes. It will also be of interest using ChIP and luciferase reporter assays to further validate the steric hindrance mechanism via which GATA4 “locks” *msl* transcription at the PP. Determining whether this is a common mode of control and how widespread it is may have relevance to general mechanism of transcriptional control and systems networking genome wide. Characterising the *in vivo* role for the UP3 module and probing whether other identified TFBS (Chapter 3) may play a regulatory role at these domains (and more importantly interact with GATA4) may also be fruitful.

In Chapter 6, the mechanisms governing skeletal muscle specific expression of *msl* was investigated. It was demonstrated that the temporal recruitment of the MRF, MyoD, at independent E-Box motifs within the distal UP1 and proximal PP domain was important for mediating correct temporal expression of *msl* during myogenic differentiation *in vitro*. The dynamics of this interaction led us to characterise the UP1 domain as a bona-fide skeletal muscle specific enhancer domain. These data implicated *msl* as a key component of a MyoD generated feed-forward regulatory circuit which functioned to temporally co-ordinate the timing of gene expression during skeletal myogenesis. We believe these findings provide the first piece of evidence to demonstrate a direct link between MyoD activity and SRF transcriptional signalling, with *msl* serving as the nodal point to integrate these two central myogenic regulatory networks. It is noteworthy that Olson’s group have demonstrated that in cardiomyocytes *msl* serves a similar function in that it integrates the Mef2 and SRF signalling networks, providing a link for crosstalk between Mef2 and SRF (Kuwahara et al, 2007). We have also demonstrated in Chapter 5 that *msl* integrates the GATA4 cardiogenic network SRF activity. Thus this appears to be a conserved emerging paradigm for *msl* function both in cardiac and skeletal muscle and suggests *msl* plays a pivotal regulatory role with respect to modulating striated muscle biology. In the future we would like to confirm the dynamics of MyoD binding at the UP1 domain *in vivo*, during skeletal muscle development and myogenesis. It would also be interesting to manipulate endogenous *msl* levels and characterise the effect of this on the activity of the MyoD GRN and subsequently myogenesis

Throughout this thesis we have constantly paid attention to the regulatory roles played by epigenetic phenomena, specifically histone modification. It was clearly demonstrated both in Chapters 4 and 6 that histone modification and specifically the dynamic role of histone acetylation (regulated by Class I and Class II HDACs) plays in determining correct context specific expression of *msl*. These findings may be of particular importance as many chemotherapeutic agents now exist which can specifically inhibit the various HDAC enzymes. It would be of interest to determine if such modulation of *msl* regulating HDACs could serve to normalise *msl* gene expression and therefore affect the cardiac pathological phenotype.

At the outset of this project the primary aims were to characterise the regulatory mechanisms governing *msl* expression. Taken together, the findings presented demonstrate that *msl* is regulated in a complex manner by the dynamic interplay of both proximal and distal *cis* regulatory motifs. These independent regulatory motifs ultimately collaborate to mediate the correct context specific transcriptional output and therefore ultimately give us a novel insight in *msl* function. These findings have also illuminated our current understanding of striated muscle specific regulatory contexts and have given potential mechanisms for developmental and disease phenotypes. To conclude, the findings here provide the foundations and framework for future studies to further interrogate *msl* regulation potentially allowing for the focussed development of novel therapeutic compounds for intervention in *msl* associated pathological phenotypes.

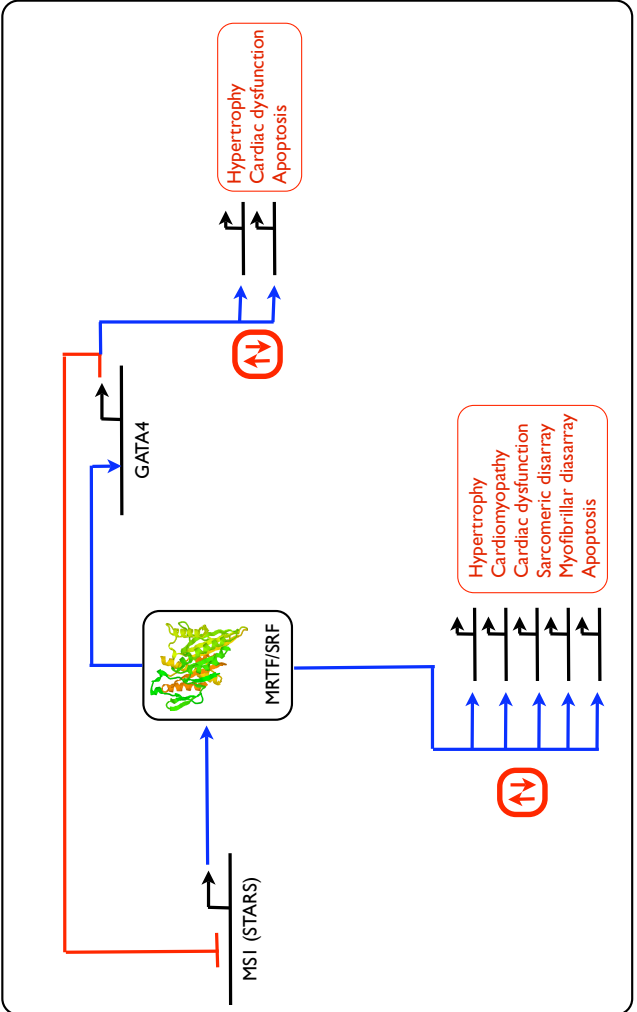


Figure 7.1 Schematic summary illustrating the interaction of the GATA4 and SRF GRNs and the formation of a negative feedback loop via MS1/STARS. Red and blue arrows represents repressive and activating interactions respectively.

APPENDIX 1

The IUPAC (International union of pure applied chemistry) code for representing degenerate nucleotide sequence patterns.

N = A, C, T or G

W = A or T

S = C or G

R = A or G

Y = C or T

K = G or T

M = A or C

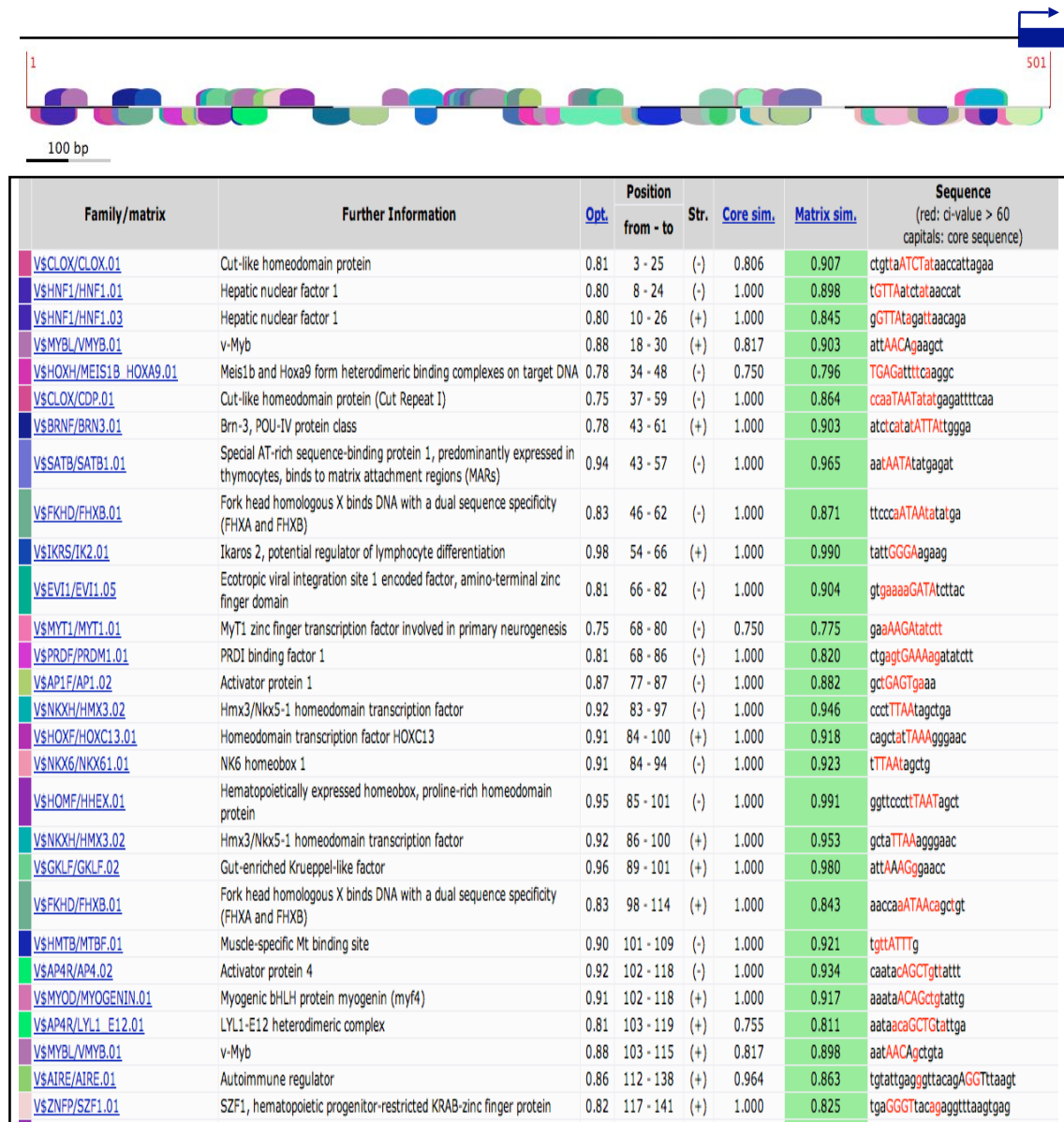
B = C, G, or T (not A)

D = A, G, or T (not C)

H = A, C, or T (not G)

V = A, C, or G (not T)

APPENDIX 2



V\$HOMF/EN1.01	Homeobox protein engrailed (en-1)	0.77	125 - 141	(+)	1.000	0.806	acagaggTTTAgtgag
V\$CP2F/CP2.02	LBP-1c (leader-binding protein-1c), LSF (late SV40 factor), CP2, SEF (SAA3 enhancer factor)	0.84	141 - 159	(-)	1.000	0.850	gACTGttccaggttgtctc
V\$BRNF/BRN5.01	Brn-5, POU-VI protein class (also known as emb and CNS-1)	0.74	159 - 177	(-)	0.750	0.782	agtgcACAaattgtcctag
V\$GREF/GRE.01	Glucocorticoid receptor, C2C2 zinc finger protein binds glucocorticoid dependent to GREs, IR3 sites	0.85	159 - 177	(-)	0.893	0.914	agtgacacaattGTCCtag
V\$MYBL/CMYB.02	c-Myb, important in hematopoiesis, cellular equivalent to avian myoblastosis virus oncogene v-myb	0.96	175 - 187	(+)	1.000	0.989	acTAAcgtctac
V\$TBP/TATA.01	Cellular and viral TATA box elements	0.90	188 - 204	(+)	1.000	0.961	tcctaTAAAtgittgga
V\$RUSH/SMARCA3.01	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3	0.96	191 - 201	(-)	1.000	0.975	aaCCATttata
V\$HOXH/MEIS1B HOXA9.01	Meis1b and Hoxa9 form heterodimeric binding complexes on target DNA	0.78	205 - 219	(+)	0.750	0.786	TCACattttaaggctc
V\$EREF/ERR.01	Estrogen related receptor	0.87	207 - 225	(+)	1.000	0.899	acattttAAGGtcaaggaa
V\$RORA/RORA.01	RAR-related orphan receptor alpha1	0.93	207 - 229	(+)	1.000	0.966	acattttaAGGtCaaggaaaaca
V\$RXRF/LXRE.02	Highly conserved DR1 element selected by LXRbeta/RXR heterodimers	0.69	207 - 231	(+)	1.000	0.692	acatttttaAGGtCaaggaaaacata
V\$NR2F/COUP.02	Chicken ovalbumin upstream promoter (COUP-TF), DR0 sites	0.84	208 - 232	(+)	1.000	0.969	cattttaaggTCAAggaaaacataa
V\$NBRE/NBRE.01	Monomers of the nur subfamily of nuclear receptors (nur77, nur1, nor-1)	0.86	210 - 224	(+)	1.000	0.924	ttttAAGGtcaaggaa
V\$IRFF/IRF4.01	Interferon regulatory factor (IRF)-related protein (NF-EM5, PIP, LSIRF, ICSAT)	0.94	213 - 233	(+)	1.000	0.960	taaggctcagGAAAcataaa
V\$CSEN/DREAM.01	Downstream regulatory element-antagonist modulator, Ca2+-binding protein of the neuronal calcium sensors family that binds DRE (downstream regulatory element) sites as a tetramer	0.95	215 - 225	(+)	1.000	0.960	agGTCAaggaa
V\$FKHD/ILF1.01	Winged-helix transcription factor IL-2 enhancer binding factor (ILF), forkhead box K2 (FOKK2)	0.98	218 - 234	(+)	1.000	0.988	tcaaggaaAACAataag
V\$NFAT/NFAT5.01	Nuclear factor of activated T-cells 5	0.83	219 - 237	(+)	1.000	0.842	caaGGAaAacataaagcta
V\$MITF/MIT.01	MIT (microphthalmia transcription factor) and TFE3	0.81	234 - 252	(-)	1.000	0.933	attgaatCATGtgttagc
V\$AP4R/PARAXIS.01	Paraxis (TCF15), member of the Twist subfamily of Class B bHLH factors, forms heterodimers with E12	0.86	235 - 251	(+)	0.882	0.893	ctaAGCacatgattcaa
V\$EBOX/MYCMAX.02	c-Myc/Max heterodimer	0.92	236 - 248	(+)	0.860	0.934	taagcaCATGatt
V\$AP1F/AP1.01	Activator protein 1	0.94	242 - 252	(+)	0.833	0.945	catgATTCaat
V\$AP1F/AP1.01	Activator protein 1	0.94	242 - 252	(-)	0.880	0.945	attgAATCatg
V\$HOXC/PBX1.01	Homeo domain factor Pbx-1	0.78	242 - 258	(-)	1.000	0.811	tactaGATTgaatcatg
V\$NFAT/NFAT.01	Nuclear factor of activated T-cells	0.95	249 - 267	(-)	1.000	0.954	ggtGGAaagtactagattg
V\$RU49/RU49.01	Zinc finger transcription factor RU49 (zinc finger proliferation 1 - Zipro 1). RU49 exhibits a strong preference for binding to tandem repeats of the minimal RU49 consensus binding site.	0.98	255 - 261	(-)	1.000	0.988	aAGTAct

VSPAX6/PAX6.02	PAX6 paired domain and homeodomain are required for binding to this site	0.87	262 - 280	(-)	1.000	0.882	ctccaagtgCAGggtgga
VSMYOD/MYOD.01	Myogenic regulatory factor MyoD (myf3)	0.88	266 - 282	(+)	1.000	0.881	ccctGGCacttgagaa
VSEBOX/USF.04	Upstream stimulating factor 1/2	0.90	268 - 280	(+)	0.851	0.927	ctggCACTtgag
VSEV11/EV11.02	Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain	0.83	276 - 292	(+)	1.000	0.834	tggagAAGAagaggg
VSPAX6/PAX6.04	PAX6 paired domain binding site	0.83	278 - 296	(-)	0.777	0.833	tgtTCCtcttctctc
VSGKLF/GKLF.01	Gut-enriched Krueppel-like factor	0.86	280 - 292	(+)	1.000	0.914	gaagaaaggAGGG
VSGATA/GATA.01	GATA binding factor	0.93	292 - 304	(-)	1.000	0.938	cactGATAtgttc
VSNR2F/HNF4.03	Hepatic nuclear factor 4, DR1 sites	0.83	297 - 321	(-)	1.000	0.899	tctctggtggCAAAGgacactgata
VSNBRE/NBRE.01	Monomers of the nur subfamily of nuclear receptors (nur77, nur1, nor-1)	0.86	299 - 313	(-)	1.000	0.862	ggcaAAGGacactga
VSPERO/PPAR_RXR.02	PPAR/RXR heterodimers, DR1 sites	0.69	300 - 322	(-)	1.000	0.748	atctctggtggCAAGgacactg
VSNRSF/NRSE.01	Neural-restrictive-silencer-element	0.67	321 - 341	(-)	1.000	0.741	ttcatcaccCGGAggacat
VSPARF/DBP.01	Albumin D-box binding protein	0.84	330 - 346	(+)	0.807	0.863	cgcggtGATgaaatggg
VSPARF/TEF_HLF.01	Thyrotrophic embryonic factor / hepatic leukemia factor	0.78	331 - 347	(-)	0.784	0.789	ttccaTTTCatcacgc
VSAARF/AARE.01	Amino acid response element, ATF4 binding site	0.95	335 - 343	(-)	1.000	0.995	aTTTCatca
VSTALE/MEIS1.01	Binding site for monomeric Meis1 homeodomain protein	0.95	347 - 357	(+)	1.000	0.991	accTGTcaggg
VSEKLF/EKLF.01	Erythroid krueppel like factor (EKLF)	0.89	348 - 364	(+)	1.000	0.911	cctgtcaGGGTttaaaa
VSCSEN/DREAM.01	Downstream regulatory element-antagonist modulator, Ca2+-binding protein of the neuronal calcium sensors family that binds DRE (downstream regulatory element) sites as a tetramer	0.95	349 - 359	(+)	1.000	0.984	ctGTCAgggtt
VSPAX8/PAX8.01	PAX 2/5/8 binding site	0.88	349 - 361	(+)	0.850	0.906	ctgTCAggttta
VSTBPF/MTATA.01	Muscle TATA box	0.84	350 - 366	(-)	1.000	0.884	gttttTAAAcctgaca
VSPAX2/PAX2.01	Zebrafish PAX2 paired domain protein	0.78	355 - 377	(-)	1.000	0.840	gacgtgttctgttttTAACcc
VSMYBL/MYB.01	v-Myb	0.88	361 - 373	(+)	0.817	0.882	aaaAACAgacac
VSCREB/ATF.01	Activating transcription factor	0.90	364 - 384	(-)	1.000	0.930	tggctcTGACggtgtctgtt
VSGREF/PRE.01	Progesterone receptor binding site, IR3 sites	0.84	365 - 383	(-)	1.000	0.846	ggcctcgacgtGTTCdtgt
VSMOKF/MOK2.01	Ribonucleoprotein associated zinc finger protein MOK-2 (mouse)	0.74	369 - 389	(+)	0.750	0.758	aacaccgtcagagCCATagca
VSOCT1/OCT1.02	Octamer-binding factor 1	0.85	406 - 420	(-)	1.000	0.870	actATGCaaggcatt
VSNFAT/NFAT.01	Nuclear factor of activated T-cells	0.95	409 - 427	(-)	1.000	0.950	ggtGGAaactatgcaaggc
VSOCTP/OCT1P.01	Octamer-binding factor 1, POU-specific domain	0.86	410 - 422	(-)	0.789	0.880	aaaCTATgcaagg
VSBPF/RFX1.02	X-box binding protein RFX1	0.90	416 - 434	(-)	0.881	0.938	ctgtgacggtgGAAActat
VSGABF/GAGA.01	GAGA-Box	0.78	432 - 456	(-)	0.750	0.796	qqqtqTGAAGtqaaqaaqaaqccctq
VSNFNP/SZF1.01	SZF1, hematopoietic progenitor-restricted KRAB-zinc finger protein	0.82	435 - 459	(-)	1.000	0.848	gcaGGGTgtgagtggaaggagcgc
VSRBPF/RBPJ.02	Mammalian transcriptional repressor RBP-Jkappa/CBF1	0.94	437 - 451	(-)	1.000	0.955	tgagTGGGaggagc
VSMEF2/SL1.01	Member of the RSRF (related to serum response factor) protein family from Xenopus laevis	0.84	455 - 477	(+)	1.000	0.866	cctgcttCTATttaatcccagg
VSLHXF/LHX3.01	Homeodomain binding site in LIM/Homeodomain factor LHX3	0.81	460 - 476	(-)	1.000	0.821	ctgggatTTAAatagaa
VSTBPF/ATATA.01	Avian C-type LTR TATA box	0.78	460 - 476	(+)	0.750	0.830	ttctattTAAAtcccag
VSTBPF/MTATA.01	Muscle TATA box	0.84	462 - 478	(+)	1.000	0.923	ctattTAAAtcccaggc
VSGF11/GF11.01	Growth factor independence 1 zinc finger protein acts as transcriptional repressor	0.96	466 - 480	(+)	1.000	0.967	ttaAATCccaggcaa
VSHMTB/MTBF.01	Muscle-specific Mt binding site	0.90	467 - 475	(-)	1.000	0.901	tgggATTTa
VSMYT1/MYT1L.01	Myelin transcription factor 1-like, neuronal C2HC zinc finger factor 1	0.92	474 - 486	(-)	1.000	0.944	agagAGTTgctcg
VSPBXC/PBX1_MEIS1.02	Binding site for a Pbx1/Meis1 heterodimer	0.77	480 - 496	(-)	0.750	0.776	aatgTGGTgagagagt
VSHAML/AML3.01	Runt-related transcription factor 2 / CBFA1 (core-binding factor, runt domain, alpha subunit 1)	0.84	483 - 497	(-)	1.000	0.870	gaatGTGgtgagag

UP1

Family/matrix	Further Information	Opt.	Position from - to	Str.	Core sim.	Matrix sim.	Sequence (red: ci-value > 60 capital: core sequence)
V\$PAX2/PAX2.01	Zebrafish PAX2 paired domain protein	0.78	5 - 27	(-)	1.000	0.818	tggtatataccagattaaAAGct
V\$HOXF/CRX.01	Cone-rod homeobox-containing transcription factor / otx-like homeobox gene	0.94	7 - 23	(+)	1.000	0.950	gtttTAATctggataaa
V\$MYOD/TAL1_E2A.01	Complex of Lmo2 bound to Tal-1, E2A proteins, and GATA-1, half-site 1	0.98	27 - 43	(+)	1.000	0.984	actgaaCAGGtgcgtgtt
V\$AP4R/TAL1ALPHA47.01	Tal-1alpha/E47 heterodimer	0.87	28 - 44	(+)	0.795	0.881	ctgaaCAGGtgcgtgtt
V\$ZFHX/AREB6.03	AREB6 (Atp1a1 regulatory element binding factor 6)	0.96	30 - 42	(-)	1.000	0.982	acagCACCGtttc
V\$ZFHX/AREB6.04	AREB6 (Atp1a1 regulatory element binding factor 6)	0.98	36 - 48	(+)	1.000	0.994	gtgctGTTTctct
V\$BNCF/BNC.01	Basonuclin, cooperates with UBF1 in rDNA PolI transcription	0.85	38 - 56	(+)	1.000	0.850	gctgtttctTGTCcaca
V\$MITF/MIT.01	MIT (microphthalmia transcription factor) and TFE3	0.81	46 - 64	(-)	1.000	0.825	gataagCATGtgacaga
V\$RORA/REV-ERBA.03	Orphan nuclear receptor rev-erb alpha (NR1D1), homodimer DR2 binding site	0.72	47 - 69	(-)	1.000	0.748	gaaaggaataGTCAtgtggacag
V\$AP1R/BACH1.01	BTB/POZ-bZIP transcription factor BACH1 forms heterodimers with the small Maf protein family	0.82	48 - 68	(-)	0.750	0.863	aaaggaTAAGtcagtggaca
V\$EBOX/USF.02	Upstream stimulating factor	0.90	48 - 60	(+)	0.837	0.912	tgctCACAtgact
V\$BRNF/BRN3.01	Brn-3, POU-IV protein class	0.78	50 - 68	(+)	0.750	0.796	tcacatgaCTTAtccctt
V\$BRNF/BRN5.01	Brn-5, POU-VI protein class (also known as emb and CNS-1)	0.74	50 - 68	(-)	0.750	0.778	aaagGATAagtcagtgtga
V\$HOXC/PBX_HOXA9.01	PBX - HOXA9 binding site	0.79	52 - 68	(+)	0.750	0.840	cacaTGACtattccctt
V\$EVI1/EVI1.05	Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain	0.81	55 - 71	(-)	1.000	0.833	ctgaaagGATAagtcac
V\$GATA/GATA.01	GATA binding factor	0.93	56 - 68	(-)	1.000	0.937	aaagGATAagtcac
V\$GREF/ARE.02	Androgene receptor binding site, IR3 sites	0.89	59 - 77	(+)	1.000	0.907	cttatcctttcaGTTctct
V\$IRFF/IRF4.01	Interferon regulatory factor (IRF)-related protein (NF-EM5, PIP, LSIRF, ICSAT)	0.94	59 - 79	(-)	1.000	0.959	taagagaactGAAAggataag
V\$EVI1/EVI1.04	Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain	0.73	60 - 76	(-)	1.000	0.804	gagaactgaaagGATAa
V\$MEF2/MEF2.01	Myocyte-specific enhancer factor 2	0.77	70 - 92	(+)	1.000	0.773	agttctctTAAaataatttt
V\$FKHD/FHXB.01	Fork head homologous X binds DNA with a dual sequence specificity (FHXA and FHXB)	0.83	74 - 90	(+)	0.818	0.849	ctcttaAAAAattttat
V\$FKHD/HNF3B.01	Hepatocyte nuclear factor 3beta (FOXA2)	0.94	79 - 95	(-)	1.000	0.995	agtaataAATAttttt
V\$BRNF/BRN2.03	Brn-2, POU-III protein class	0.92	80 - 98	(+)	1.000	0.960	aaaataattATTtactga
V\$NKXH/NKX3.1.01	Prostate-specific homeodomain protein NKX3.1	0.84	80 - 94	(-)	0.760	0.853	gtaaatAAATAttttt
V\$FKHD/FREAC3.01	Fork head related activator-3 (FOXC1)	0.84	83 - 99	(-)	1.000	0.960	atcaaGTAaataaatat
V\$HOMF/EN1.01	Homeobox protein engrailed (en-1)	0.77	83 - 99	(+)	1.000	0.787	atattttaTTAActtgat

Family/matrix	Further Information	Opt.	Position		Str.	Core sim.	Matrix sim.	Sequence (red: ci-value > 60 capitals: core sequence)
			from	to				
V\$BCL6/BCL6.02	POZ/zinc finger protein, transcriptional repressor, translocations observed in diffuse large cell lymphoma	0.77	3	19	(-)	1.000	0.784	agt gtat TAGaaaggca
V\$OCTB/TST1.01	POU-factor Tst-1/Oct-6	0.90	6	18	(-)	1.000	0.940	gtgt ATT Agaag
V\$HOXF/PHOX2.01	Phox2a (ARIX) and Phox2b	0.87	7	23	(+)	1.000	0.891	tttc TAAT acactggc
V\$SATB/SATB1.01	Special AT-rich sequence-binding protein 1, predominantly expressed in thymocytes, binds to matrix attachment regions (MARs)	0.94	9	23	(+)	1.000	0.947	tct AATA cacttggc
V\$NKXH/NKX3.1.01	Prostate-specific homeodomain protein NKX3.1	0.84	22	36	(+)	1.000	0.909	gctag AAGT aaagg
V\$CLOX/CUT2.01	Cut-like homeodomain protein (Cut repeat II)	0.67	26	48	(+)	0.750	0.679	gtaagt aaaggaaaATT acag
V\$PLZF/PLZF.01	Promyelocytic leukemia zinc finger (TF with nine Krueppel-like zinc fingers)	0.86	28	42	(+)	0.958	0.870	aag TAA Aggaaaatt
V\$SORY/HMGY.01	HMG(Y) high-mobility-group protein I (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA transcriptional complex	0.92	30	46	(-)	1.000	0.924	gtaa AATT tcctttac
V\$MYT1/MYT1.01	MyT1 zinc finger transcription factor involved in primary neurogenesis	0.75	33	45	(-)	0.750	0.798	taa AATT ttcctt
V\$BPF/TATA.02	Mammalian C-type LTR TATA box	0.89	34	50	(-)	1.000	0.890	tcctg TAAA attttct
V\$SORY/HMGY.01	HMG(Y) high-mobility-group protein I (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA transcriptional complex	0.92	35	51	(+)	1.000	0.920	ggaa AATT ttacaggat
V\$IRFF/IRF1.01	Interferon regulatory factor 1	0.87	48	68	(+)	1.000	0.905	ggat caaat GAAAgcatgg
V\$PRDF/BLIMP1.01	Transcriptional repressor B lymphocyte-induced maturation protein-1 (Blimp-1, prdm1)	0.81	52	70	(+)	1.000	0.837	caaatt GAA Agcatggct
V\$XBBF/RFX1.01	X-box binding protein RFX1	0.89	62	80	(+)	0.881	0.892	gtcatggctaa GAA Actct
V\$HNF6/HNF6.01	Liver enriched Cut - Homeodomain transcription factor HNF6 (ONECUT)	0.82	70	86	(+)	0.785	0.863	taaga aacTCT Atatt
V\$MEF2/SL1.01	Member of the RSRF (related to serum response factor) protein family from Xenopus laevis	0.84	72	94	(+)	1.000	0.957	agaa actCTAT atttagatcttc
V\$MEF2/MEF2.03	Myocyte-specific enhancer factor 2	0.89	73	95	(-)	1.000	0.905	cgaagat ctaaat ATAGagtttc
V\$GATA/GATA3.02	GATA-binding factor 3	0.91	83	95	(-)	1.000	0.947	cga GAT ctaaat
V\$STAT/STAT6.01	STAT6: signal transducer and activator of transcription 6	0.84	87	105	(-)	0.793	0.884	aagt TCT Tacgaagatct
V\$STAT/STAT6.01	STAT6: signal transducer and activator of transcription 6	0.84	88	106	(+)	0.793	0.907	gatc TTCG taagaaacttg
V\$MYT1/MYT1.02	MyT1 zinc finger transcription factor involved in primary neurogenesis	0.88	96	108	(-)	1.000	0.881	aac AAGT ttctta
V\$GRHL/GRHL3.01	Grainyhead-like 3 (sister-of-mammalian grainyhead - SOM)	0.82	98	110	(-)	0.783	0.844	taa acaAGT Ttct
V\$FKHD/FREAC2.01	Fork head related activator-2 (FOX2)	0.84	100	116	(-)	1.000	0.899	actttc TAA Acaagttt
V\$CREB/ATF2.01	Activating transcription factor 2	0.87	109	129	(-)	1.000	0.938	agcttg TGACG ccactttcta
V\$CREB/CREB.02	cAMP-responsive element binding protein	0.89	112	132	(-)	1.000	0.913	acaagcttg TGACG ccacttt

V\$WHNF/WHN.01	Winged helix protein, involved in hair keratinization and thymus epithelium differentiation	0.95	114 - 124	(-)	1.000	0.960	gtgACGCact
V\$P53F/P53.01	Tumor suppressor p53	0.73	123 - 145	(+)	0.844	0.739	acaagCTGTgaagcttgcct
V\$P53F/P53.06	Tumor suppressor p53	0.77	124 - 146	(-)	0.938	0.780	aaggcaAAGacttcaagcttg
V\$BNCF/BNC.01	Basonuclin, cooperates with UBF1 in rDNA PolI transcription	0.85	130 - 148	(+)	1.000	0.882	tgtgaagctcTGCTcttga
V\$EREF/ERR.01	Estrogen related receptor	0.87	135 - 153	(-)	1.000	0.932	cacctcAAGGacaagact
V\$SMAD/SMAD4.01	Smad4 transcription factor involved in TGF-beta signaling	0.94	135 - 143	(+)	1.000	0.943	aGTCTgtc
V\$SF1F/SF1.01	SF1 steroidogenic factor 1	0.95	139 - 151	(-)	1.000	0.954	ccctCAAGgaca
V\$EKLF/EKLF.01	Erythroid krueppel like factor (EKLF)	0.89	142 - 158	(+)	1.000	0.907	tcctgaGGGTggggga
V\$ZBPF/ZNF219.01	Kruppel-like zinc finger protein 219	0.91	142 - 164	(-)	1.000	0.939	ctttctCCCCaccctcaagg
V\$GCMF/GCM1.01	Gilal cells missing homolog 1, chorion-specific transcription factor GCMa	0.85	143 - 153	(-)	1.000	0.895	caCCCTcaagg
V\$SP1F/GC.01	GC box elements	0.88	145 - 159	(+)	0.872	0.881	ttgagGTGTggggag
V\$ZBPF/ZBP89.01	Zinc finger transcription factor ZBP-89	0.93	145 - 167	(-)	1.000	0.972	ttgtttctCCCCaccctcaa
V\$EGRF/WT1.01	Wilms Tumor Suppressor	0.92	147 - 163	(+)	1.000	0.997	gaggTGGGgaggaaa
V\$EKLF/BKLF.01	Basic krueppel-like factor (KLF3)	0.95	147 - 163	(+)	1.000	0.971	gaGGGTggggaggaaa
V\$EGRF/CKROX.01	Collagen krox protein (zinc finger protein 67 - zfp67)	0.88	151 - 167	(+)	1.000	0.885	gtgGGGAggaagcaa
V\$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	0.90	153 - 165	(+)	1.000	0.910	ggggGAGGaaagc
V\$NFAT/NFAT.01	Nuclear factor of activated T-cells	0.95	156 - 174	(+)	1.000	0.970	ggaGGAAGgcaagagaaa
V\$LEFF/LEF1.01	TCF/LEF-1, involved in the Wnt signal transduction pathway	0.86	158 - 174	(+)	1.000	0.896	aggaaagCAAAAgagaaa
V\$IRFF/IRF2.01	Interferon regulatory factor 2	0.85	161 - 181	(+)	1.000	0.889	aaagcaagaGAAAggacaaa
V\$PRDF/PRDM1.01	PRDM1 binding factor 1	0.81	165 - 183	(+)	1.000	0.820	caaagaGAAAggacaaaat
V\$NBRE/NBRE.01	Monomers of the nur subfamily of nuclear receptors (nur77, nurr1, nor-1)	0.86	169 - 183	(+)	1.000	0.872	gagaAAGGacaaaat
V\$GREF/PRE.01	Progesterone receptor binding site, IR3 sites	0.84	172 - 190	(+)	0.750	0.845	aaaggacaaaaTTTctta
V\$MYT1/MYT1.01	MyT1 zinc finger transcription factor involved in primary neurogenesis	0.75	177 - 189	(+)	0.750	0.800	acaAAATttctt
V\$CP2F/CP2.02	LBP-1c (leader-binding protein-1c), LSF (late SV40 factor), CP2, SEF (SAA3 enhancer factor)	0.84	192 - 210	(-)	1.000	0.903	aACTGctgaagctctgggtc
V\$SMAD/SMAD3.01	Smad3 transcription factor involved in TGF-beta signaling	0.99	193 - 201	(-)	1.000	0.997	aGTCTgggt

Family/matrix	Further Information	Opt.	Position		Str.	Core sim.	Matrix sim.	Sequence (red: ci-value > 60 capital: core sequence)
			from	to				
V\$GATA/GATA1.04	GATA-binding factor 1	0.91	1	13	(-)	1.000	0.996	gacaGATAacagc
V\$FKHD/FREAC2.01	Fork head related activator-2 (FOXF2)	0.84	21	37	(-)	1.000	0.892	agggctTAACacactt
V\$NKXH/HMX3.02	Hmx3/Nkx5-1 homeodomain transcription factor	0.92	30	44	(-)	1.000	0.944	gctgTTAAgggtcta
V\$PAX2/PAX2.01	Zebrafish PAX2 paired domain protein	0.78	30	52	(+)	0.789	0.788	tagacccttaacagctTAATtt
V\$NKXH/HMX3.02	Hmx3/Nkx5-1 homeodomain transcription factor	0.92	33	47	(+)	1.000	0.937	accctTTAAcagctta
V\$HOXH/MEIS1B_HOXA9.01	Meis1b and Hoxa9 form heterodimeric binding complexes on target DNA	0.78	38	52	(+)	0.750	0.781	TAACagcttaaat
V\$TBPF/ATATA.01	Avian C-type LTR TATA box	0.78	38	54	(-)	1.000	0.790	ctaaattTAAGctgtta
V\$NOLF/OLF1.01	Olfactory neuron-specific factor	0.82	71	93	(+)	1.000	0.824	ccagctTCCctgtgggtatgag
V\$SORY/HMGA.01	HMGA family of architectural transcription factors (HMGA1, HMGA2)	0.88	92	108	(-)	1.000	0.924	agtCATTtaacacct
V\$HOXF/HOXB9.01	Abd-B-like homeodomain protein Hoxb-9	0.88	93	109	(+)	1.000	0.919	gtgggtgtTAAtgactg
V\$SORY/HBP1.01	HMG box-containing protein 1	0.86	98	114	(+)	0.800	0.860	gttaaatgACTGacaga
V\$PBXC/PBX1_MEIS1.01	Binding site for a Pbx1/Meis1 heterodimer	0.74	100	116	(+)	0.767	0.905	taaaTGACtgcacagatg
V\$TCFF/TCF1.01	TCF1/LCR-F1/Nrf1 homodimers	1.00	101	107	(-)	1.000	1.000	GTCatt
V\$BRAC/BRACH.01	Brachyury	0.66	103	123	(+)	0.750	0.667	atgactgacAGATgttaaca
V\$TALE/TGIF.01	TG-interacting factor belonging to TALE class of homeodomain factors	1.00	105	115	(-)	1.000	1.000	atctGTCAgtc
V\$AP4R/TAL1ALPHA47.01	Tal-1alpha/E47 heterodimer	0.87	106	122	(+)	1.000	0.962	actgaCAGAtgttaaac
V\$HOXH/MEIS1A_HOXA9.01	Meis1a and Hoxa9 form heterodimeric binding complexes on target DNA	0.77	108	122	(+)	1.000	0.835	TGACagatgttaaac
V\$RP58/RP58.01	Zinc finger protein RP58 (ZNF238), associated preferentially with heterochromatin	0.84	108	120	(-)	1.000	0.936	ttaaCATCtctga
V\$GATA/GATA3.02	GATA-binding factor 3	0.91	109	121	(+)	1.000	0.948	gacAGATgttaaa
V\$FKHD/FREAC2.01	Fork head related activator-2 (FOXF2)	0.84	112	128	(+)	1.000	0.905	agatgtTAACaaagat
V\$LEFF/LEF1.01	TCF/LEF-1, involved in the Wnt signal transduction pathway	0.86	115	131	(+)	1.000	0.862	tgtaaaCAAAGatacc
V\$MYT1/MYT1.01	MyT1 zinc finger transcription factor involved in primary neurogenesis	0.75	121	133	(+)	0.750	0.776	acaAGAtacctt
V\$HMTB/MTBF.01	Muscle-specific Mt binding site	0.90	124	132	(-)	0.884	0.901	aggtATCTt
V\$RBPJ/RBPJK.02	Mammalian transcriptional repressor RBP-Jkappa/CBF1	0.94	127	141	(-)	1.000	0.968	tggtTGGGaaaggtat
V\$ETSF/PEA3.01	Polyomavirus enhancer A binding protein 3, ETV4 (Ets variant gene 4)	0.94	135	155	(-)	1.000	0.956	gagtcacAGGAagctgtgtgg
V\$GKLF/GKLF.02	Gut-enriched Krueppel-like factor	0.96	141	153	(-)	1.000	0.977	gtcAAAGaaagct
V\$LEFF/LEF1.02	TCF/LEF-1, involved in the Wnt signal transduction pathway	0.94	142	158	(-)	1.000	0.941	aaagagCAAAGgaagc
V\$AP1R/MARE.03	Binding sites for heterodimers with small Maf-proteins	0.82	150	170	(-)	0.775	0.827	taaaGCTCactcaagagtc
V\$HOXH/MEIS1B_HOXA9.01	Meis1b and Hoxa9 form heterodimeric binding complexes on target DNA	0.78	150	164	(+)	1.000	0.816	TGACtcttgagtg
V\$AP1F/AP1.02	Activator protein 1	0.87	156	166	(+)	1.000	0.894	tttGAGTgagc
V\$NR2F/PNR.01	Photoreceptor-specific nuclear receptor subfamily 2, group E, member 3 (Nr2e3), DR1 sites	0.79	156	180	(-)	0.758	0.798	ccccacagTAAAgctcactcaaa

REFERENCES

Aasum E, Hafstad AD, Severson DL, Larsen TS (2003) Age-dependent changes in metabolism, contractile function, and ischemic sensitivity in hearts from db/db mice. *Diabetes* 52(2): 434-441

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, WoodageT, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287 (5461): 2185-2195

Andres V, Walsh K (1996) Myogenin expression, cell cycle withdrawal, and phenotypic differentiation are temporally separable events that precede cell fusion upon myogenesis. *J Cell Biol* 132(4): 657-666

Antos CL, McKinsey TA, Frey N, Kutschke W, McAnally J, Shelton JM, Richardson JA, Hill JA, Olson EN (2002) Activated glycogen synthase-3 beta suppresses cardiac hypertrophy in vivo. *Proc Natl Acad Sci U S A* 99(2): 907-912

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F,

Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297(5585): 1301-1310

Arai A, Spencer JA, Olson EN (2002) STARS, a striated muscle activator of Rho signaling and serum response factor-dependent transcription. *J Biol Chem* 277(27): 24453-24459

Araujo-Jorge TC, Waghbi MC, Hasslocher-Moreno AM, Xavier SS, Higuchi Mde L, Keramidas M, Bailly S, Feige JJ (2002) Implication of transforming growth factor-beta1 in Chagas disease myocardiopathy. *J Infect Dis* 186(12): 1823-1828

Aries A, Paradis P, Lefebvre C, Schwartz RJ, Nemer M (2004) Essential role of GATA-4 in cell survival and drug-induced cardiotoxicity. *Proc Natl Acad Sci U S A* 101(18): 6975-6980

Arsenian S, Weinhold B, Oelgeschlager M, Ruther U, Nordheim A (1998) Serum response factor is essential for mesoderm formation during mouse embryogenesis. *EMBO J* 17(21): 6289-6299

Asakura A, Fujisawa-Sehara A, Komiya T, Nabeshima Y (1993) MyoD and myogenin act on the chicken myosin light-chain 1 gene as distinct transcriptional factors. *Mol Cell Biol* 13(11): 7153-7162

Atkins GB, Jain MK (2007) Role of Kruppel-like transcription factors in endothelial biology. *Circ Res* 100(12): 1686-1695

Backs J, Olson EN (2006) Control of cardiac growth by histone acetylation/deacetylation. *Circ Res* 98(1): 15-24

Badorff C, Seeger FH, Zeiher AM, Dimmeler S (2005) Glycogen synthase kinase 3beta inhibits myocardin-dependent transcription and hypertrophy induction through site-specific phosphorylation. *Circ Res* 97(7): 645-654

Bagheri-Fam S, Ferraz C, Demaille J, Scherer G, Pfeifer D (2001) Comparative genomics of the SOX9 region in human and *Fugu rubripes*: conservation of short regulatory sequence elements within large intergenic regions. *Genomics* 78(1-2): 73-82

Balza RO, Jr., Misra RP (2006) Role of the serum response factor in regulating contractile apparatus gene expression and sarcomeric integrity in cardiomyocytes. *J Biol Chem* 281(10): 6498-6510

Barrientos T, Frank D, Kuwahara K, Bezprozvannaya S, Pipes GC, Bassel-Duby R, Richardson JA, Katus HA, Olson EN, Frey N (2007) Two novel members of the

ABLIM protein family, ABLIM-2 and -3, associate with STARS and directly bind F-actin. *J Biol Chem* 282(11): 8393-8403

Beals CR, Sheridan CM, Turck CW, Gardner P, Crabtree GR (1997) Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3. *Science* 275(5308): 1930-1934

Belaguli NS, Schildmeyer LA, Schwartz RJ (1997) Organization and myogenic restricted expression of the murine serum response factor gene. A role for autoregulation. *J Biol Chem* 272(29): 18222-18231

Belaguli NS, Sepulveda JL, Nigam V, Charron F, Nemer M, Schwartz RJ (2000) Cardiac tissue enriched factors serum response factor and GATA-4 are mutual coregulators. *Mol Cell Biol* 20(20): 7550-7558

Bhalla SS, Robitaille L, Nemer M (2001) Cooperative activation by GATA-4 and YY1 of the cardiac B-type natriuretic peptide promoter. *J Biol Chem* 276(14): 11439-11445

Bi W, Drake CJ, Schwarz JJ (1999) The transcription factor MEF2C-null mouse exhibits complex vascular malformations and reduced cardiac expression of angiopoietin 1 and VEGF. *Dev Biol* 211(2): 255-267

Bingham AJ, Ooi L, Kozera L, White E, Wood IC (2007) The repressor element 1-silencing transcription factor regulates heart-specific gene expression using multiple chromatin-modifying complexes. *Mol Cell Biol* 27(11): 4082-4092

Bingham AJ, Ooi L, Wood IC (2006) Multiple chromatin modifications important for gene expression changes in cardiac hypertrophy. *Biochem Soc Trans* 34(Pt 6): 1138-1140

Black BL, Olson EN (1998) Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu Rev Cell Dev Biol* 14: 167-196

Blais A, Dynlacht BD (2005) Constructing transcriptional regulatory networks. *Genes Dev* 19(13): 1499-1511

Blau HM, Pavlath GK, Hardeman EC, Chiu CP, Silberstein L, Webster SG, Miller SC, Webster C (1985) Plasticity of the differentiated state. *Science* 230(4727): 758-766

Bogoyevitch MA, Andersson MB, Gillespie-Brown J, Clerk A, Glennon PE, Fuller SJ, Sugden PH (1996a) Adrenergic receptor stimulation of the mitogen-activated protein kinase cascade and cardiac hypertrophy. *Biochem J* 314 (Pt 1): 115-121

Bogoyevitch MA, Gillespie-Brown J, Ketterman AJ, Fuller SJ, Ben-Levy R, Ashworth A, Marshall CJ, Sugden PH (1996b) Stimulation of the stress-activated mitogen-activated protein kinase subfamilies in perfused heart. p38/RK mitogen-activated

protein kinases and c-Jun N-terminal kinases are activated by ischemia/reperfusion. *Circ Res* 79(2): 162-173

Bogoyevitch MA, Sugden PH (1996) The role of protein kinases in adaptational growth of the heart. *Int J Biochem Cell Biol* 28(1): 1-12

Brand-Saberi B (2005) Genetic and epigenetic control of skeletal muscle development. *Ann Anat* 187(3): 199-207

Braz JC, Bueno OF, Liang Q, Wilkins BJ, Dai YS, Parsons S, Braunwart J, Glascock BJ, Klevitsky R, Kimball TF, Hewett TE, Molkentin JD (2003) Targeted inhibition of p38 MAPK promotes hypertrophic cardiomyopathy through upregulation of calcineurin-NFAT signaling. *J Clin Invest* 111(10): 1475-1486

Brown CO, 3rd, Chi X, Garcia-Gras E, Shirai M, Feng XH, Schwartz RJ (2004) The cardiac determination factor, Nkx2-5, is activated by mutual cofactors GATA-4 and Smad1/4 via a novel upstream enhancer. *J Biol Chem* 279(11): 10659-10669

Brown SA, Imbalzano AN, Kingston RE (1996) Activator-dependent regulation of transcriptional pausing on nucleosomal templates. *Genes Dev* 10(12): 1479-1490

Buckingham M, Bajard L, Chang T, Daubas P, Hadchouel J, Meilhac S, Montarras D, Rocancourt D, Relaix F (2003) The formation of skeletal muscle: from somite to limb. *J Anat* 202(1): 59-68

Bueno OF, Molkentin JD (2002) Involvement of extracellular signal-regulated kinases 1/2 in cardiac hypertrophy and cell death. *Circ Res* 91(9): 776-781

Bueno OF, van Rooij E, Molkentin JD, Doevendans PA, De Windt LJ (2002a) Calcineurin and hypertrophic heart disease: novel insights and remaining questions. *Cardiovasc Res* 53(4): 806-821

Bueno OF, Wilkins BJ, Tymitz KM, Glascock BJ, Kimball TF, Lorenz JN, Molkentin JD (2002b) Impaired cardiac hypertrophic response in Calcineurin Abeta⁻deficient mice. *Proc Natl Acad Sci U S A* 99(7): 4586-4591

Callis TE, Cao D, Wang DZ (2005) Bone morphogenetic protein signaling modulates myocardin transactivation of cardiac genes. *Circ Res* 97(10): 992-1000

Cantley LC (2002) The phosphoinositide 3-kinase pathway. *Science* 296(5573): 1655-1657

Cao D, Wang Z, Zhang CL, Oh J, Xing W, Li S, Richardson JA, Wang DZ, Olson EN (2005) Modulation of smooth muscle gene expression by association of histone acetyltransferases and deacetylases with myocardin. *Mol Cell Biol* 25(1): 364-376

- Carnac G, Primig M, Kitzmann M, Chafey P, Tuil D, Lamb N, Fernandez A (1998) RhoA GTPase and serum response factor control selectively the expression of MyoD without affecting Myf5 in mouse myoblasts. *Mol Biol Cell* 9(7): 1891-1902
- Carson JA, Schwartz RJ, Booth FW (1996) SRF and TEF-1 control of chicken skeletal alpha-actin gene during slow-muscle hypertrophy. *Am J Physiol* 270(6 Pt 1): C1624-1633
- Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 21(13): 2933-2942
- Catala F, Wanner R, Barton P, Cohen A, Wright W, Buckingham M (1995) A skeletal muscle-specific enhancer regulated by factors binding to E and CArG boxes is present in the promoter of the mouse myosin light-chain 1A gene. *Mol Cell Biol* 15(8): 4585-4596
- Cen B, Selvaraj A, Burgess RC, Hitzler JK, Ma Z, Morris SW, Prywes R (2003) Megakaryoblastic leukemia 1, a potent transcriptional coactivator for serum response factor (SRF), is required for serum induction of SRF target genes. *Mol Cell Biol* 23(18): 6597-6608
- Cen B, Selvaraj A, Prywes R (2004) Myocardin/MKL family of SRF coactivators: key regulators of immediate early and muscle specific gene expression. *J Cell Biochem* 93(1): 74-82
- Charron F, Nemer M (1999) GATA transcription factors and cardiac development. *Semin Cell Dev Biol* 10(1): 85-91
- Charron F, Paradis P, Bronchain O, Nemer G, Nemer M (1999) Cooperative interaction between GATA-4 and GATA-6 regulates myocardial gene expression. *Mol Cell Biol* 19(6): 4355-4365
- Charron F, Tsimiklis G, Arcand M, Robitaille L, Liang Q, Molkentin JD, Meloche S, Nemer M (2001) Tissue-specific GATA factors are transcriptional effectors of the small GTPase RhoA. *Genes Dev* 15(20): 2702-2719
- Charvet C, Houbron C, Parlakian A, Giordani J, Lahoute C, Bertrand A, Sotiropoulos A, Renou L, Schmitt A, Melki J, Li Z, Daegelen D, Tuil D (2006) New role for serum response factor in postnatal skeletal muscle growth and regeneration via the interleukin 4 and insulin-like growth factor 1 pathways. *Mol Cell Biol* 26(17): 6664-6674
- Chen CY, Schwartz RJ (1996) Recruitment of the tinman homolog Nkx-2.5 by serum response factor activates cardiac alpha-actin gene transcription. *Mol Cell Biol* 16(11): 6372-6384

- Chen F, Kook H, Milewski R, Gitler AD, Lu MM, Li J, Nazarian R, Schnepf R, Jen K, Biben C, Runke G, Mackay JP, Novotny J, Schwartz RJ, Harvey RP, Mullins MC, Epstein JA (2002a) Hop is an unusual homeobox gene that modulates cardiac development. *Cell* 110(6): 713-723
- Chen J, Kitchen CM, Streb JW, Miano JM (2002b) Myocardin: a component of a molecular switch for smooth muscle differentiation. *J Mol Cell Cardiol* 34(10): 1345-1356
- Chien KR, Knowlton KU, Zhu H, Chien S (1991) Regulation of cardiac gene expression during myocardial growth and hypertrophy: molecular studies of an adaptive physiologic response. *FASEB J* 5(15): 3037-3046
- Chien KR, Olson EN (2002) Converging pathways and principles in heart development and disease: CV@CSH. *Cell* 110(2): 153-162
- Christ B, Ordahl CP (1995) Early stages of chick somite development. *Anat Embryol (Berl)* 191(5): 381-396
- Clemmitson JR, Dixon RJ, Haines S, Bingham AJ, Patel BR, Hall L, Lo M, Sassard J, Charchar FJ, Samani NJ (2007) Genetic dissection of a blood pressure quantitative trait locus on rat chromosome 1 and gene expression analysis identifies SPON1 as a novel candidate hypertension gene. *Circ Res* 100(7): 992-999
- Clerk A, Cullingford TE, Fuller SJ, Giraldo A, Markou T, Pikkarainen S, Sugden PH (2007) Signaling pathways mediating cardiac myocyte gene expression in physiological and stress responses. *J Cell Physiol* 212(2): 311-322
- Clerk A, Sugden PH (2000) Small guanine nucleotide-binding proteins and myocardial hypertrophy. *Circ Res* 86(10): 1019-1023
- Condorelli G, Drusco A, Stassi G, Bellacosa A, Roncarati R, Iaccarino G, Russo MA, Gu Y, Dalton N, Chung C, Latronico MV, Napoli C, Sadoshima J, Croce CM, Ross J, Jr. (2002) Akt induces enhanced myocardial contractility and cell size in vivo in transgenic mice. *Proc Natl Acad Sci U S A* 99(19): 12333-12338
- Corey LL, Weirich CS, Benjamin IJ, Kingston RE (2003) Localized recruitment of a chromatin-remodeling activity by an activator in vivo drives transcriptional elongation. *Genes Dev* 17(11): 1392-1401
- Croissant JD, Kim JH, Eichele G, Goering L, Lough J, Prywes R, Schwartz RJ (1996) Avian serum response factor expression restricted primarily to muscle cell lineages is required for alpha-actin gene transcription. *Dev Biol* 177(1): 250-264

- Cucoranu I, Clempus R, Dikalova A, Phelan PJ, Ariyan S, Dikalov S, Sorescu D (2005) NAD(P)H oxidase 4 mediates transforming growth factor-beta1-induced differentiation of cardiac fibroblasts into myofibroblasts. *Circ Res* 97(9): 900-907
- Cullingford TE, Markou T, Fuller SJ, Giraldo A, Pikkarainen S, Zoumpoulidou G, Alsafi A, Ekere C, Kemp TJ, Dennis JL, Game L, Sugden PH, Clerk A (2008) Temporal regulation of expression of immediate early and second phase transcripts by endothelin-1 in cardiomyocytes. *Genome Biol* 9(2): R32
- Davis RL, Cheng PF, Lassar AB, Thayer M, Tapscott S, Weintraub H (1989) MyoD and achaete-scute: 4-5 amino acids distinguishes myogenesis from neurogenesis. *Princess Takamatsu Symp* 20: 267-278
- Davis RL, Weintraub H, Lassar AB (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51(6): 987-1000
- de Ruijter AJ, van Gennip AH, Caron HN, Kemp S, van Kuilenburg AB (2003) Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J* 370 (Pt 3): 737-749
- Dedieu S, Mazeres G, Cottin P, Brustis JJ (2002) Involvement of myogenic regulator factors during fusion in the cell line C2C12. *Int J Dev Biol* 46(2): 235-241
- Delgado I, Huang X, Jones S, Zhang L, Hatcher R, Gao B, Zhang P (2003) Dynamic gene expression during the onset of myoblast differentiation in vitro. *Genomics* 82(2): 109-121
- Derynck R, Zhang YE (2003) Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* 425(6958): 577-584
- Di Padova M, Caretti G, Zhao P, Hoffman EP, Sartorelli V (2007) MyoD acetylation influences temporal patterns of skeletal muscle gene expression. *J Biol Chem* 282(52): 37650-37659
- Dixon IM, Hao J, Reid NL, Roth JC (2000) Effect of chronic AT(1) receptor blockade on cardiac Smad overexpression in hereditary cardiomyopathic hamsters. *Cardiovasc Res* 46(2): 286-297
- Du KL, Ip HS, Li J, Chen M, Dandre F, Yu W, Lu MM, Owens GK, Parmacek MS (2003) Myocardin is a critical serum response factor cofactor in the transcriptional program regulating smooth muscle cell differentiation. *Mol Cell Biol* 23(7): 2425-2437
- Durocher D, Charron F, Warren R, Schwartz RJ, Nemer M (1997) The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. *EMBO J* 16(18): 5687-5696

- Edmondson DG, Olson EN (1989) A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. *Genes Dev* 3(5): 628-640
- Feinberg MW, Lin Z, Fisch S, Jain MK (2004) An emerging role for Kruppel-like factors in vascular biology. *Trends Cardiovasc Med* 14(6): 241-246
- Fickett JW, Hatzigeorgiou AG (1997) Eukaryotic promoter recognition. *Genome Res* 7(9): 861-878
- Firulli AB (2003) A HANDful of questions: the molecular biology of the heart and neural crest derivatives (HAND)-subclass of basic helix-loop-helix transcription factors. *Gene* 312: 27-40
- Firulli AB, Olson EN (1997) Modular regulation of muscle gene transcription: a mechanism for muscle cell diversity. *Trends Genet* 13(9): 364-369
- Fisch S, Gray S, Heymans S, Halder SM, Wang B, Pfister O, Cui L, Kumar A, Lin Z, Sen-Banerjee S, Das H, Petersen CA, Mende U, Burleigh BA, Zhu Y, Pinto YM, Liao R, Jain MK (2007) Kruppel-like factor 15 is a regulator of cardiomyocyte hypertrophy. *Proc Natl Acad Sci U S A* 104(17): 7074-7079
- Fischle W, Wang Y, Allis CD (2003) Histone and chromatin cross-talk. *Curr Opin Cell Biol* 15(2): 172-183
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32(Web Server issue): W273-279
- Frey N, Barrientos T, Shelton JM, Frank D, Rutten H, Gehring D, Kuhn C, Lutz M, Rothermel B, Bassel-Duby R, Richardson JA, Katus HA, Hill JA, Olson EN (2004) Mice lacking calstabin-1 are sensitized to calcineurin signaling and show accelerated cardiomyopathy in response to pathological biomechanical stress. *Nat Med* 10(12): 1336-1343
- Fu Y, Weng Z (2005) Improvement of TRANSFAC matrices using multiple local alignment of transcription factor binding site sequences. *Genome Inform* 16(1): 68-72
- Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K, Matsuoka R, Cohen JC, Srivastava D (2003) GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* 424(6947): 443-447
- Gerdes AM (1992) Remodeling of ventricular myocytes during cardiac hypertrophy and heart failure. *J Fla Med Assoc* 79(4): 253-255

- Gossett LA, Kelvin DJ, Sternberg EA, Olson EN (1989) A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. *Mol Cell Biol* 9(11): 5022-5033
- Haider AW, Larson MG, Benjamin EJ, Levy D (1998) Increased left ventricular mass and hypertrophy are associated with increased risk for sudden death. *J Am Coll Cardiol* 32(5): 1454-1459
- Hammes A, Oberdorf S, Strehler EE, Stauffer T, Carafoli E, Vetter H, Neyses L (1994) Differentiation-specific isoform mRNA expression of the calmodulin-dependent plasma membrane Ca(2+)-ATPase. *FASEB J* 8(6): 428-435
- Han J, Jiang Y, Li Z, Kravchenko VV, Ulevitch RJ (1997) Activation of the transcription factor MEF2C by the MAP kinase p38 in inflammation. *Nature* 386(6622): 296-299
- Han S, Lu J, Zhang Y, Cheng C, Han L, Wang X, Li L, Liu C, Huang B (2006) Recruitment of histone deacetylase 4 by transcription factors represses interleukin-5 transcription. *Biochem J* 400(3): 439-448
- Hao J, Ju H, Zhao S, Junaid A, Scammell-La Fleur T, Dixon IM (1999) Elevation of expression of Smads 2, 3, and 4, decorin and TGF-beta in the chronic phase of myocardial infarct scar healing. *J Mol Cell Cardiol* 31(3): 667-678
- Hao J, Wang B, Jones SC, Jassal DS, Dixon IM (2000) Interaction between angiotensin II and Smad proteins in fibroblasts in failing heart and in vitro. *Am J Physiol Heart Circ Physiol* 279(6): H3020-3030
- Haq S, Choukroun G, Kang ZB, Ranu H, Matsui T, Rosenzweig A, Molkentin JD, Alessandrini A, Woodgett J, Hajjar R, Michael A, Force T (2000) Glycogen synthase kinase-3beta is a negative regulator of cardiomyocyte hypertrophy. *J Cell Biol* 151(1): 117-130
- Hasegawa K, Lee SJ, Jobe SM, Markham BE, Kitsis RN (1997) cis-Acting sequences that mediate induction of beta-myosin heavy chain gene expression during left ventricular hypertrophy due to aortic constriction. *Circulation* 96(11): 3943-3953
- Hautala N, Tokola H, Luodonpaa M, Puhakka J, Romppanen H, Vuolteenaho O, Ruskoaho H (2001) Pressure overload increases GATA4 binding activity via endothelin-1. *Circulation* 103(5): 730-735
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39(3): 311-318

Hendrix JA, Wamhoff BR, McDonald OG, Sinha S, Yoshida T, Owens GK (2005) 5' CArG degeneracy in smooth muscle alpha-actin is required for injury-induced gene suppression in vivo. *J Clin Invest* 115(2): 418-427

Herzig TC, Jobe SM, Aoki H, Molkentin JD, Cowley AW, Jr., Izumo S, Markham BE (1997) Angiotensin II type1a receptor gene expression in the heart: AP-1 and GATA-4 participate in the response to pressure overload. *Proc Natl Acad Sci U S A* 94(14): 7543-7548

Hill CS, Wynne J, Treisman R (1995) The Rho family GTPases RhoA, Rac1, and CDC42Hs regulate transcriptional activation by SRF. *Cell* 81(7): 1159-1170

Hogan PG, Chen L, Nardone J, Rao A (2003) Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes Dev* 17(18): 2205-2232

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298(5591): 129-149

Iezzi S, Di Padova M, Serra C, Caretti G, Simone C, Maklan E, Minetti G, Zhao P, Hoffman EP, Puri PL, Sartorelli V (2004) Deacetylase inhibitors increase muscle cell size by promoting myoblast recruitment and fusion through induction of follistatin. *Dev Cell* 6(5): 673-684

Imbalzano AN, Kwon H, Green MR, Kingston RE (1994) Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* 370(6489): 481-485

Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293(5532): 1074-1080

- Kaczynski J, Cook T, Urrutia R (2003) Sp1- and Kruppel-like transcription factors. *Genome Biol* 4(2): 206
- Kaneda R, Ueno S, Yamashita Y, Choi YL, Koinuma K, Takada S, Wada T, Shimada K, Mano H (2005) Genome-wide screening for target regions of histone deacetylases in cardiomyocytes. *Circ Res* 97(3): 210-218
- Kasahara H, Bartunkova S, Schinke M, Tanaka M, Izumo S (1998) Cardiac and extracardiac expression of Csx/Nkx2.5 homeodomain protein. *Circ Res* 82(9): 936-946
- Kato Y, Kravchenko VV, Tapping RI, Han J, Ulevitch RJ, Lee JD (1997) BMK1/ERK5 regulates serum-induced early gene expression through transcription factor MEF2C. *EMBO J* 16(23): 7054-7066
- Kawamura T, Ono K, Morimoto T, Wada H, Hirai M, Hidaka K, Morisaki T, Heike T, Nakahata T, Kita T, Hasegawa K (2005) Acetylation of GATA-4 is involved in the differentiation of embryonic stem cells into cardiac myocytes. *J Biol Chem* 280(20): 19682-19688
- Kee HJ, Sohn IS, Nam KI, Park JE, Qian YR, Yin Z, Ahn Y, Jeong MH, Bang YJ, Kim N, Kim JK, Kim KK, Epstein JA, Kook H (2006) Inhibition of histone deacetylation blocks cardiac hypertrophy induced by angiotensin II infusion and aortic banding. *Circulation* 113(1): 51-59
- Kerkela R, Pikkarainen S, Majalahti-Palviainen T, Tokola H, Ruskoaho H (2002) Distinct roles of mitogen-activated protein kinase pathways in GATA-4 transcription factor-mediated regulation of B-type natriuretic peptide gene. *J Biol Chem* 277(16): 13752-13760
- Kimes BW, Brandt BL (1976) Properties of a clonal muscle cell line from rat heart. *Exp Cell Res* 98(2): 367-381
- Kitta K, Clement SA, Remeika J, Blumberg JB, Suzuki YJ (2001) Endothelin-1 induces phosphorylation of GATA-4 transcription factor in the HL-1 atrial-muscle cell line. *Biochem J* 359(Pt 2): 375-380
- Klein L, Gheorghiade M (2003) Therapeutic strategies for patients hospitalized with worsening heart failure. *Ital Heart J* 4(2): 71-74
- Knuppel R, Dietze P, Lehnberg W, Frech K, Wingender E (1994) TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *J Comput Biol* 1(3): 191-198
- Kobayashi S, Lackey T, Huang Y, Bisping E, Pu WT, Boxer LM, Liang Q (2006) Transcription factor gata4 regulates cardiac BCL2 gene expression in vitro and in vivo. *FASEB J* 20(6): 800-802

Kolodziejczyk SM, Wang L, Balazsi K, DeRepentigny Y, Kothary R, Megeney LA (1999) MEF2 is upregulated during cardiac hypertrophy and is required for normal post-natal growth of the myocardium. *Curr Biol* 9(20): 1203-1206

Komuro I, Izumo S (1993) Csx: a murine homeobox-containing gene specifically expressed in the developing heart. *Proc Natl Acad Sci U S A* 90(17): 8145-8149

Kook H, Itoh H, Choi BS, Sawada N, Doi K, Hwang TJ, Kim KK, Arai H, Baik YH, Nakao K (2003) Physiological concentration of atrial natriuretic peptide induces endothelial regeneration in vitro. *Am J Physiol Heart Circ Physiol* 284(4): H1388-1397

Kook H, Lee J, Kim SW, Baik YH (2002) Augmented natriuretic peptide-induced guanylyl cyclase activity and vasodilation in experimental hyperglycemic rats. *Jpn J Pharmacol* 88(2): 167-173

Krejci A, Bruce AW, Dolezal V, Tucek S, Buckley NJ (2004) Multiple promoters drive tissue-specific expression of the human M muscarinic acetylcholine receptor gene. *J Neurochem* 91(1): 88-98

Kuo H, Chen J, Ruiz-Lozano P, Zou Y, Nemer M, Chien KR (1999) Control of segmental expression of the cardiac-restricted ankyrin repeat protein gene by distinct regulatory pathways in murine cardiogenesis. *Development* 126(19): 4223-4234

Kuwahara K, Barrientos T, Pipes GC, Li S, Olson EN (2005) Muscle-specific signaling mechanism that links actin dynamics to serum response factor. *Mol Cell Biol* 25(8): 3173-3181

Kuwahara K, Teg Pipes GC, McAnally J, Richardson JA, Hill JA, Bassel-Duby R, Olson EN (2007) Modulation of adverse cardiac remodeling by STARS, a mediator of MEF2 signaling and SRF activity. *J Clin Invest* 117(5): 1324-1334

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczký J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM,

Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglu S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822): 860-921

Lassar AB, Buskin JN, Lockshon D, Davis RL, Apone S, Hauschka SD, Weintraub H (1989) MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. *Cell* 58(5): 823-831

Lavallee G, Andelfinger G, Nadeau M, Lefebvre C, Nemer G, Horb ME, Nemer M (2006) The Kruppel-like transcription factor KLF13 is a novel regulator of heart development. *EMBO J* 25(21): 5201-5213

Lee Y, Shioi T, Kasahara H, Jobe SM, Wiese RJ, Markham BE, Izumo S (1998) The cardiac tissue-restricted homeobox protein Csx/Nkx2.5 physically associates with the zinc finger protein GATA4 and cooperatively activates atrial natriuretic factor gene expression. *Mol Cell Biol* 18(6): 3120-3129

Lenhard B, Sandelin A, Mendoza L, Engstrom P, Jareborg N, Wasserman WW (2003) Identification of conserved regulatory elements by comparative genome analysis. *J Biol* 2(2): 13

Levine M, Tjian R (2003) Transcription regulation and animal diversity. *Nature* 424 (6945): 147-151

Li H, Capetanaki Y (1994) An E box in the desmin promoter cooperates with the E box and MEF-2 sites of a distal enhancer to direct muscle-specific transcription. *EMBO J* 13 (15): 3580-3589

- Li S, Czubryt MP, McAnally J, Bassel-Duby R, Richardson JA, Wiebel FF, Nordheim A, Olson EN (2005) Requirement for serum response factor for skeletal muscle growth and maturation revealed by tissue-specific gene deletion in mice. *Proc Natl Acad Sci U S A* 102(4): 1082-1087
- Liang F, Gardner DG (1999) Mechanical strain activates BNP gene transcription through a p38/NF-kappaB-dependent mechanism. *J Clin Invest* 104(11): 1603-1612
- Liang F, Lu S, Gardner DG (2000) Endothelin-dependent and -independent components of strain-activated brain natriuretic peptide gene transcription require extracellular signal regulated kinase and p38 mitogen-activated protein kinase. *Hypertension* 35(1 Pt 2): 188-192
- Liang Q, Bueno OF, Wilkins BJ, Kuan CY, Xia Y, Molkentin JD (2003) c-Jun N-terminal kinases (JNK) antagonize cardiac growth through cross-talk with calcineurin-NFAT signaling. *EMBO J* 22(19): 5079-5089
- Liang Q, De Windt LJ, Witt SA, Kimball TR, Markham BE, Molkentin JD (2001a) The transcription factors GATA4 and GATA6 regulate cardiomyocyte hypertrophy in vitro and in vivo. *J Biol Chem* 276(32): 30245-30253
- Liang Q, Molkentin JD (2003) Redefining the roles of p38 and JNK signaling in cardiac hypertrophy: dichotomy between cultured myocytes and animal models. *J Mol Cell Cardiol* 35(12): 1385-1394
- Liang Q, Wiese RJ, Bueno OF, Dai YS, Markham BE, Molkentin JD (2001b) The transcription factor GATA4 is activated by extracellular signal-regulated kinase 1- and 2-mediated phosphorylation of serine 105 in cardiomyocytes. *Mol Cell Biol* 21(21): 7460-7469
- Liao P, Georgakopoulos D, Kovacs A, Zheng M, Lerner D, Pu H, Saffitz J, Chien K, Xiao RP, Kass DA, Wang Y (2001) The in vivo role of p38 MAP kinases in cardiac remodeling and restrictive cardiomyopathy. *Proc Natl Acad Sci U S A* 98(21): 12283-12288
- Lien CL, McAnally J, Richardson JA, Olson EN (2002) Cardiac-specific activity of an Nkx2-5 enhancer requires an evolutionarily conserved Smad binding site. *Dev Biol* 244(2): 257-266
- Lints TJ, Parsons LM, Hartley L, Lyons I, Harvey RP (1993) Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development* 119(2): 419-431

- Lips DJ, deWindt LJ, van Kraaij DJ, Doevendans PA (2003) Molecular determinants of myocardial hypertrophy and failure: alternative pathways for beneficial and maladaptive hypertrophy. *Eur Heart J* 24(10): 883-896
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4): 402-408
- Lomvardas S, Thanos D (2001) Nucleosome sliding via TBP DNA binding in vivo. *Cell* 106(6): 685-696
- Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* 288(5463): 136-140
- Loots GG, Ovcharenko I, Pachter L, Dubchak I, Rubin EM (2002) rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res* 12(5): 832-839
- Lorell BH, Carabello BA (2000) Left ventricular hypertrophy: pathogenesis, detection, and prognosis. *Circulation* 102(4): 470-479
- Lu J, McKinsey TA, Nicol RL, Olson EN (2000) Signal-dependent activation of the MEF2 transcription factor by dissociation from histone deacetylases. *Proc Natl Acad Sci U S A* 97(8): 4070-4075
- Mahadeva H, Brooks G, Lodwick D, Chong NW, Samani NJ (2002) ms1, a novel stress-responsive, muscle-specific gene that is up-regulated in the early stages of pressure overload-induced left ventricular hypertrophy. *FEBS Lett* 521(1-3): 100-104
- Mahadeva H, Starkey MP, Sheikh FN, Mundy CR, Samani NJ (1998) A simple and efficient method for the isolation of differentially expressed genes. *J Mol Biol* 284(5): 1391-1398
- Majalahti T, Suo-Palosaari M, Sarman B, Hautala N, Pikkarainen S, Tokola H, Vuolteenaho O, Wang J, Paradis P, Nemer M, Ruskoaho H (2007) Cardiac BNP gene activation by angiotensin II in vivo. *Mol Cell Endocrinol* 273(1-2): 59-67
- Mal A, Harter ML (2003) MyoD is functionally linked to the silencing of a muscle-specific regulatory gene prior to skeletal myogenesis. *Proc Natl Acad Sci U S A* 100(4): 1735-1739
- Mal A, Sturniolo M, Schiltz RL, Ghosh MK, Harter ML (2001) A role for histone deacetylase HDAC1 in modulating the transcriptional activity of MyoD: inhibition of the myogenic program. *EMBO J* 20(7): 1739-1753

- Marinissen MJ, Chiariello M, Pallante M, Gutkind JS (1999) A network of mitogen-activated protein kinases links G protein-coupled receptors to the c-jun promoter: a role for c-Jun NH2-terminal kinase, p38s, and extracellular signal-regulated kinase 5. *Mol Cell Biol* 19(6): 4289-4301
- Martin JF, Miano JM, Hustad CM, Copeland NG, Jenkins NA, Olson EN (1994) A Mef2 gene that generates a muscle-specific isoform via alternative mRNA splicing. *Mol Cell Biol* 14(3): 1647-1656
- Matsui T, Li L, Wu JC, Cook SA, Nagoshi T, Picard MH, Liao R, Rosenzweig A (2002) Phenotypic spectrum caused by transgenic overexpression of activated Akt in the heart. *J Biol Chem* 277(25): 22896-22901
- McBride K, Charron F, Lefebvre C, Nemer M (2003) Interaction with GATA transcription factors provides a mechanism for cell-specific effects of c-Fos. *Oncogene* 22(52): 8403-8412
- McKinsey TA, Zhang CL, Lu J, Olson EN (2000) Signal-dependent nuclear export of a histone deacetylase regulates muscle differentiation. *Nature* 408(6808): 106-111
- McKinsey TA, Zhang CL, Olson EN (2002) Signaling chromatin to make muscle. *Curr Opin Cell Biol* 14(6): 763-772
- Menard C, Pupier S, Mornet D, Kitzmann M, Nargeot J, Lory P (1999) Modulation of L-type calcium channel expression during retinoic acid-induced differentiation of H9C2 cardiac cells. *J Biol Chem* 274(41): 29063-29070
- Mercher T, Busson-Le Coniat M, Nguyen Khac F, Ballerini P, Mauchauffe M, Bui H, Pellegrino B, Radford I, Valensi F, Mugneret F, Dastugue N, Bernard OA, Berger R (2002) Recurrence of OTT-MAL fusion in t(1;22) of infant AML-M7. *Genes Chromosomes Cancer* 33(1): 22-28
- Mercher T, Coniat MB, Monni R, Mauchauffe M, Nguyen Khac F, Gressin L, Mugneret F, Leblanc T, Dastugue N, Berger R, Bernard OA (2001) Involvement of a human gene related to the Drosophila spen gene in the recurrent t(1;22) translocation of acute megakaryocytic leukemia. *Proc Natl Acad Sci U S A* 98(10): 5776-5779
- Miano JM (2003) Serum response factor: toggling between disparate programs of gene expression. *J Mol Cell Cardiol* 35(6): 577-593
- Michael A, Haq S, Chen X, Hsich E, Cui L, Walters B, Shao Z, Bhattacharya K, Kilter H, Huggins G, Andreucci M, Periasamy M, Solomon RN, Liao R, Patten R, Molkentin JD, Force T (2004) Glycogen synthase kinase-3 β regulates growth, calcium homeostasis, and diastolic function in the heart. *J Biol Chem* 279(20): 21383-21393

Minetti GC, Colussi C, Adami R, Serra C, Mozzetta C, Parente V, Fortuni S, Straino S, Sampaolesi M, Di Padova M, Illi B, Gallinari P, Steinkuhler C, Capogrossi MC, Sartorelli V, Bottinelli R, Gaetano C, Puri PL (2006) Functional and morphological recovery of dystrophic muscles in mice treated with deacetylase inhibitors. *Nat Med* 12 (10): 1147-1150

Miralles F, Posern G, Zaromytidou AI, Treisman R (2003) Actin dynamics control SRF activity by regulation of its coactivator MAL. *Cell* 113(3): 329-342

Miska EA, Karlsson C, Langley E, Nielsen SJ, Pines J, Kouzarides T (1999) HDAC4 deacetylase associates with and represses the MEF2 transcription factor. *EMBO J* 18 (18): 5099-5107

Molkentin JD (2000) The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *J Biol Chem* 275(50): 38949-38952

Molkentin JD (2004) Calcineurin-NFAT signaling regulates the cardiac hypertrophic response in coordination with the MAPKs. *Cardiovasc Res* 63(3): 467-475

Molkentin JD, Black BL, Martin JF, Olson EN (1995) Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* 83(7): 1125-1136

Molkentin JD, Dorn GW, 2nd (2001) Cytoplasmic signaling pathways that regulate cardiac hypertrophy. *Annu Rev Physiol* 63: 391-426

Molkentin JD, Lu JR, Antos CL, Markham B, Richardson J, Robbins J, Grant SR, Olson EN (1998) A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* 93(2): 215-228

Molkentin JD, Markham BE (1993) Myocyte-specific enhancer-binding factor (MEF-2) regulates alpha-cardiac myosin heavy chain gene expression in vitro and in vivo. *J Biol Chem* 268(26): 19512-19520

Molkentin JD, Olson EN (1996) Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. *Proc Natl Acad Sci U S A* 93(18): 9366-9373

Morimoto T, Hasegawa K, Kaburagi S, Kakita T, Wada H, Yanazume T, Sasayama S (2000) Phosphorylation of GATA-4 is involved in alpha 1-adrenergic agonist-responsive transcription of the endothelin-1 gene in cardiac myocytes. *J Biol Chem* 275(18): 13721-13726

Morin S, Charron F, Robitaille L, Nemer M (2000) GATA-dependent recruitment of MEF2 proteins to target promoters. *EMBO J* 19(9): 2046-2055

- Morin S, Paradis P, Aries A, Nemer M (2001) Serum response factor-GATA ternary complex required for nuclear signaling by a G-protein-coupled receptor. *Mol Cell Biol* 21(4): 1036-1044
- Morisco C, Condorelli G, Orzi F, Vigliotta G, Di Grezia R, Beguinot F, Trimarco B, Lembo G (2000) Insulin-stimulated cardiac glucose uptake is impaired in spontaneously hypertensive rats: role of early steps of insulin signalling. *J Hypertens* 18(4): 465-473
- Morisco C, Seta K, Hardt SE, Lee Y, Vatner SF, Sadoshima J (2001) Glycogen synthase kinase 3 β regulates GATA4 in cardiac myocytes. *J Biol Chem* 276(30): 28586-28597
- Moses KA, DeMayo F, Braun RM, Reecy JL, Schwartz RJ (2001) Embryonic expression of an Nkx2-5/Cre gene using ROSA26 reporter mice. *Genesis* 31(4): 176-180
- Murakami A, Ishida S, Dickson C (2002) GATA-4 interacts distinctively with negative and positive regulatory elements in the Fgf-3 promoter. *Nucleic Acids Res* 30(4): 1056-1064
- Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB, et al. (1989) Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 58(3): 537-544
- Nadruz W, Jr., Corat MA, Marin TM, Guimaraes Pereira GA, Franchini KG (2005) Focal adhesion kinase mediates MEF2 and c-Jun activation by stretch: role in the activation of the cardiac hypertrophic genetic program. *Cardiovasc Res* 68(1): 87-97
- Nadruz W, Jr., Kobarg CB, Constancio SS, Corat PD, Franchini KG (2003) Load-induced transcriptional activation of c-jun in rat myocardium: regulation by myocyte enhancer factor 2. *Circ Res* 92(2): 243-251
- Nemoto S, Sheng Z, Lin A (1998) Opposing effects of Jun kinase and p38 mitogen-activated protein kinases on cardiomyocyte hypertrophy. *Mol Cell Biol* 18(6): 3518-3526
- Niu Z, Yu W, Zhang SX, Barron M, Belaguli NS, Schneider MD, Parmacek M, Nordheim A, Schwartz RJ (2005) Conditional mutagenesis of the murine serum response factor gene blocks cardiogenesis and the transcription of downstream gene targets. *J Biol Chem* 280(37): 32531-32538
- Norman C, Runswick M, Pollock R, Treisman R (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 55(6): 989-1003

- Oh J, Richardson JA, Olson EN (2005) Requirement of myocardin-related transcription factor-B for remodeling of branchial arch arteries and smooth muscle differentiation. *Proc Natl Acad Sci U S A* 102(42): 15122-15127
- Oh J, Wang Z, Wang DZ, Lien CL, Xing W, Olson EN (2004) Target gene-specific modulation of myocardin activity by GATA transcription factors. *Mol Cell Biol* 24(19): 8519-8528
- Ohkawa Y, Marfella CG, Imbalzano AN (2006) Skeletal muscle specification by myogenin and Mef2D via the SWI/SNF ATPase Brg1. *EMBO J* 25(3): 490-501
- Ohkawa Y, Yoshimura S, Higashi C, Marfella CG, Dacwag CS, Tachibana T, Imbalzano AN (2007) Myogenin and the SWI/SNF ATPase Brg1 maintain myogenic gene expression at different stages of skeletal myogenesis. *J Biol Chem* 282(9): 6564-6570
- Oka T, Maillet M, Watt AJ, Schwartz RJ, Aronow BJ, Duncan SA, Molkentin JD (2006) Cardiac-specific deletion of Gata4 reveals its requirement for hypertrophy, compensation, and myocyte viability. *Circ Res* 98(6): 837-845
- Oka T, Xu J, Molkentin JD (2007) Re-employment of developmental transcription factors in adult heart disease. *Semin Cell Dev Biol* 18(1): 117-131
- Olson EN (2004) A decade of discoveries in cardiac biology. *Nat Med* 10(5): 467-474
- Olson EN, Schneider MD (2003) Sizing up the heart: development redux in disease. *Genes Dev* 17(16): 1937-1956
- Owens GK, Kumar MS, Wamhoff BR (2004) Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiol Rev* 84(3): 767-801
- Parlakian A, Charvet C, Escoubet B, Mericskay M, Molkentin JD, Gary-Bobo G, De Windt LJ, Ludosky MA, Paulin D, Daegelen D, Tuil D, Li Z (2005) Temporally controlled onset of dilated cardiomyopathy through disruption of the SRF gene in adult heart. *Circulation* 112(19): 2930-2939
- Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29(9): e45
- Pikkarainen S, Tokola H, Kerkela R, Ruskoaho H (2004) GATA transcription factors in the developing and adult heart. *Cardiovasc Res* 63(2): 196-207
- Pu WT, Ishiwata T, Juraszek AL, Ma Q, Izumo S (2004) GATA4 is a dosage-sensitive regulator of cardiac morphogenesis. *Dev Biol* 275(1): 235-244

- Punn A, Mockridge JW, Farooqui S, Marber MS, Heads RJ (2000) Sustained activation of p42/p44 mitogen-activated protein kinase during recovery from simulated ischaemia mediates adaptive cytoprotection in cardiomyocytes. *Biochem J* 350 Pt 3: 891-899
- Qiu P, Feng XH, Li L (2003) Interaction of Smad3 and SRF-associated complex mediates TGF-beta1 signals to regulate SM22 transcription during myofibroblast differentiation. *J Mol Cell Cardiol* 35(12): 1407-1420
- Rajagopal SK, Ma Q, Obler D, Shen J, Manichaikul A, Tomita-Mitchell A, Boardman K, Briggs C, Garg V, Srivastava D, Goldmuntz E, Broman KW, Benson DW, Smoot LB, Pu WT (2007) Spectrum of heart disease associated with murine and human GATA4 mutation. *J Mol Cell Cardiol* 43(6): 677-685
- Riley PR, Gertsenstein M, Dawson K, Cross JC (2000) Early exclusion of hand1-deficient cells from distinct regions of the left ventricular myocardium in chimeric mouse embryos. *Dev Biol* 227(1): 156-168
- Roeder RG (2005) Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett* 579(4): 909-915
- Rosenkranz S, Flesch M, Amann K, Haeuseler C, Kilter H, Seeland U, Schluter KD, Bohm M (2002) Alterations of beta-adrenergic signaling and cardiac hypertrophy in transgenic mice overexpressing TGF-beta(1). *Am J Physiol Heart Circ Physiol* 283(3): H1253-1262
- Rossini AA, Like AA, Chick WL, Appel MC, Cahill GF, Jr. (1977a) Studies of streptozotocin-induced insulinitis and diabetes. *Proc Natl Acad Sci U S A* 74(6): 2485-2489
- Rossini AA, Like AA, Dulin WE, Cahill GF, Jr. (1977b) Pancreatic beta cell toxicity by streptozotocin anomers. *Diabetes* 26(12): 1120-1124
- Rothermel BA, Berenji K, Tannous P, Kutschke W, Dey A, Nolan B, Yoo KD, Demetroulis E, Gimbel M, Cabuay B, Karimi M, Hill JA (2005) Differential activation of stress-response signaling in load-induced cardiac hypertrophy and failure. *Physiol Genomics* 23(1): 18-27
- Roulet E, Fisch I, Junier T, Bucher P, Mermod N (1998) Evaluation of computer tools for the prediction of transcription factor binding sites on genomic DNA. *In Silico Biol* 1(1): 21-28
- Saadane N, Alpert L, Chalifour LE (1999) Expression of immediate early genes, GATA-4, and Nkx-2.5 in adrenergic-induced cardiac hypertrophy and during regression in adult mice. *Br J Pharmacol* 127(5): 1165-1176

- Sahai E, Alberts AS, Treisman R (1998) RhoA effector mutants reveal distinct effector pathways for cytoskeletal reorganization, SRF activation and transformation. *EMBO J* 17(5): 1350-1361
- Salma N, Xiao H, Mueller E, Imbalzano AN (2004) Temporal recruitment of transcription factors and SWI/SNF chromatin-remodeling enzymes during adipogenic induction of the peroxisome proliferator-activated receptor gamma nuclear hormone receptor. *Mol Cell Biol* 24(11): 4651-4663
- Sanbe A, Gulick J, Hanks MC, Liang Q, Osinska H, Robbins J (2003) Reengineering inducible cardiac-specific transgenesis with an attenuated myosin heavy chain promoter. *Circ Res* 92(6): 609-616
- Sartorelli V, Puri PL, Hamamori Y, Ogryzko V, Chung G, Nakatani Y, Wang JY, Kedes L (1999) Acetylation of MyoD directed by PCAF is necessary for the execution of the muscle program. *Mol Cell* 4(5): 725-734
- Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, Maron BJ, Seidman CE, Seidman JG (1998) Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* 281(5373): 108-111
- Schubert W, Yang XY, Yang TT, Factor SM, Lisanti MP, Molkentin JD, Rincon M, Chow CW (2003) Requirement of transcription factor NFAT in developing atrial myocardium. *J Cell Biol* 161(5): 861-874
- Schultheiss TM, Burch JB, Lassar AB (1997) A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev* 11(4): 451-462
- Schultz Jel J, Witt SA, Glascock BJ, Nieman ML, Reiser PJ, Nix SL, Kimball TR, Doetschman T (2002) TGF-beta1 mediates the hypertrophic cardiomyocyte growth induced by angiotensin II. *J Clin Invest* 109(6): 787-796
- Schulz RA, Yutzey KE (2004) Calcineurin signaling and NFAT activation in cardiovascular and skeletal muscle development. *Dev Biol* 266(1): 1-16
- Selvaraj A, Prywes R (2003) Megakaryoblastic leukemia-1/2, a transcriptional co-activator of serum response factor, is required for skeletal myogenic differentiation. *J Biol Chem* 278(43): 41977-41987
- Selvaraj A, Prywes R (2004) Expression profiling of serum inducible genes identifies a subset of SRF target genes that are MKL dependent. *BMC Mol Biol* 5: 13
- Sepulveda JL, Belaguli N, Nigam V, Chen CY, Nemer M, Schwartz RJ (1998) GATA-4 and Nkx-2.5 coactivate Nkx-2 DNA binding targets: role for regulating early cardiac gene expression. *Mol Cell Biol* 18(6): 3405-3415

- Sepulveda JL, Vlahopoulos S, Iyer D, Belaguli N, Schwartz RJ (2002) Combinatorial expression of GATA4, Nkx2-5, and serum response factor directs early cardiac gene activity. *J Biol Chem* 277(28): 25775-25782
- Shin CH, Liu ZP, Passier R, Zhang CL, Wang DZ, Harris TM, Yamagishi H, Richardson JA, Childs G, Olson EN (2002) Modulation of cardiac growth and development by HOP, an unusual homeodomain protein. *Cell* 110(6): 725-735
- Shiojima I, Komuro I, Oka T, Hiroi Y, Mizuno T, Takimoto E, Monzen K, Aikawa R, Akazawa H, Yamazaki T, Kudoh S, Yazaki Y (1999) Context-dependent transcriptional cooperation mediated by cardiac transcription factors Csx/Nkx-2.5 and GATA-4. *J Biol Chem* 274(12): 8231-8239
- Shore P, Sharrocks AD (1995) The MADS-box family of transcription factors. *Eur J Biochem* 229(1): 1-13
- Shyu KG, Ko WH, Yang WS, Wang BW, Kuan P (2005) Insulin-like growth factor-1 mediates stretch-induced upregulation of myostatin expression in neonatal rat cardiomyocytes. *Cardiovasc Res* 68(3): 405-414
- Sipido KR, Marban E (1991) L-type calcium channels, potassium channels, and novel nonspecific cation channels in a clonal muscle cell line derived from embryonic rat ventricle. *Circ Res* 69(6): 1487-1499
- Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. *Annu Rev Biochem* 72: 449-479
- Song K, Backs J, McAnally J, Qi X, Gerard RD, Richardson JA, Hill JA, Bassel-Duby R, Olson EN (2006) The transcriptional coactivator CAMTA2 stimulates cardiac growth by opposing class II histone deacetylases. *Cell* 125(3): 453-466
- Sotiropoulos A, Gineitis D, Copeland J, Treisman R (1999) Signal-regulated activation of serum response factor is mediated by changes in actin dynamics. *Cell* 98(2): 159-169
- Soutoglou E, Talianidis I (2002) Coordination of PIC assembly and chromatin remodeling during differentiation-induced gene activation. *Science* 295(5561): 1901-1904
- Sparrow DB, Miska EA, Langley E, Reynaud-Deonauth S, Kotecha S, Towers N, Spohr G, Kouzarides T, Mohun TJ (1999) MEF-2 function is modified by a novel co-repressor, MITR. *EMBO J* 18(18): 5085-5098
- Srivastava D (1999) HAND proteins: molecular mediators of cardiac development and congenital heart disease. *Trends Cardiovasc Med* 9(1-2): 11-18

- Srivastava D, Thomas T, Lin Q, Kirby ML, Brown D, Olson EN (1997) Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat Genet* 16(2): 154-160
- Stanley EG, Biben C, Elefanty A, Barnett L, Koentgen F, Robb L, Harvey RP (2002) Efficient Cre-mediated deletion in cardiac progenitor cells conferred by a 3'UTR-ires-Cre allele of the homeobox gene Nkx2-5. *Int J Dev Biol* 46(4): 431-439
- Stormo GD, Schneider TD, Gold L, Ehrenfeucht A (1982) Use of the 'Perceptron' algorithm to distinguish translational initiation sites in E. coli. *Nucleic Acids Res* 10(9): 2997-3011
- Sugden PH, Clerk A (1998a) Cellular mechanisms of cardiac hypertrophy. *J Mol Med* 76(11): 725-746
- Sugden PH, Clerk A (1998b) Regulation of mitogen-activated protein kinase cascades in the heart. *Adv Enzyme Regul* 38: 87-98
- Suske G, Bruford E, Philipsen S (2005) Mammalian SP/KLF transcription factors: bring in the family. *Genomics* 85(5): 551-556
- Takimoto E, Mizuno T, Terasaki F, Shimoyama M, Honda H, Shiojima I, Hiroi Y, Oka T, Hayashi D, Hirai H, Kudoh S, Toko H, Kawamura K, Nagai R, Yazaki Y, Komuro I (2000) Up-regulation of natriuretic peptides in the ventricle of Csx/Nkx2-5 transgenic mice. *Biochem Biophys Res Commun* 270(3): 1074-1079
- Temsah R, Nemer M (2005) GATA factors and transcriptional regulation of cardiac natriuretic peptide genes. *Regul Pept* 128(3): 177-185
- ten Dijke P, Hill CS (2004) New insights into TGF-beta-Smad signalling. *Trends Biochem Sci* 29(5): 265-273
- Thompson JT, Rackley MS, O'Brien TX (1998) Upregulation of the cardiac homeobox gene Nkx2-5 (CSX) in feline right ventricular pressure overload. *Am J Physiol* 274(5 Pt 2): H1569-1573
- Tomczak KK, Marinescu VD, Ramoni MF, Sanoudou D, Montanaro F, Han M, Kunkel LM, Kohane IS, Beggs AH (2004) Expression profiling and identification of novel genes involved in myogenic differentiation. *FASEB J* 18(2): 403-405
- Uchida S, Tanaka Y, Ito H, Saitoh-Ohara F, Inazawa J, Yokoyama KK, Sasaki S, Marumo F (2000) Transcriptional regulation of the CLC-K1 promoter by myc-associated zinc finger protein and kidney-enriched Kruppel-like factor, a novel zinc finger repressor. *Mol Cell Biol* 20(19): 7319-7331

- Ureta-Vidal A, Ettwiller L, Birney E (2003) Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nat Rev Genet* 4(4): 251-262
- van Vliet J, Crofts LA, Quinlan KG, Czolij R, Perkins AC, Crossley M (2006) Human KLF17 is a new member of the Sp/KLF family of transcription factors. *Genomics* 87(4): 474-482
- Vandromme M, Gauthier-Rouviere C, Carnac G, Lamb N, Fernandez A (1992) Serum response factor p67SRF is expressed and required during myogenic differentiation of both mouse C2 and rat L6 muscle cell lines. *J Cell Biol* 118(6): 1489-1500
- Walsh K (2006) Akt signaling and growth of the heart. *Circulation* 113(17): 2032-2034
- Wang AH, Bertos NR, Vezmar M, Pelletier N, Crosato M, Heng HH, Th'ng J, Han J, Yang XJ (1999) HDAC4, a human histone deacetylase related to yeast HDA1, is a transcriptional corepressor. *Mol Cell Biol* 19(11): 7816-7827
- Wang B, Hao J, Jones SC, Yee MS, Roth JC, Dixon IM (2002a) Decreased Smad 7 expression contributes to cardiac fibrosis in the infarcted rat heart. *Am J Physiol Heart Circ Physiol* 282(5): H1685-1696
- Wang D, Chang PS, Wang Z, Sutherland L, Richardson JA, Small E, Krieg PA, Olson EN (2001) Activation of cardiac gene expression by myocardin, a transcriptional cofactor for serum response factor. *Cell* 105(7): 851-862
- Wang DZ, Li S, Hockemeyer D, Sutherland L, Wang Z, Schratt G, Richardson JA, Nordheim A, Olson EN (2002b) Potentiation of serum response factor activity by a family of myocardin-related transcription factors. *Proc Natl Acad Sci U S A* 99(23): 14855-14860
- Wang J, Paradis P, Aries A, Komati H, Lefebvre C, Wang H, Nemer M (2005a) Convergence of protein kinase C and JAK-STAT signaling on transcription factor GATA-4. *Mol Cell Biol* 25(22): 9829-9844
- Wang J, Xu N, Feng X, Hou N, Zhang J, Cheng X, Chen Y, Zhang Y, Yang X (2005b) Targeted disruption of Smad4 in cardiomyocytes results in cardiac hypertrophy and heart failure. *Circ Res* 97(8): 821-828
- Wang Y, Huang S, Sah VP, Ross J, Jr., Brown JH, Han J, Chien KR (1998) Cardiac muscle cell hypertrophy and apoptosis induced by distinct members of the p38 mitogen-activated protein kinase family. *J Biol Chem* 273(4): 2161-2168
- Wasserman WW, Sandelin A (2004) Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* 5(4): 276-287

- Watt AJ, Battle MA, Li J, Duncan SA (2004) GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. *Proc Natl Acad Sci U S A* 101(34): 12573-12578
- Wei L, Zhou W, Croissant JD, Johansen FE, Prywes R, Balasubramanyam A, Schwartz RJ (1998) RhoA signaling via serum response factor plays an obligatory role in myogenic differentiation. *J Biol Chem* 273(46): 30287-30294
- Wilkins BJ, Dai YS, Bueno OF, Parsons SA, Xu J, Plank DM, Jones F, Kimball TR, Molkentin JD (2004) Calcineurin/NFAT coupling participates in pathological, but not physiological, cardiac hypertrophy. *Circ Res* 94(1): 110-118
- Wilkins BJ, De Windt LJ, Bueno OF, Braz JC, Glascock BJ, Kimball TF, Molkentin JD (2002) Targeted disruption of NFATc3, but not NFATc4, reveals an intrinsic defect in calcineurin-mediated cardiac hypertrophic growth. *Mol Cell Biol* 22(21): 7603-7613
- Wilkins BJ, Molkentin JD (2004) Calcium-calcineurin signaling in the regulation of cardiac hypertrophy. *Biochem Biophys Res Commun* 322(4): 1178-1191
- Xiao G, Mao S, Baumgarten G, Serrano J, Jordan MC, Roos KP, Fishbein MC, MacLellan WR (2001) Inducible activation of c-Myc in adult myocardium in vivo provokes cardiac myocyte hypertrophy and reactivation of DNA synthesis. *Circ Res* 89(12): 1122-1129
- Xing W, Zhang TC, Cao D, Wang Z, Antos CL, Li S, Wang Y, Olson EN, Wang DZ (2006) Myocardin induces cardiomyocyte hypertrophy. *Circ Res* 98(8): 1089-1097
- Xu J, Gong NL, Bodi I, Aronow BJ, Backx PH, Molkentin JD (2006a) Myocyte enhancer factors 2A and 2C induce dilated cardiomyopathy in transgenic mice. *J Biol Chem* 281(14): 9152-9162
- Xu L, Renaud L, Muller JG, Baicu CF, Bonnema DD, Zhou H, Kappler CS, Kubalak SW, Zile MR, Conway SJ, Menick DR (2006b) Regulation of Ncx1 expression. Identification of regulatory elements mediating cardiac-specific expression and up-regulation. *J Biol Chem* 281(45): 34430-34440
- Yaffe D, Saxel O (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* 270(5639): 725-727
- Yanazume T, Hasegawa K, Wada H, Morimoto T, Abe M, Kawamura T, Sasayama S (2002) Rho/ROCK pathway contributes to the activation of extracellular signal-regulated kinase/GATA-4 during myocardial cell hypertrophy. *J Biol Chem* 277(10): 8618-8625
- Yanazume T, Morimoto T, Wada H, Kawamura T, Hasegawa K (2003) Biological role of p300 in cardiac myocytes. *Mol Cell Biochem* 248(1-2): 115-119

Zechner D, Thuerauf DJ, Hanford DS, McDonough PM, Glembotski CC (1997) A role for the p38 mitogen-activated protein kinase pathway in myocardial cell growth, sarcomeric organization, and cardiac-specific gene expression. *J Cell Biol* 139(1): 115-127

Zeisberg EM, Ma Q, Juraszek AL, Moses K, Schwartz RJ, Izumo S, Pu WT (2005) Morphogenesis of the right ventricle requires myocardial expression of Gata4. *J Clin Invest* 115(6): 1522-1531

Zhang CL, McKinsey TA, Chang S, Antos CL, Hill JA, Olson EN (2002) Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell* 110(4): 479-488

Zhang X, Azhar G, Chai J, Sheridan P, Nagano K, Brown T, Yang J, Khrapko K, Borrás AM, Lawitts J, Misra RP, Wei JY (2001a) Cardiomyopathy in transgenic mice with cardiac-specific overexpression of serum response factor. *Am J Physiol Heart Circ Physiol* 280(4): H1782-1792

Zhang X, Chai J, Azhar G, Sheridan P, Borrás AM, Furr MC, Khrapko K, Lawitts J, Misra RP, Wei JY (2001b) Early postnatal cardiac changes and premature death in transgenic mice overexpressing a mutant form of serum response factor. *J Biol Chem* 276(43): 40033-40040