

**H3K36me3 in muscle differentiation:
Regulation of tissue-specific gene expression
by H3K36-specific histone methyltransferases**

Tarunpreet Dhaliwal

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in
partial fulfillment of the requirements for the Masters of Science program in
Cellular and Molecular Medicine

Department of Cellular and Molecular Medicine
Faculty of Medicine
University of Ottawa
September 2012

© Tarunpreet Dhaliwal, Ottawa, Canada, 2012.

ABSTRACT

The dynamic changes in chromatin play a significant role in lineage commitment and differentiation. These epigenetic modifications control gene expression through recruitment of transcription factors. While the active mark H3K4me3 is present around the transcription start site on the gene, the function of the H3K36me3 mark is unknown. A number of H3K36-specific histone methyltransferases (HMTs) have been identified, however the focus of this study is the HMT Hypb. To elucidate the role of H3K36me3 in mediating expression of developmentally-regulated loci, native chromatin immunoprecipitation (N-ChIP) was performed at a subset of genes. Upon differentiation, we observe that H3K36me3 becomes enriched at the 3' end of several muscle-specific genes. To further investigate the role of H3K36me3 in myogenesis, a lentiviral-mediated knockdown of the H3K36 HMT Hypb was performed in muscle myoblasts using shRNA. Upon Hypb knockdown, we were surprised to observe enhanced myogenesis. N-ChIP was also performed on differentiated Hypb knockdown cell lines in order to look at H3K36me3 enrichment on genes involved in muscle differentiation. N-ChIP data show a drop in H3K36me3 enrichment levels on myogenin and Ckm genes. The possible occupancy of Hypb on the coding regions of muscle-specific genes was experimentally observed by cross-linked chromatin immunoprecipitation (X-ChIP) on differentiated C2C12 cells and subsequently confirmed by X-ChIP on knockdown

lines where the occupancy was lost. A model is proposed that links the observed phenotype with H3K36me3.

TABLE OF CONTENTS

ABSTRACT	ii-iii
TABLE OF CONTENTS	iii-vi
LIST OF FIGURES AND TABLES	vii-ix
LIST OF ABBREVIATIONS	x-xi
ACKNOWLEDGMENT	xii

CHAPTER 1

1. INTRODUCTION	1-13
1.1 The genome and epigenome	1
1.2 Muscle differentiation	2-3
1.3 Chromatin organization and histone variants	4-6
1.4 Epigenetic modifications	6-8
1.4 Trithorax and Polycomb group of proteins	8-9
1.5 Trimethylation of histone H3 at Lys 36	9-11
1.6 The HMT Hypb and its role in transcription	11-13
RESEARCH RATIONALE	14
HYPOTHESIS AND OBJECTIVES	15

CHAPTER 2

MATERIALS AND METHODS	16-25
2.1 Antibodies	16

2.2 Cell culture	16
2.3Lentivirus production and lentiviral-mediated shRNA knockdown	16-17
2.4 RNA isolation and RT-qPCR assay	17-20
2.5 Western blot analysis	21
2.6 Histone extraction	21-22
2.7 Immunocytochemistry	22
2.8 Native chromatin immunoprecipitation	23
2.9 Cross-linked chromatin immunoprecipitation	23-24
2.10 Statistical analysis	24-25

CHAPTER 3

RESULTS	26-51
3.1 The H3K36me3 mark is enriched on the coding regions of muscle-specific genes during upon differentiation of C2C12 cells	26-31
3.2 Stable knockdown of histone methyltransferase Hypb in C2C12 cells	32-33
3.3 Hypb is responsible for majority of the global H3K36me3 methylation marks in differentiated C2C12 cells	34-35
3.4 Enhanced myogenesis phenotype upon lentiviral-mediated stable knockdown of Hypb in C2C12 cells	36-40
3.5 The H3K36me3 mark is lost on the coding regions of muscle-specific genes in the shHypb cell line	41-43
3.6 Enrichment of the HMT Hypb occurs in the coding regions of muscle-specific genes	44-48

<u>3.7 Enrichment of Hypb at the coding regions of Myog and Ckm genes is lost upon its knockdown</u>	<u>49-51</u>
---	---------------------

CHAPTER 4

<u>DISCUSSION</u>	<u>52-58</u>
--------------------------	---------------------

<u>FUTURE DIRECTIONS</u>	<u>59-62</u>
---------------------------------	---------------------

<u>REFERENCES</u>	<u>63-68</u>
--------------------------	---------------------

<u>APPENDIX</u>	<u>69-80</u>
------------------------	---------------------

LIST OF FIGURES

<u>1.1 Epigenetic landscape and cell commitment as depicted by Waddington</u>	<u>1</u>
<u>1.2 The process of muscle differentiation</u>	<u>2</u>
<u>1.3 Establishment of muscle program by histone methylation</u>	<u>3</u>
<u>1.4 Simple to Complex: From a DNA helix to a chromosome</u>	<u>5</u>
<u>1.5 Epigenetic modifications and their gene landscape</u>	<u>7</u>
<u>1.6 Inhibition of spurious transcription in yeast by H3K36me3</u>	<u>12</u>
<u>3.1A Chromatin resolution for N-ChIP experiments</u>	<u>29</u>
<u>3.1B (i) Amplicon sizes of Myog ChIP-qPCR primer sets</u>	<u>30</u>
<u>3.1B (ii) Amplicon sizes of Ckm ChIP-qPCR primer sets</u>	<u>30</u>
<u>3.1C H3K36me3 N-ChIP on differentiation time course in C2C12 cells</u>	<u>31</u>
<u>3.2A Lentiviral-mediated shRNA knockdown of histone methyltransferase Hypb in C2C12 cells</u>	<u>33</u>
<u>3.2B shRNA knockdown of Hypb at the protein level</u>	<u>33</u>
<u>3.3 Western blot of global histone extract</u>	<u>35</u>
<u>3.4A Lentiviral-mediated stable knockdown of Hypb comparing them to the control (TRC)</u>	<u>38</u>
<u>3.4B RT-qPCR showing mRNA expression of differentiation markers in stable TRC and shHypb lines</u>	<u>38</u>
<u>3.4C Western blot displaying levels of differentiation markers in 72-hour differentiated TRC and shHypb lines</u>	<u>39</u>

<u>3.4D Fusion index graph showing the average number of nuclei per myotube in TRC and shHypb knockdown lines</u>	39
<u>3.5 H3K36me3 enrichment on the Myog gene by N-ChIP</u>	42
<u>3.6A (i) Sonication optimization of chromatin for X-ChIP</u>	46
<u>3.6A (ii) Sonication optimization for X-ChIP: Ideal sample for X-ChIP experiments</u>	46
<u>3.6B Hypb enrichment on the Myog gene in wild-type differentiated (day 3) C2C12 cells by X-ChIP</u>	47
<u>3.7 Hypb enrichment on the Myog and Ckm genes in 72h differentiated, stable TRC and shHypb knockdown lines by X-ChIP</u>	50
<u>4.4 Proposed model for the enhanced myogenesis phenotype upon Hypb knockdown</u>	56

Appendix

<u>A1 Lentiviral-mediated stable knockdown of Ash1l comparing them to the control (TRC)</u>	69
<u>A2 Fusion index graph showing the average number of nuclei per myotube in TRC and shAsh1l knockdown lines</u>	70
<u>A3 Lentiviral-mediated shRNA knockdown of histone methyltransferase Ash1l in C2C12 cells</u>	71
<u>A4 H3K36me3 enrichment on the Myog and Ckm genes by N-ChIP in TRC and shAsh1l cell lines</u>	72

<u>A5 H3K36me3 enrichment on the Myog and Ckm genes by N-ChIP in the stable double knockdown line (shHyppb+shAsh1l)</u>	74
<u>A6 H3K36me3 enrichment on the Myh3 gene by N-ChIP</u>	76
<u>A7 Ash1l enrichment on the Ckm gene (Top) and Myog gene (Bottom) in wild-type differentiated (day 3) C2C12 cells by X-ChIP</u>	77
<u>A8 Ash1l enrichment on the Myog gene (Top) and Ckm gene (Bottom) in 72h differentiated, stable TRC and shAsh1l knockdown lines by X-ChIP</u>	79

LIST OF TABLES

<u>Table 1. Substrates recognized by different H3K36-HMTs</u>	10
--	-----------

LIST OF ABBREVIATIONS

5hmC – 5-Hydroxymethylcytosine
Ash2l – Absent, small, or homeotic-2 like
Ash1l1 – Absent, small, or homeotic-1 like
ChIP-Seq – Chromatin immunoprecipitation sequencing
Ckm – Muscle creatine kinase
COMPASS – Complex proteins associated with Set1
DAPI – 4',6'-diamidino-2-phenylindole dihydrochloride hydrate
DM – Differentiation medium
DMEM – Dulbecco's Modified Eagle Medium
DNA – Deoxyribonucleic acid
ESCs – Embryonic stem cells
FBS – Fetal bovine serum
GM – Growth medium
H3K27me – Trimethylation of histone H3 on lysine (K) 27
H3K36me3 – Trimethylation of histone H3 on lysine (K) 36
H3K4me3 – Trimethylation of histone H3 on lysine (K) 4
H3K9me – Monomethylation of histone H3 on lysine (K) 9
HATs – Histone acetyltransferases
HDACs – Histone deacetylases
HMT – Histone methyltransferase
HOX - Homeobox
HYPB – Huntington yeast partner B
KDM5A – Lysine (K) specific demethylase 5A
KDM5B = Lysine (K) specific demethylase 5B
MAPK – Mitogen-activated protein kinase
MEF2 – Myocyte enhancer factor 2
MHC – Myosin heavy chain
MLL – Mixed lineage leukemia
MNase – Micrococcal nuclease
MRF4 – Myogenic regulatory factor 4
Myf5 – Myogenic factor 5
Myh1 – Myosin heavy chain isoform 1
MyoD – Myogenic differentiation 1
Myog – Myogenic factor 4
N-ChIP – Native chromatin immunoprecipitation
Pax7 – Paired box protein 7
PBS – Phosphate buffered saline
PcG – Polycomb group of proteins

PHD – Plant Homeo Domain
PMSF- Phenylmethanesulfonylfluoride
RNA – Ribonucleic acid
RNA Pol II – RNA Polymerase II
RT-qPCR – Real time quantitative polymerase chain reaction
SDS-PAGE – Sodium dodecylsulphate – polyacrylamide gel electrophoresis
shRNA – short-hairpin ribonucleic acid
TET – Ten eleven translocation
TrxG – Trithorax group of proteins
X-ChIP – Cross-linked chromatin immunoprecipitation

ACKNOWLEDGMENT

Firstly, I would like to thank my thesis supervisor, Dr. F. Jeffrey Dilworth for providing me with the opportunity to work in his laboratory as a graduate student. I consider myself fortunate and honoured to have had him as a mentor. From reviewing progress reports and critiquing work-in-progress talks to helping me make important decisions regarding projects, he has offered continual guidance and support throughout the two years of my research experience. Without his professional and able guidance, this journey would have remained incomplete. I cannot thank you enough Jeff.

Next, I like to thank my thesis advisory committee members, Dr. Lynn Megeney and Dr. Duncan Stewart who have steered my thesis progress in the right direction by providing constructive comments during advisory committee meetings.

Next, I would like to thank all the Dilworth lab members, and by that I mean all the post-doctoral fellows—Arif, Herve, Kulwant, Soji, and Qi-Cai, who with their rich experience and enthusiasm made the research environment less intimidating and more knowledgeable. I also like to thank our humble and extremely cooperative lab manager, Lifang Li, who spent countless hours getting all the reagents in time for my experiments.

Last but not the least, I like to thank my parents, who have always motivated me towards my career goals and instilled in me the confidence to do the same. I also like to thank my sister who has always been at the giving end whenever I needed helpful advice. Finally, I like to thank my late grandfather who has been a role model and a source of motivation throughout my time as a graduate student.

INTRODUCTION

The genome and the epigenome

All cells in a multicellular organism possess the same genome, but what differs one cell type from the other is the epigenome. This disparity in cell structure and function is owed, in part, to epigenetics. The epigenome provides an additional rank of gene regulation to the underlying genome of an organism. This concept of epigenetic landscape was first proposed by Waddington **(1)(Fig. 1.1)**, whereby he mentioned that the processes of cell commitment and differentiation is an epigenetic phenomenon rather than a genetic one. He also went on to define epigenetics as a study of stable, heritable changes in gene expression without any change in the genetic makeup **(2)**. These epigenetic modifications are classified into many different types namely, DNA methylation, histone methylation, acetylation, ubiquitinylation, phosphorylation, and sumoylation **(3)**.

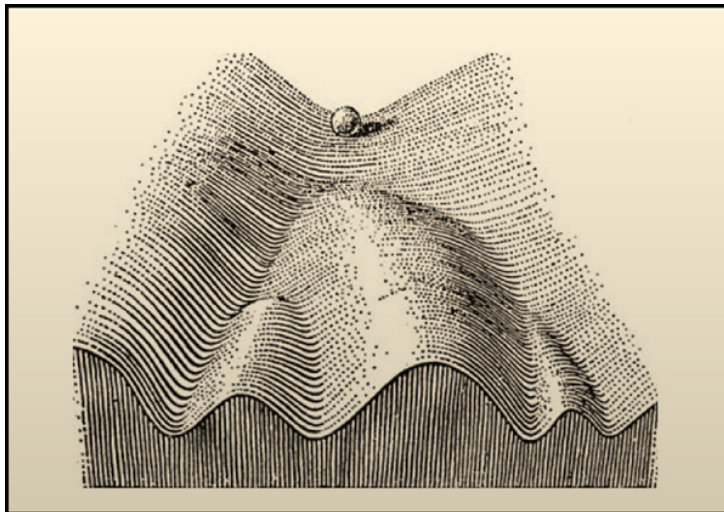


Fig. 1.1. Epigenetic landscape and cell commitment as depicted by Waddington

The ball in the figure represents a cell before an epigenetic landscape whereby it encounters different paths that result in different fates (cellular lineages) **(2)** (© Cell publishing group, used with permission, all rights reserved. See Ref. **2** for details).

Muscle differentiation

Muscle differentiation is an intricate process that begins with a muscle satellite stem cell undergoing activation upon injury or activity that prompts cell cycle exit resulting in myoblasts. These cycling myoblasts undergo further differentiation and fuse to form multinucleate myotubes. These events of commitment, differentiation, and fusion are under the control of muscle-specific transcription factors. Myf5 and MyoD act redundantly in muscle cell fate specification, whereas, myogenin and MRF4 in conjunction with MEF2 regulate muscle differentiation (4) (Fig.1.2).

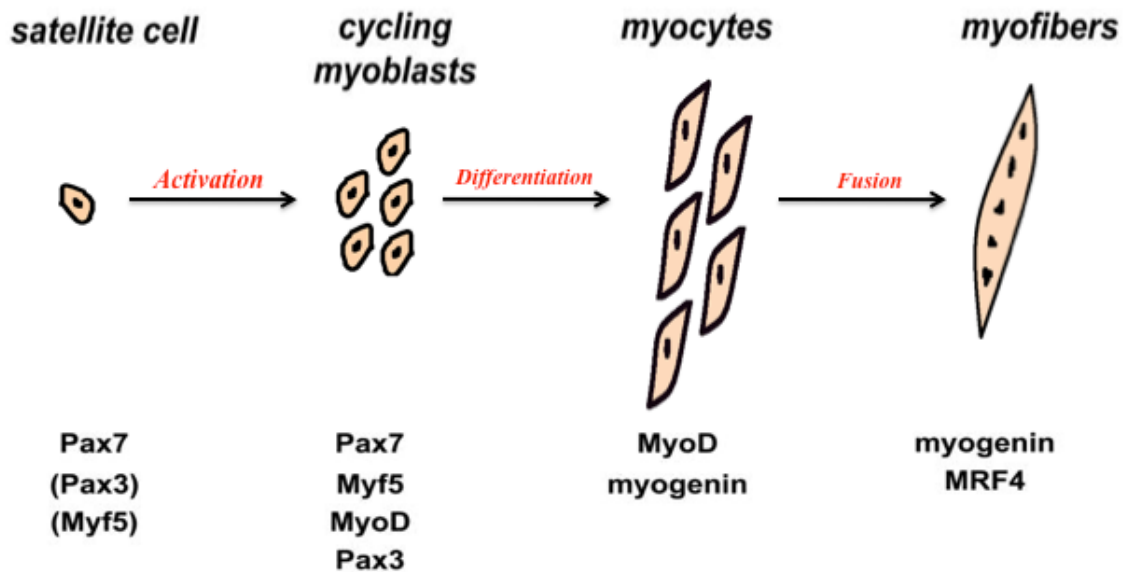


Fig. 1.2 The process of muscle differentiation

Upon injury satellite stem cells are activated and form myoblasts which exit upon cell cycle withdrawal, differentiate to form myocytes. These myocytes then fuse to form long terminally-differentiated myofibers. Pax7 identifies stem cells, while Myog, MyoD, and MRF4 play a major role in maintaining the muscle program upon differentiation (Adapted from LeGrand et al., 2007) (5).

Bennett et al (1997) **(6)** and Bengal et al (1999) **(7)** have shown that p38 MAPK signaling plays an important role in establishing muscle-specific gene expression. In 2007, our group showed that a trithorax group protein Ash2L, a component of a histone methyltransferase (HMT) complex, helps p38 in establishing muscle-specific gene expression **(8)**. Prior to the activation of p38, the muscle-specific promoter is acetylated on histone H4. Upon activation, p38 MAPK phosphorylates Mef2d, which then recruits Ash2L-containing HMT complexes to muscle-specific promoters. Once at the promoters, these chromatin-remodeling enzymes act to establish H3K4me3 epigenetic mark that leads to the activation of muscle-specific gene expression **(8)** **(Fig.1.3)**.

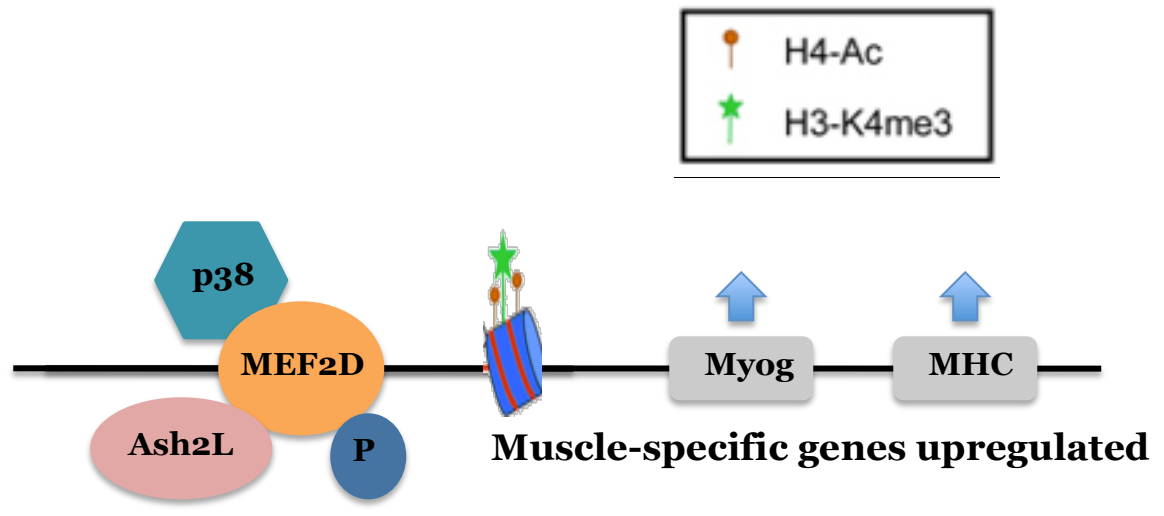


Fig.1.3 Establishment of muscle program by histone methylation

Cell-cell contact induces cells to exit cell cycle and undergo differentiation where activated p38 recruits Ash2L-containing HMT complex to muscle gene promoters. These complexes trimethylate lysine 4 on histone 3, hence, activating the muscle program **(8)**.

Chromatin organization and variant histones

The most basic structural and functional unit of chromatin is called a nucleosome – a 146 bp fragment of negatively-charged DNA wrapped around an octamer of positively-charged four histones (H2A, H2B, H3, H4) **(9)**. Due to the opposite charges, histones bind DNA very tightly. The core histone particle consists of tetramers of H3 and H4 and dimers of H2A and H2B **(10)**. Each histone consists of a core globular domain and an N-terminal tail protruding from the former **(10)**. The nucleosomes interact with each other forming compact secondary structures spanning 30nm as shown in **Fig. 1.4 (9)**. These secondary structures come together to form complex tertiary structures called chromosomes **(Fig. 1.4)**. This compaction of DNA in complex chromatin structures presents a problem in that it renders genes inaccessible to proteins involved in various processes such as transcription, DNA repair, and replication. This problem, however, is circumvented by the presence of histone-modifying enzymes (discussed in brief later) and chromatin remodelers, that mark the histones and recruit various complexes, respectively, resulting in a highly dynamic nature of chromatin, thus, making it a key player in epigenetic regulation of gene expression **(11-13)**.

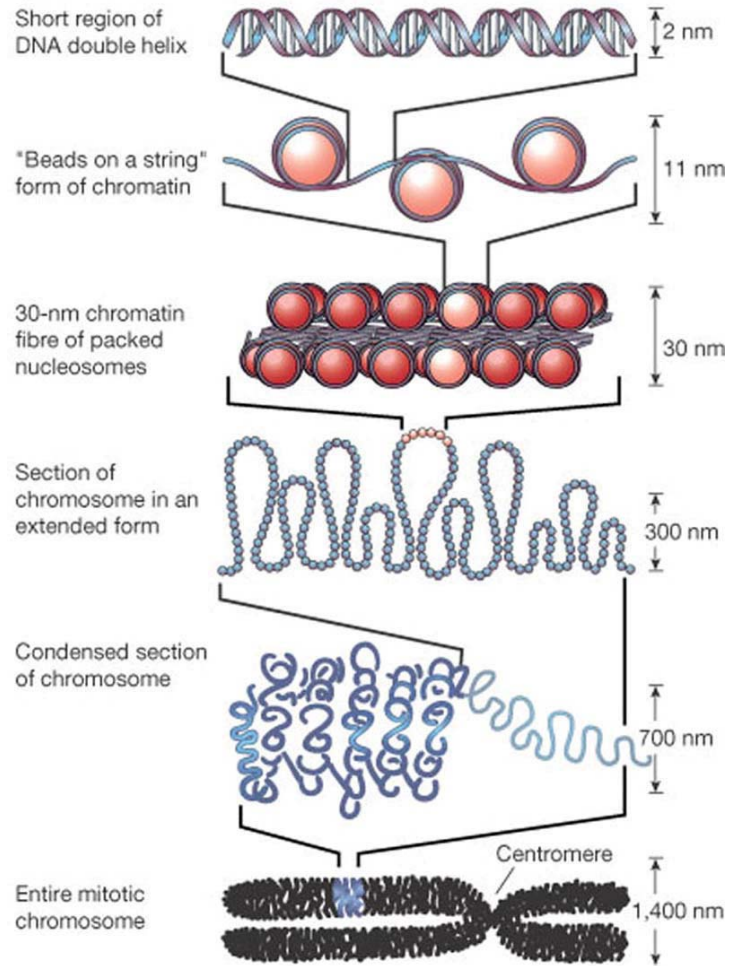


Fig. 1.4. Simple to Complex: From a DNA helix to a chromosome

The simplest organization in chromatin is the nucleosome which is 146 bps of helical DNA wrapped around histone proteins. These nucleosomes are interconnected through linker DNA and compact to form structures that span 30nm. These then further condense to form higher-order structures eventually culminating in a chromosome that is a dense, compact mass of nucleosomes (9) (© Nature publishing group, used with permission, all rights reserved. See Ref. 9 for details).

In addition to histone modifications and chromatin remodeling, cells create variation in the chromatin by way of introduction of variant histones (14). In mammals, three histone H3 variants exist—H3.1, H3.2, and H3.3 (15,16). In cells of limited potency, histone variant H3.3 is enriched on actively transcribed genes (17-

19) The histone chaperone Hira associates with H3.3 and mediates replication-independent nucleosome assembly **(20,21)**, whereas, another histone chaperone CAF-1 associates with H3.1 replication-coupled assembly of nucleosomes **(21,22)**. Variant H3.2 is found to be coupled to replication **(23)**.

Epigenetic modifications

The globular domain and the histone tail are subject to various post-translational modifications such as methylation of lysine residues, methylation of arginine residues, acetylation, ubiquitination, sumoylation of lysines, and phosphorylation of serines and threonines **(24)**. Acetylation of H3 and H4 and di- or trimethylation of H3K4 are markers of transcribed genes and hence are called euchromatin modifications, whereas, H3K27me and H3K9me are markers of inactive or repressed genes and are referred to as heterochromatin modifications **(25)**. Most histone modifications have characteristic patterns of localization on the gene—either at the 5'end, 3'end, or anywhere upstream of the promoter region. These distinct patterns are considered crucial for transcription. For example, H3K4me1 is present on the enhancer elements and 3' end of the genes, while the promoter region is marked by H3K4me3 **(26)**. All methylation marks are catalyzed by HMTs. The methylation marks established by HMTs are erased by histone demethylase enzymes **(27)**. The acetylation marks on the other hand are carried out by histone acetyltransferases (HATs) **(28)**, whose effect is antagonized by histone deacetylases (HDACs). The function of a HAT is dependent on the number of lysines modified. All acetylation and phosphorylation marks are known to be characteristic

of actively transcribed regions of the gene (29). This project focuses on the histone methylation mark H3K36me3.

Histone methylation is a type of histone modification catalyzed by HMTs that transfer one to three methyl groups from S-adenosyl methionine to lysine and arginine residues of histones. Some of the well-known modifications are shown in Fig. 1.5 along with their distribution on the gene and resultant gene transcription activity (active/repressive) (30).

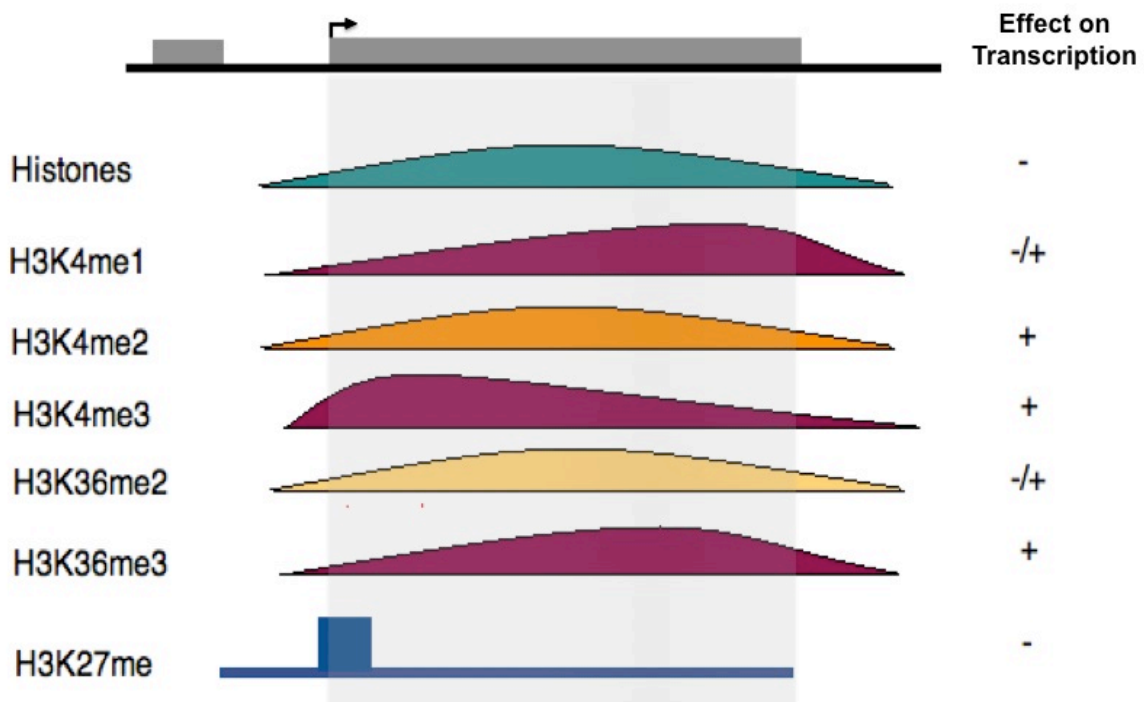


Fig. 1.5 Epigenetic modifications and their gene landscape

Distribution of histone modifications mapped onto a hypothetical gene. The arrow on the gene marks the promoter region and the direction points towards the 3' end of the gene. The data is representative of yeast genome (Adapted from Li et al., 2007) (30).

These epigenetic modifications control gene expression through recruitment of various transcription factors.

Relatively recently, a new epigenetic modification called hydroxymethylation has been discovered. It is previously known in literature that 5-Methylcytosine (5mC) is a DNA mark associated with gene silencing and stability that is present on dinucleotides of CpG islands **(31)**. Recently, it has been shown that in higher mammals, 5mC can be oxidized to 5-hydroxymethylcytosine (5hmC) by ten-eleven-translocation (TET) enzymes **(32,33)**. It has also been shown that in differentiating mouse neural and adipocyte cells, the mark 5hmC associates with transcription factor binding at distal regulatory regions (enhancers) of respective genes **(34)**. Therefore, it could have a putative role in establishing tissue-specific gene expression through enhancer elements.

Trithorax and Polycomb group of proteins

The regulation of Homeobox genes (HOX genes) during early embryogenesis in *Drosophila melanogaster* is under the control of transcription factors that disappear at later stages in development. The HOX gene memory is maintained by two antagonistic groups of proteins—Trithorax group (TrxG) proteins and Polycomb group (PcG) proteins that activate and repress transcription, respectively **(35)**. Trimethylation of histone 3 at lysine 4 (H3K4me3) is a hallmark of actively

transcribing gene promoters, whereas trimethylation of histone 3 at lysine 27 (H3K27me3) is a hallmark of repressive genes. TRX was the first member of the TrxG family to be characterized and it is known to be a positive regulator of HOX genes **(36,37)**. Although its function was not required to initiate gene expression, it was needed to maintain HOX gene expression in the body segments later on in development **(38)**. Both TrxG and PcG proteins regulate many genes in addition to HOX genes **(39)**.

TrxG proteins are evolutionarily conserved regulators of gene expression that are categorized into three main classes—ATP-dependent chromatin remodeling complexes, sequence-specific DNA binding proteins, and histone-modifying complexes. Set1 was the first HMT shown to catalyze mono-, di-, and trimethylation of H3K4 in the COMPASS (complex proteins associated with Set1) protein complex in yeast **(40)**. Subsequent studies have found at least six mammalian homologs of COMPASS exist namely, MLL1-4, hSET1a, and hSET1b **(41,42)**. It has also been shown that the recruitment of these HMT complexes results in H3K4me3 mark on active gene promoter regions in various cell types analyzed **(40-42)**.

Trimethylation of histone H3 at lysine 36

H3K36 methylation is a form of post-translational modification that is conserved from unicellular organisms such as yeast to complex, multicellular organisms such as humans **(43)**. While trimethylation at H3K4 is present around

the transcription start site towards the 5' end of the gene, trimethylation at H3K36 is predominantly found in the coding region towards the 3' end of actively transcribed genes (44). The former is a marker of transcriptional initiation, whereas, the latter is characteristic of transcriptional elongation (44). In addition to this function, this mark has been implicated in alternative splicing, DNA methylation, and evidently in pathologies, all of which are discussed briefly later in this section.

According to current literature, there are six known H3K36-specific lysine HMTs as listed in **Table 1** with the specific histone methylations they catalyze. All these HMTs have a catalytic SET domain in common, but differ in methylation preferences (me1, me2, or me3). Some of these enzymes have chromatin interacting domains that recognize the K36 mark and other methylated histone residues, for example, the PHD finger domains (45).

Table 1. Substrates recognized by different H3K36-HMTs

HMT	Histone targets	Reference
NSD1	H3K36me1, H3K36me2, H4K20	46-49
NSD2	H3K36me2	50
NSD3	H3K36me2	52-54
SETD2/ HYPB	H3K36me3	55-57
ASH1L	H3K4me3, H3K36me2, H3K36me3	22-24

H3K36me3 has been shown to modulate DNA methylation by interacting with the PWWP domain of the DNA methyltransferase Dnmt3a and increasing its methylation activity **(58)**.

Luco et al. (2010) demonstrated a role of H3K36me3 in alternative splicing **(59)**. They proposed a presence of an adaptor system that constitutes a histone modification, a chromatin-binding protein that is responsible for reading the histone mark, and a splicing regulator. This complex serves as epigenetic information that is transmitted to the pre-mRNA processing machinery where specific splicing regulators are recruited, eventually helping decide the splicing outcome **(59)**.

The HMT Hypb and its role in transcription

Set2, the yeast ortholog of human Hypb was the first H3K36-specific HMT to be discovered **(60)**. In addition to *Saccharomyces cerevisiae*, it is the only known HMT to be present in *Neurospora crassa* (bread mould) and *Schizosaccharomyces pombe* (fission yeast) **(61,62)**. Two reports have shown that in *in vitro* experiments, Hypb is able to act as a mono-, di-, and tri-methyltransferase **(63,64)**, however, a report *in vivo* has shown it to be a sole H3K36me3 HMT **(64)**. This raises a question whether it is mandatory to have a dimethyl mark before it is trimethylated or all three methyl marks are placed by an H3K36me3 HMT at once on the lysine substrates. The answer to this is not yet clear.

In yeast, the nucleosomes are methylated on H3K36 by Set2, a HMT that travels with RNA Pol II by binding to the C-terminal domain of the latter during transcriptional elongation (65). These methylation marks provide for a transcriptional memory mark that signals recruitment of HDACs such as Rpd3S (in yeast) to actively transcribed regions of the gene. This deacetylates the histones in these regions that erases transcription-elongation associated acetylation and prevents spurious initiation of transcription at these sites (65).

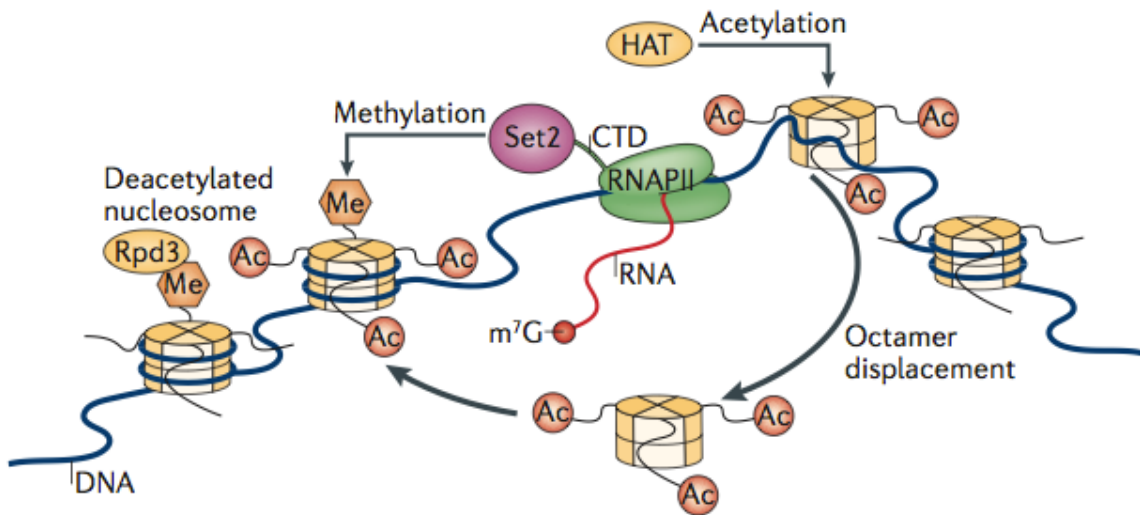


Fig. 1.6 Inhibition of spurious transcription in yeast by H3K36me3

Set2 binds active elongating Pol II and marks the nucleosomes with H3K36me3. This signals binding of the Rpd3S complex facilitating deacetylation of the nucleosomes, hence, preventing spurious activation of transcription from sites other than the promoter region during elongation (45) (© Nature publishing group, used with permission, all rights reserved. See Ref. 45 for details).

In a recent paper, it was shown that one of the HMTs SETD3 plays a role in muscle differentiation **(66)**. Although recently, it has been shown that the HMT Hypb plays a role in embryonic vasculature development **(67)**, there has not been any direct evidence suggesting the role of H3K36me3 mark in muscle development. As the H3K36me3 mark is catalyzed by the HMT Hypb, we want to determine whether trimethylation of H3K36 plays a role in myogenic differentiation.

RESEARCH RATIONALE

Establishment of the muscle differentiation program is a complex event that starts with a satellite stem cell committing itself to form myoblasts that differentiate into multinucleate myofibers. Epigenetic mechanisms play a significant role in regulating gene expression programs not just in muscle, but in almost every cell lineage of the body. Histone methylations are key modifications that can mark genes for repression by the H3K27me3 mark or for activation by H3K4me3 or H3K36me3. This study looks to delineate the role of the H3K36me3 in muscle differentiation. In order to connect the two, we need a middleman that in this project is Hypb, a histone methyltransferase. Our approach is to generate cell lines that show stable knockdown of Hypb and then observe the effects on H3K36me3 and muscle differentiation by targeting muscle-specific genes that have previously been shown to be epigenetically-regulated. This would help identify the role of Hypb and HMTs, in general, in the process of muscle differentiation and enhance knowledge in the public domain regarding muscle development.

Hypothesis

As the coding regions of muscle-specific genes are marked by H3K36me3, we propose that this mark plays an important role in modulating transcriptional events necessary for muscle differentiation.

Objectives

1. Determine the regions of H3K36me3 enrichment on the muscle-specific genes by native chromatin immunoprecipitation (N-ChIP) experiments
2. Identify H3K36-specific methyltransferases involved in myogenesis based on their knockdown phenotype in differentiation conditions and identify their targets based on previous literature and cross-linked chromatin immunoprecipitation (X-ChIP) experiments
3. Determine the role of the H3K36me3 mark in myogenic differentiation.

CHAPTER 2

MATERIALS AND METHODS

2.1 Antibodies

Commercial antibodies used in the experiments include H3K36me3 (Abcam ab31356), H3 antibody (Millipore 06-755), Hypb (Abcam ab31358), Ash1l (Abcam ab4477), Myog (Santa Cruz SC-576), and MHC (MF-20).

2.2 Cell culture

The mouse myoblast cell line C2C12 was maintained at less than 70% confluency in DMEM (10% FBS, 1:100 Pen/Strep) and differentiated in DMEM (2% horse serum, 1:100 Pen/Strep, no insulin or transferrin).

2.3 Lentivirus production and lentiviral-mediated shRNA knockdown

Low-passage 293T cells were plated and propagated in DMEM(10% FBS, 1:100 Pen/Strep) prior to transfection. 3.6ug of pMD2.G plasmid (Addgene #12259), 5.4ug of psPax2 plasmid (Addgene #12260), and 3ug of a lentiviral vector pLKO.1 were added to a tube. In a separate tube, DNA-Optimem and Lipo-Optimem solutions were combined and mixed gently. After 20 minutes, DNA-Lipofectamine 2000 complex was added to the 10cm tissue culture plate with GM (w/out

antibiotics). Trypsinized 293T cells were added and the plate gently rocked. Next day, GM was replaced with new GM (w/antibiotics). Supernatant was collected 48 hours later and centrifuged at 1600 g for 5 minutes at 4°C. 1mL virus aliquots were stored at -80°C for infections. These lentivirus expressing shRNA targeting Hypb or Ash1l were used to infect C2C12 cells that were plated at 40% confluency on the day of infection. 1mL of viral supernatant was added with 9mL of GM, followed by addition of polybrene (4ug/mL). 24-36 hours were allowed for lentiviral genes to be expressed before selecting cells for puromycin (72-96h). The level of stable shRNA knockdown was later analyzed using RT-qPCR.

2.4 RNA isolation and RT-qPCR assays

Total RNA was extracted using the RNA STAT-60 protocol which a single-step protocol **(68,69)**. RNA was homogenized in RNA STAT-60 solution. Chloroform was then added to separate the solution into two phases—the aqueous phase and organic phase. Total RNA content is found in the former phase, whereas DNA and proteins are found in the latter. Isopropanol is added to the aqueous phase in a new tube to precipitate the DNA. The DNA pellet was then washed with 75% ethanol, and solubilized in water. 500 ng of total RNA was reverse transcribed into cDNA using random primers and MMLV reverse transcriptase enzyme **(69)**.

Gene expression for all experiments was quantitated by RT-qPCR using SYBR-Green mix and primer sets. 35 cycles of qPCR were run in order to quantify mRNA

transcript level data. GAPDH (constitutively-expressed gene) was used in all experiments as an internal control and all the data was normalized to it.

The following are the sequences of the primers (in 5'-3' orientation) used for gene expression analysis:-

Tnnt2

Forward primer: AGG TTG GTC CTG ATG AAG AAG CCA

Reverse primer: ATT CAG GTC CTT CTC CAC ACG CTT

Ckm

Forward primer: GAG CCT CTG CTC CTC GCT AAG AGT

Reverse primer: AGG CAG AGT GTA ACC CTT GAT GCT

a-Actin

Forward Primer: TTG TGC GCG ACA TCA AAG AGA AGC

Reverse Primer: GAA ACG CTC ATT GCC GAT GGT GAT

Myog

Forward Primer: CAT CCA GTA CAT TGA GCG CCT ACA

Reverse Primer: AGC AAA TGA TCT CCT GGG TTG GGA

MHC

Forward Primer: ACC TTG CCA AGA AGA AGG ACT CCA

Reverse Primer: TGG ATG CGG ATG AAC TTG CCA AAG

DDX5

Forward Primer: TTC TGA TTG CTA CCG ATG TGG CCT

Reverse Primer: TGG TAC TGC GAG CAG TTC TTC CAA

The following are the sequences of the primers (in 5'-3' orientation) used for ChIP-qPCR analysis:-

Myog Prom

Forward Primer: TCA CAT GTA ATC CAC TGG AAA CG

Reverse Primer: CCT GAG CCC CCC TCT AAG C

Myog +1kb

Forward Primer: TCC CAA CCC AGG AGG TAA GTG AAT

Reverse Primer: AAG GGT GAT TGG GAC AGG GCT TAT

Myog +2kb

Forward Primer: AGA CGA AAC CAT GCC CAA CTG AGA

Reverse Primer: AGG AGT CAG CTA AAT TCC CTC GCT

Ckm Prom

Forward Primer: TAG TCA CAC CCT GTA GGC TCC TCT AT

Reverse Primer: ATT TCT CTC AGT CCC TAC CTG GCT

Ckm +4kb

Forward Primer: AAC CCA AGC TTC CTA TGC TGA CCT

Reverse Primer: TCG GTC TTA TGC TTG TCT GTG GGT

Ckm +5kb

Forward Primer: TTT GAT GTC CGG GTG AGG AGT GTT

Reverse Primer: ATC AAT GAG CTG CTG TTC CTG

IgH Enh

Forward Primer: GCC TGG GCT TGC TTT GTC T

Reverse Primer: TGA CCC CTT TCG CTC ATT G

Bmaj Prom

Forward Primer: CTG CTC ACA CAG GAT AGA GAG GG

Reverse Primer: GCA AAT GTG AGG AGC AAC TGA TC

2.5 Western Blot analysis

Three-day differentiated C2C12 cells were washed with 1X PBS and harvested. Cells were transferred to an eppendorf tube and centrifuged at max. speed (11,500g) for 17 seconds at 4°C. Cell pellet was diluted in lysis buffer and vortexed briefly every 10 minutes for 30 minutes. Cell lysate was centrifuged at 11,500g for 10 minutes and the supernatant transferred to a new eppendorf tube and stored at -80°C until future use. These whole cell protein extracts were thawed and Bradford assay performed on each sample. 10ug of protein sample was loaded in each well. 5 uL of molecular weight marker was also loaded in one well. Gels were run at a constant voltage of 75V for 1.5-2 hours or until the dye runs out. Samples were transferred onto a membrane after running transfer for 2 hours (75V) with ice-cold blocks. Blots were then blocked for 1 hour in 3% milk in 1X PBS and incubated in primary antibody overnight at 4°C. Next day, blots were washed for 30 minutes at room temperature with PBS. Following this, the blots were incubated in secondary antibody (1:2500) for 1 hour at room temperature. Washes were performed again for 30 minutes. The blots were finally developed and analyzed for protein level expression.

2.6 Histone extraction

Histones were extracted using the protocol by Abcam **(70)**. Low-passage C2C12 were propagated and differentiated for 72 hours. They were then harvested

and washed twice with ice-cold PBS. Cells were then resuspended in TEB(0.5% Triton X 100 v/v) and 2mM PMSF and lysed on ice for 10 minutes with light stirring. They were centrifuged at 1200g for 10 minutes at 4°C. The pellet was then resuspended in 0.2N HCl. Histones were acid extracted overnight at 4°C. Next day, the samples were centrifuged at 1200g for 10 minutes at 4°C. Supernatant was removed and protein content determined using Bradford assay. This was followed by western blot analysis of the histone proteins on a 15% SDS-PAGE gel.

2.7 Immunocytochemistry

Low-passage C2C12 cells were differentiated for 72 hours before being harvested for immunofluorescence. Cells were washed with 1X PBS 2 times and fixed with 3.7% paraformaldehyde for 10 minutes at room temperature. Cells were washed three times with PBS and then permeabilized using 0.3% Triton X-100 at room temperature for 10 minutes. Cells were again washed with PBS two times before blocking with 10% FBS in PBS for 1 hour at room temperature. Cells were then incubated in primary antibody for 2 hours (subsequent washing with PBS three times) followed by 45 minutes in the secondary antibody (1:1000). Cells were then washed with PBS three times and counterstained with DAPI nuclear stain for 5 minutes. Cells were then taken for imaging at the AxioCam inverted microscopy facility.

2.8 Native chromatin immunoprecipitation (N-ChIP)

Approximately 1×10^7 C2C12 cells obtained at three days of differentiation were lysed in buffer A (15 mM Tris (pH 7.5), 15 mM NaCl, 60 mM KCl, 250 mM sucrose, 5 mM MgCl₂, 1.5 mM CaCl₂, 1 mM DTT and 1 mM PMSF). Nuclear pellet was obtained by centrifugation and quantified by measuring total ribonucleic acid content. Chromatin (2 mg) was digested with 1 U MNase (Sigma) for 10 min at 37°C to give a maximum visible fragment size of 550-600 base pairs (mono-, di-, and tri-nucleosomes). NaCl was then added to a final concentration of 600 mM, and the nuclei were incubated with 10 mg of hydroxyapatite resin to form slurry. After extensive washing, nucleosomes were eluted from the resin with a buffer containing 300 mM NaPO₄ (pH 7.2). Eluted nucleosomes were then subjected to immunoprecipitation with either an anti-H3K36me₃ or rabbit IgG. Immunoprecipitated DNA was then purified and subjected to qPCR analysis. To calculate relative enrichment, we subtracted the signal observed in the control immunoprecipitation experiment from that observed with the specific antibody, then divided the resulting difference by the ChIP input material.

2.9 Cross-linked chromatin immunoprecipitation (X-ChIP)

Enrichment of transcriptional regulators (Hypb, Ash1l) with specific regions of the genome was performed using X-ChIP. Three-day differentiated C2C12 cells were harvested and cross-linked for 20 minutes at room temperature with 1%

formaldehyde. The effect of formaldehyde was neutralized by adding 2.5M Glycine for 10 minutes at room temperature. Cells were washed with cold PBS and swelling buffer added (Buffer N – 25mM HEPES pH 8.0, 10mM KCl, 1.5mM MgCl₂, 0.1% NP-40, 1mM DTT, 2mM PMSF, and 1x protease inhibitor cocktail). Cells were then lysed in lysis buffer (1% NP-40) and resuspended in Buffer N after which they were MNase-digested for 10mins at 37°C. Cells were then sheared in sonication buffer using a Bioruptor (Diagenode) to obtain a resolution of 150-400bp. Supernatant was then immunoprecipitated with 5ug of the respective antibodies (Hypb or Ash1l and Rb IgG) overnight at 4°C. Next day, beads were subjected to low salt (20mM Tris-HCl pH 8.0, 150mM NaCl, 2mM EDTA, 1% Triton X-100), high salt (20mM Tris-hCl pH 8.0, 500mM NaCl, 2mM EDTA, 1% Triton-X 100), and LiCl buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 250mM LiCl, 1% NP-40, 1% Na-deoxycholate). Finally, the beads were washed twice in TE buffer for 10 mins. Following this, beads were eluted for 10 mins at 65°C and crosslinking reversed overnight. Next day, Immunoprecipitated DNA was then ethanol-purified and subjected to qPCR analysis. To determine relative enrichment **(71)**, qPCR amplification of immunoprecipitated samples was normalized to genomic DNA for each individual primer set. Normalized values obtained from a mock IP (Rb IgG) and then corrected for sample variation through division by ChIP input.

2.10 Statistical analysis

All values are represented as means \pm SEM of at least three biological replicates. Statistical significance of the difference in gene expression between

control and shHypb cells was estimated by a paired student t-test.

CHAPTER 3

RESULTS

3.1 The H3K36me3 mark is enriched on the coding regions of muscle-specific genes upon differentiation of C2C12 cells

The mouse myoblast C2C12 cell line was used to conduct all experiments. This line was generated as a result of serial passaging of myoblasts isolated from the thigh muscle of post-injury C3H mice **(72)**. These myoblasts have the potential to differentiate in culture and form myotubes.

It has been previously reported in the literature that the H3K36me3 mark is predominantly found in the coding region of the genes towards the distal 3' end in various cell lineages **(44)**. However, the same has not yet been shown in muscle. Therefore, we used the high-resolution technique of N-ChIP in order to further our understanding of this mark in muscle. This technique is used to determine epigenetic modifications on histones.

As with any protocol, we needed to optimize the N-ChIP procedure for the H3K36me3 antibody. Wild-type C2C12 cells were used for optimizing the protocol. The chromatin was fragmented by MNase digestion **(Refer to Materials and Methods section)**, after which the DNA was purified and run on an agarose gel. As

shown in **Fig. 3.1A**, the chromatin resolution (mono- and di-nucleosomes) was optimal to proceed to subsequent N-ChIP experiments.

For the purpose of ChIP experiments, we chose to study two genes – Myog and Ckm. We have shown previously that these two muscle-specific genes become transcriptionally active upon H3K4me3 marking by an HMT complex **(8)**. Three different regions spanning each gene were selected and primer sets were designed against these regions. For the Myog gene, **Myog promoter** (TSS), **Myog +1kb** (1000bp downstream of the Myog TSS), and **Myog +2kb** (2000bp downstream of the Myog TSS) were used and for the Ckm gene, **Ckm promoter** (TSS), **Ckm +4kb** (4000bp downstream of the Ckm TSS), and **Ckm +5kb**(5000bp downstream of the Ckm TSS) were used for the purposes of ChIP-qPCR.

In ChIP-qPCR, the amplicon size (size of the fragment that is amplified) is important to consider. The smaller the amplicon size, the greater the chance to observe amplification from short DNA fragments - in order for both forward and reverse primers to bind, the amplicon should be shorter than the average DNA fragment size. This was also taken into consideration during the optimization of N-ChIP experiments. The amplicon sizes of the primer sets used is shown in **Fig. 3.1B (i) and (ii)**.

In order to determine whether H3K36me3 enrichment levels changed at muscle specific-genes upon differentiation, we performed N-ChIP in wild-type

C2C12 cells. One set of cells were harvested under growth conditions while the other set was allowed to undergo three days of differentiation. DNA was purified and subjected to RT-qPCR experiments. Under growth conditions, no significant enrichment was observed at any loci on Myog and Ckm genes (**Fig.3.1C**). However, upon differentiation, high levels of H3K36me3 enrichment were observed at Myog +2kb and Ckm +4kb loci on the Myog and Ckm genes, respectively (**Fig.3.1C**) and almost no enrichment on Myog and Ckm promoter regions. Rb IgG was used as an antibody control. To control for non-specific binding, we examined H3K36me3 enrichment at the IgH enhancer as this gene is not regulated in proliferating or differentiating muscle cells.

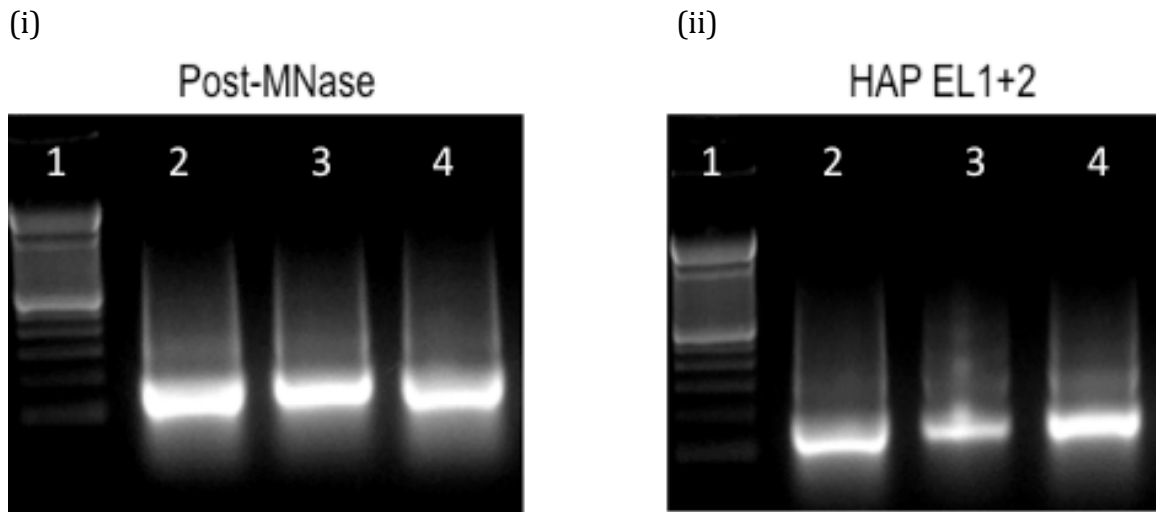
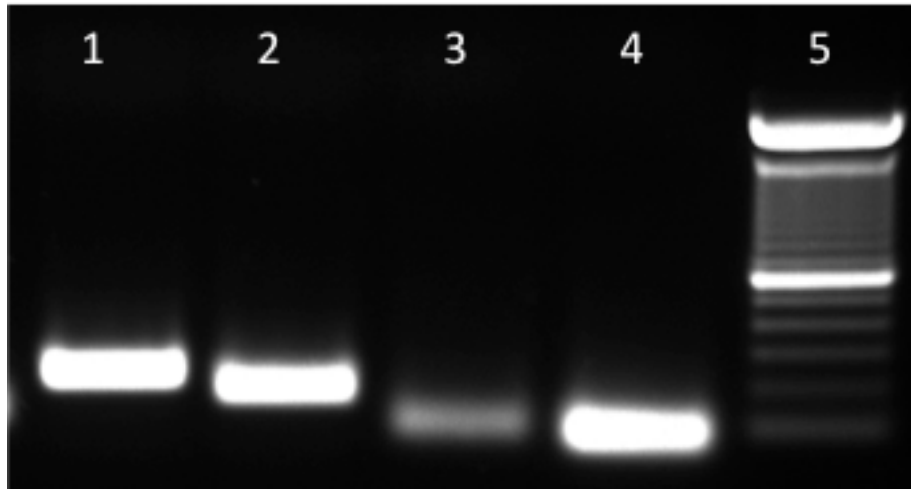


Figure 3.1A. Chromatin resolution for N-ChIP experiments Lane 1 = 100bp DNA ladder; Lane 2 = WT C2C12 cells; Lane 3 = TRC cell line; Lane 4 = shHypb cell line. Immediately post-MNase digestion **(i)**, after elutions 1 and 2 - used for Ab-beads binding **(ii)**.

(i)



(ii)

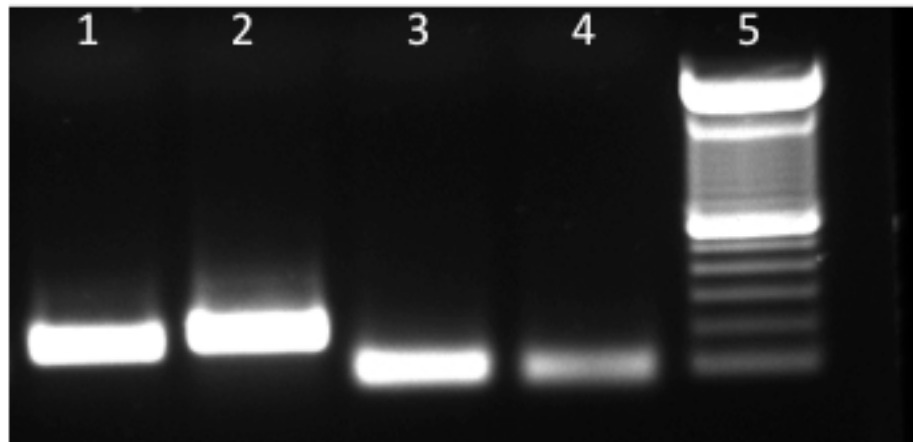


Fig. 3.1B (i) – Amplicon sizes of Myog ChIP-qPCR primer sets. From left to right: Lane 1= Myog +2kb; Lane 2= Myog +1kb; Lane 3= Bmaj Prom; Lane 4 = Myog Prom; Lane 5= 100bp DNA ladder.

Fig. 3.1B (ii) – Amplicon sizes of Ckm ChIP-qPCR primer sets. From left to right: Lane 1= Ckm +5kb; Lane 2= Ckm +4kb; Lane 3= Ckm Prom; Lane 4 = Bmaj Prom; Lane 5= 100bp DNA ladder.

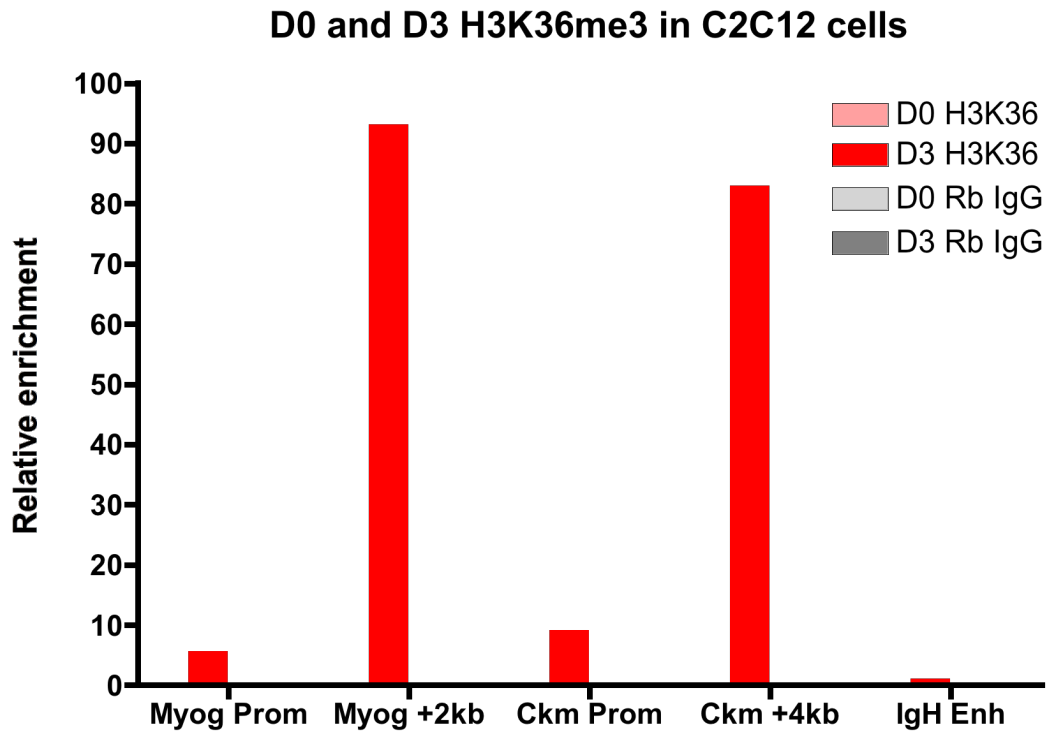
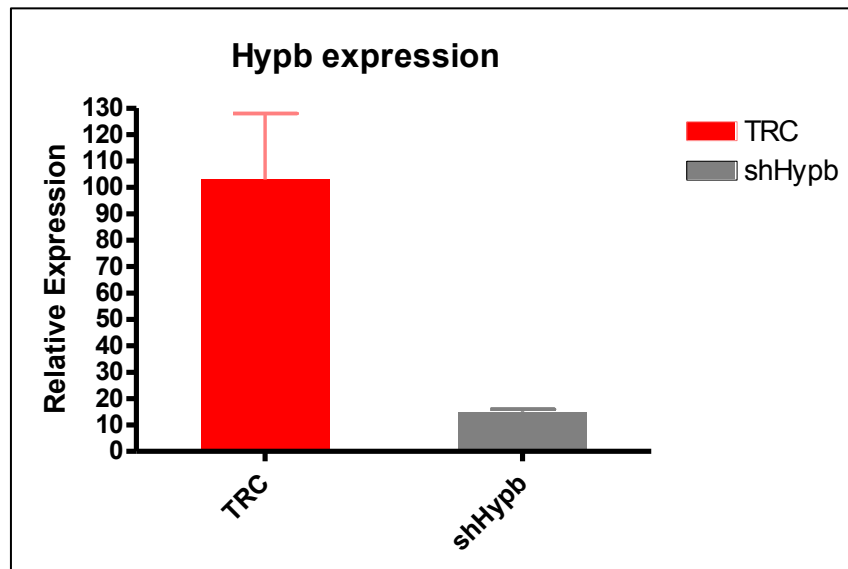


Figure 3.1C. N-ChIP on differentiation time course (day 0 and 3) on 10×10^6 C2C12 cells showing H3K36me3 enrichment levels on Myog and Ckm gene coding regions (Myog +2kb and Ckm +4kb, respectively) followed by RT-qPCR. IgH enhancer region is the negative binding region and hence acts as a negative control. Rb IgG acts as a negative control for non-specific binding to the antibody.

3.2 Stable knockdown of histone methyltransferase Hypb in C2C12 cells

To understand the function of the H3K36-specific histone methyltransferase Hypb in differentiation of C2C12 cells, we knocked it down using a lentiviral shRNA-based approach and generated a stable knockdown line. Before assessing the functional effects on C2C12 cells, the knockdown was confirmed at the transcript level (80-85% knockdown) as well as protein level by RT-qPCR and Western blot, respectively. TFIIH was used as the loading control (**Fig. 3.2A and B**).

A



B

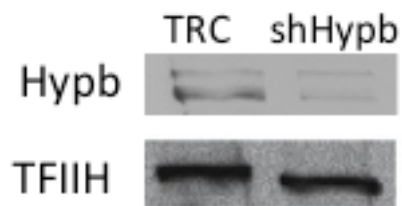


Figure 3.2A. Lentiviral-mediated shRNA knockdown of histonemethyltransferase Hypb in C2C12 cells. Expression of the respective genes in the stable knockdown lines was compared to the TRC (control) line and plotted in a bar graph as relative expression. RNA from 100,000 cells/condition was isolated, reverse transcription performed. The cDNA was then used as a template for RT-qPCR. Samples were run in triplicates. Error bars represent standard error.

Figure 3.2B. shRNA knockdown of Hypb at the protein level in TRC and shHypb cell lines. TFIIH was used as a loading control.

3.3 Hypb is responsible for majority of the global H3K36me3 methylation marks in differentiated C2C12 cells

In order to look at the global histone methylation levels, histones were acid extracted overnight (at 4°C) from three day-differentiated (day 3) lines – the TRC and shHypb cell lines. The histone proteins were immunoprecipitated with the H3K36me3 antibody. H3 acted as a loading control (**Fig.3.3**). Global levels of H3K36me3 were reduced in the shHypb line (**Fig.3.3**). Therefore, Hypb seems to be responsible for a majority of all the detectable H3K36me3 in C2C12 cells. This has also been previously shown by Edmunds et al. in 2008 (**73**).

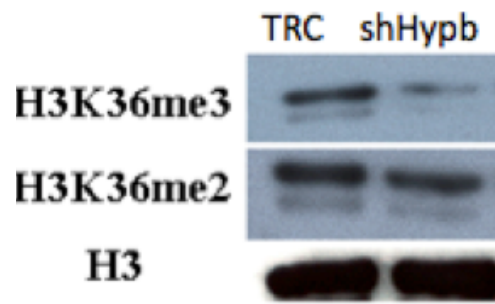


Figure 3.3. Western blot of global histone extract. Histones from three-day differentiated (day 3) TRC and shHypb lines were acid extracted overnight at 4°C and immunoprecipitated with H3K36me3, and H3 antibodies. H3 acts as a loading control.

3.4 Enhanced myogenesis phenotype upon lentiviral-mediated stable knockdown of Hypb in C2C12 cells

After establishing stable knockdown of Hypb, cells were differentiated for three days in differentiation medium at which point the cells were harvested for immunofluorescence, gene expression analysis, and western blot.

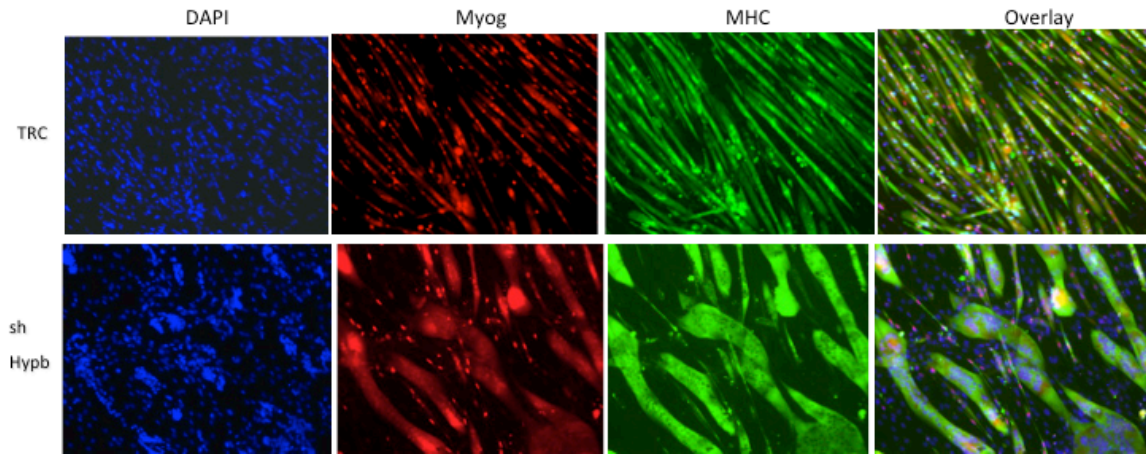
Previous work in our laboratory has shown that the Myog and Ckm genes play an important role in establishing muscle-specific gene expression. During development, Myog expression is restricted to the skeletal muscle lineage (74). During myogenesis, their expression levels differ from each other in a temporal fashion, which makes them a good model for studying muscle-specific gene expression. The TRC (control) line and the Hypb knockdown (shHypb) line were fixed and stained for differentiation markers Myog and MHC. The former stains the nucleus, whereas the latter stains the myotube cytoplasm. DAPI was used as a fluorescent nuclear counterstain (DAPI binds the A-T rich regions of the DNA) (Fig. 3.4A). After three days of differentiation, a significant difference in phenotype between the TRC and shHypb lines was seen. A 'enhanced myogenesis' phenotype was observed in the shHypb line where in contrast to the TRC cell line, the cells formed thicker caliber and longer myotubes with more number of nuclei (Fig. 3.4A). This effect was quantitated by comparing fusion indices (number of nuclei/myotube) of TRC and shHypb lines (Fig. 3.4D). The fusion index was almost 4-fold higher in the shHypb line (n=50, p<0.05).

Next, we looked at the gene expression levels of early and late differentiation markers comparing TRC and shHypb lines after subjecting cells to three days of differentiation. There was a significant increase in the expression of Myog, MHC, Ckm, and α -Actin upon Hypb knockdown (**Fig. 3.4B**). Tnnt2 expression did not change significantly. Relative expression for all genes was normalized with the housekeeping gene GAPDH.

At the protein level, MHC expression was higher in the shHypb line. The Myog protein expression levels did not seem to differ significantly upon Hypb knockdown. Beta-tubulin was used as the loading control (**Fig. 3.4C**).

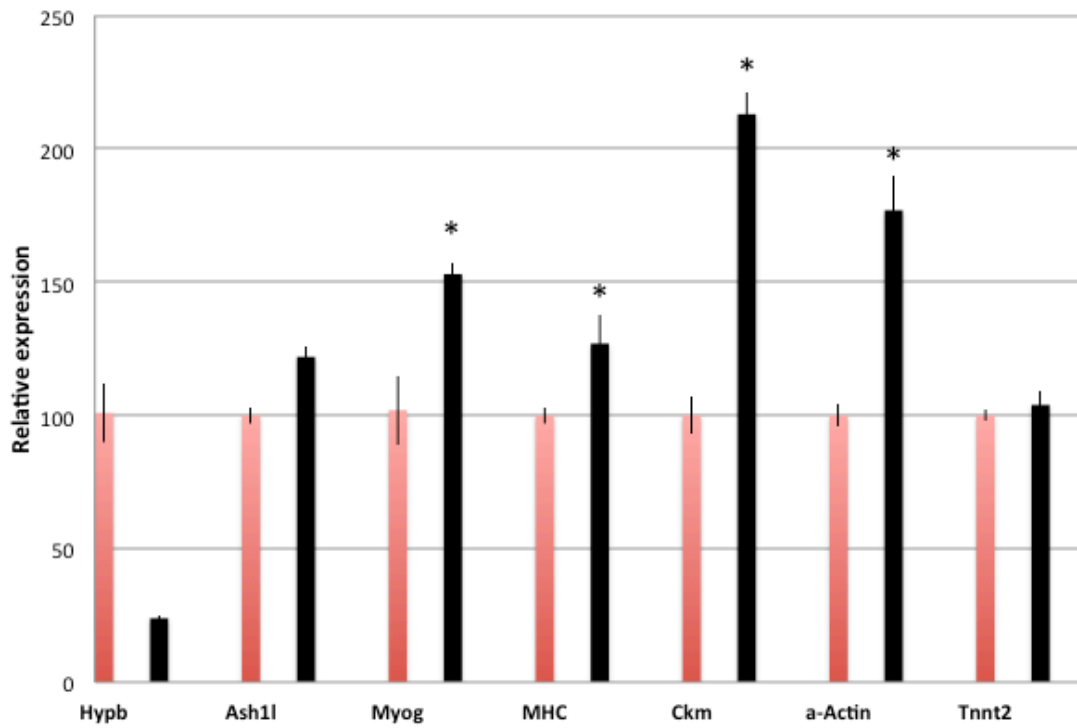
A

Phenotype characterization

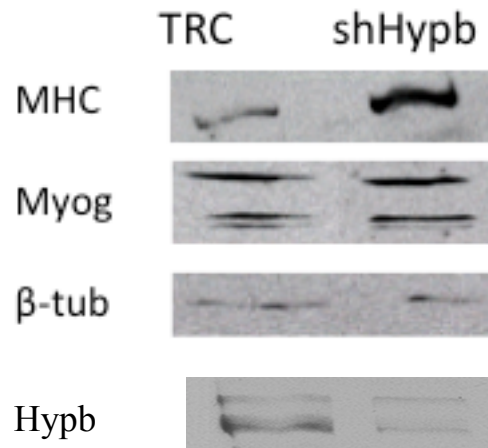


B

Gene expression levels in stable TRC and shHypb lines



C



D

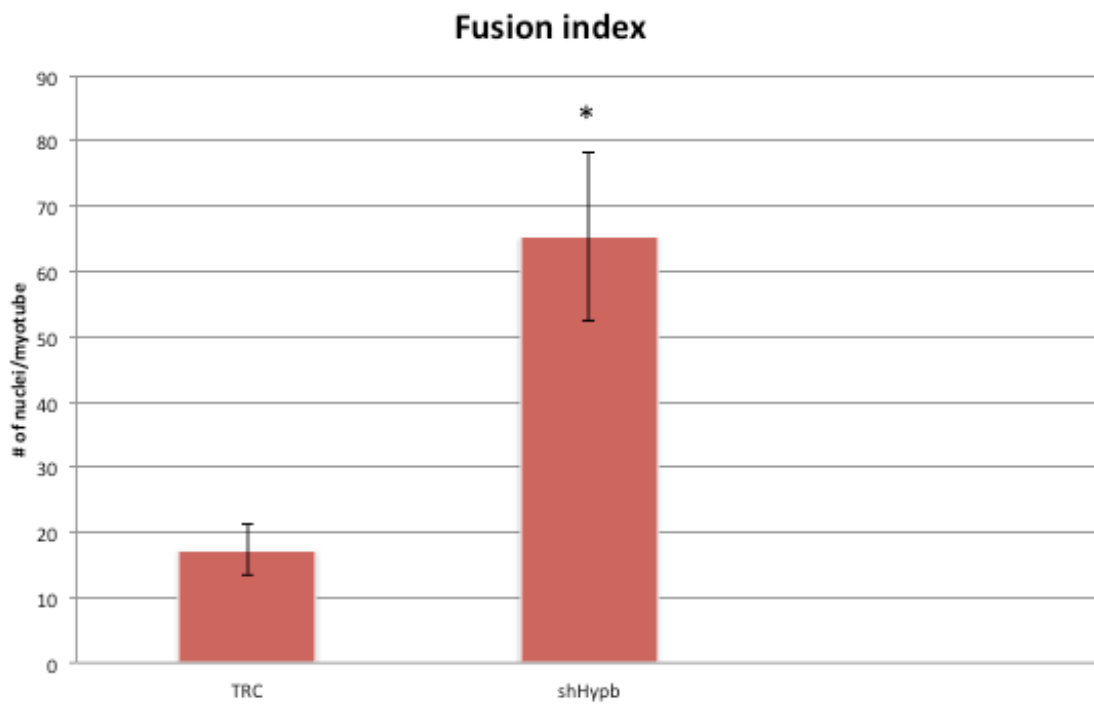


Figure 3.4A- Lentiviral-mediated stable knockdown of Hypb comparing them to the control (TRC). All lines were differentiated for 72 hours before being fixed for immunofluorescence and stained for myogenin (Red), myosin heavy chain(Green) and nucleus (Blue).

Figure 3.4B- RT-qPCR showing mRNA expression of differentiation markers in stable TRC and shHypb lines. Expression was normalized with GAPDH The experiment was performed in triplicates. Changes that are significant ($p < 0.05$) compared to the control are denoted by an asterisk.

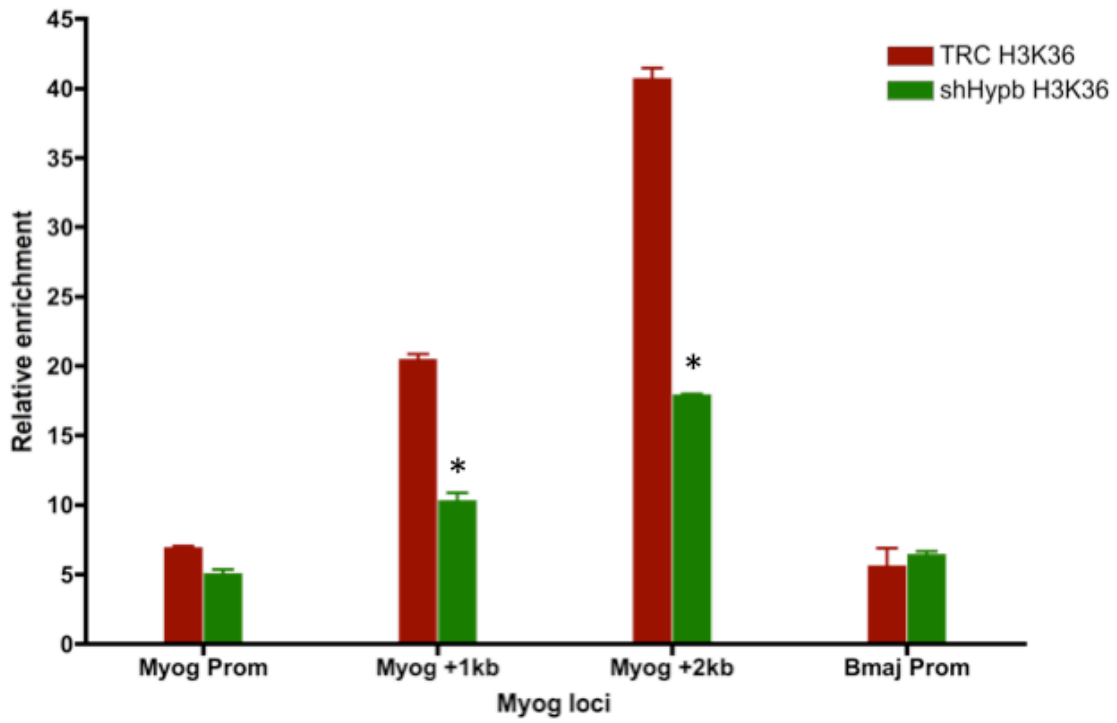
Figure 3.4C- Western blot displaying levels of differentiation markers in 72-hour differentiated TRC and shHypb lines. Beta-tubulin is the loading control.

Figure 3.4D. Fusion index graph showing the average number of nuclei per myotube in TRC and shHypb knockdown lines ($n=50$). Changes that are significant ($p < 0.05$) compared to the control are denoted by an asterisk.

3.5 The H3K36me3 mark is lost on the coding regions of muscle-specific genes in the shHypb cell line

To determine the effect of Hypb knockdown on H3K36me3 mark in muscle-specific genes, we performed N-ChIP comparing stable TRC and shHypb lines in differentiation conditions. As expected, H3K36me3 enrichment increased in the TRC line as we moved away from the promoter region and towards the 3' end of the Myog and Ckm genes (n=3, p<0.05) **(Fig.3.5)**. This increase was significant as compared to the promoter regions for both genes as well as the negative control binding region Bmaj promoter (Beta-major promoter is one of the regulatory regions of the beta-globin gene that regulates transcription in the hematopoietic system but is not regulated in the muscle lineage). The enrichment was significantly higher at the Myog +1kb and yet higher on the Myog +2kb region. Similarly, the enrichment on the Ckm gene was observed to be the highest, farthest from the promoter, at the +5kb loci. Upon Hypb knockdown, a significant loss of the H3K36me3 was observed at all the loci where significant enrichment was seen in the TRC line. This was seen at the Myog +1kb and +2kb on the Myog gene as well as at the Ckm +4kb and +5kb on the Ckm gene (n=3, p<0.05) **(Fig.3.5)**. Therefore, upon Hypb knockdown, there is a subsequent loss of H3K36me3 in the coding regions of Myog and Ckm genes in differentiating C2C12 cells.

H3K36me3 enrichment on Myog gene



H3K36me3 enrichment on the Ckm gene

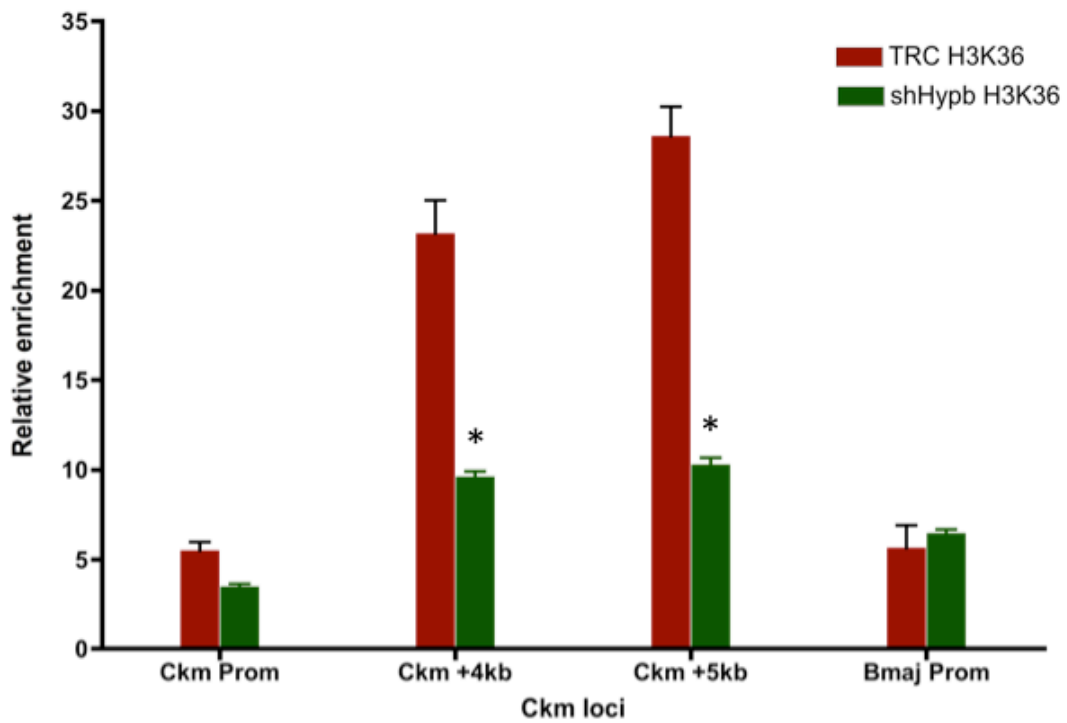


Figure 3.5. H3K36me3 enrichment on the Myog gene by N-ChIP. ChIP-qPCR graph displaying relative enrichment levels of H3K36me3 at three positions along the Myog gene (**Top**) and three positions along the Ckm gene (**Bottom**) in differentiated (day 3), stable TRC and shHypb knockdown cell lines. Rb IgG was used as a negative control for antibody-binding and the levels have been accounted for in relative enrichment calculations. Beta major promoter is a negative control binding region. Error bars represent standard error. The experiment was reproduced three times with similar results. Changes that are significant ($p < 0.05$) compared to the control are denoted by an asterisk.

3.6 Enrichment of the HMT Hypb is observed in the coding regions of muscle-specific genes

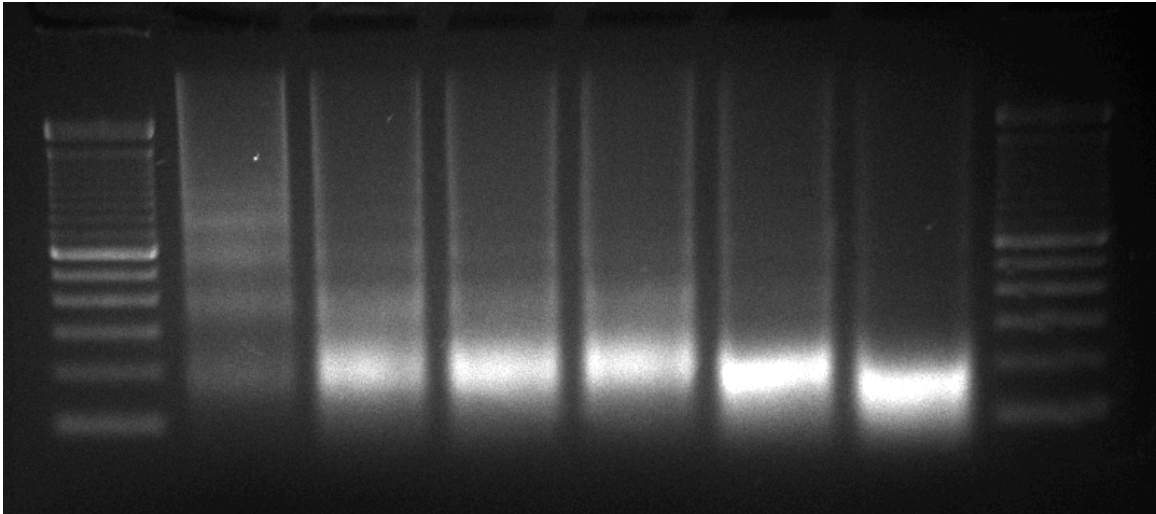
In order to investigate the loss of H3K36me3 mark upon Hypb knockdown, Hypb X-ChIP was undertaken as reduction in H3K36me3 levels on the coding regions of Myog and Ckm genes could be due to a direct interaction of this histone methyltransferase with the coding regions of muscle-development specific genes.

Similar to the N-ChIP procedure described earlier, optimization for X-ChIP was required, mainly due to the additional step of cross-linking in X-ChIP. Cross-linking of cells ensures tight-binding of otherwise loosely-bound transcription factors to the DNA such as Hypb. Therefore, in addition to MNase digestion, sonication of the DNA was also performed. Various sonication conditions were tried ranging from 10-60 cycles [**Fig.3.6A (i)**], however, the most optimal size of DNA fragments was obtained at 40 cycles of sonication post-MNase digestion [**Fig.3.6A (ii)**]. These conditions were therefore used for all subsequent X-ChIP experiments.

In order to observe whether we observe enrichment of Hypb levels on muscle-specific genes upon differentiation, X-ChIP was performed on wild-type C2C12 cells. Cells were differentiated, after which they were harvested and cross-linked (**Refer to Materials and Methods**). Chromatin-IP was performed, DNA purified, and ChIP-qPCR was undertaken. The loci chosen for ChIP-qPCR analysis were the same as the N-ChIP experiments. Hypb was enriched at +1kb and +2kb

regions of the Myog gene. Beta major promoter region was used as a negative control binding region **(Fig.3.6B)**. Similarly, for the Ckm gene, same loci as N-ChIP experiments were chosen. Hypb was enriched at +4kb and +5kb regions, consistent with H3K36me3 enrichment data at these loci. No Hypb enrichment is observed at Ckm promoter and beta major promoter regions, which is as expected **(Fig.3.6B)**. Therefore, upon differentiation Hypb appears to occupy the coding regions of muscle-specific genes.

(i)



(ii)

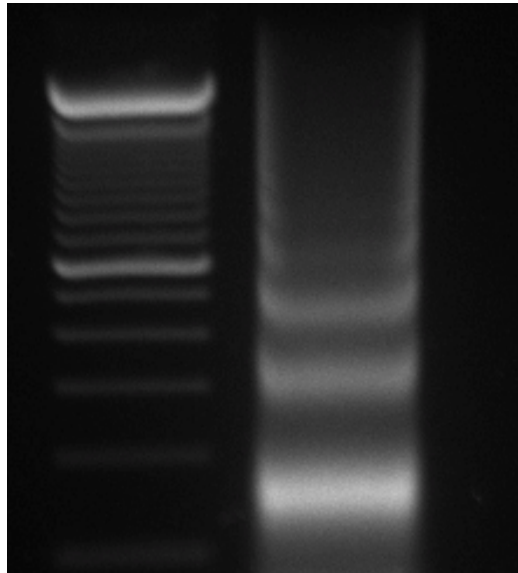


Fig.3.6A (i) Sonication optimization of chromatin for X-ChIP. From Left to Right: Sonication cycles ranging from 10-60 were preceded by MNase-digestion.

Fig. 3.6A (ii) Sonication optimization for X-ChIP. Ideal sample for X-ChIP experiments (40 cycles of sonication with MNase digestion). Reproduced multiple times and used it for all subsequent X-ChIPs.

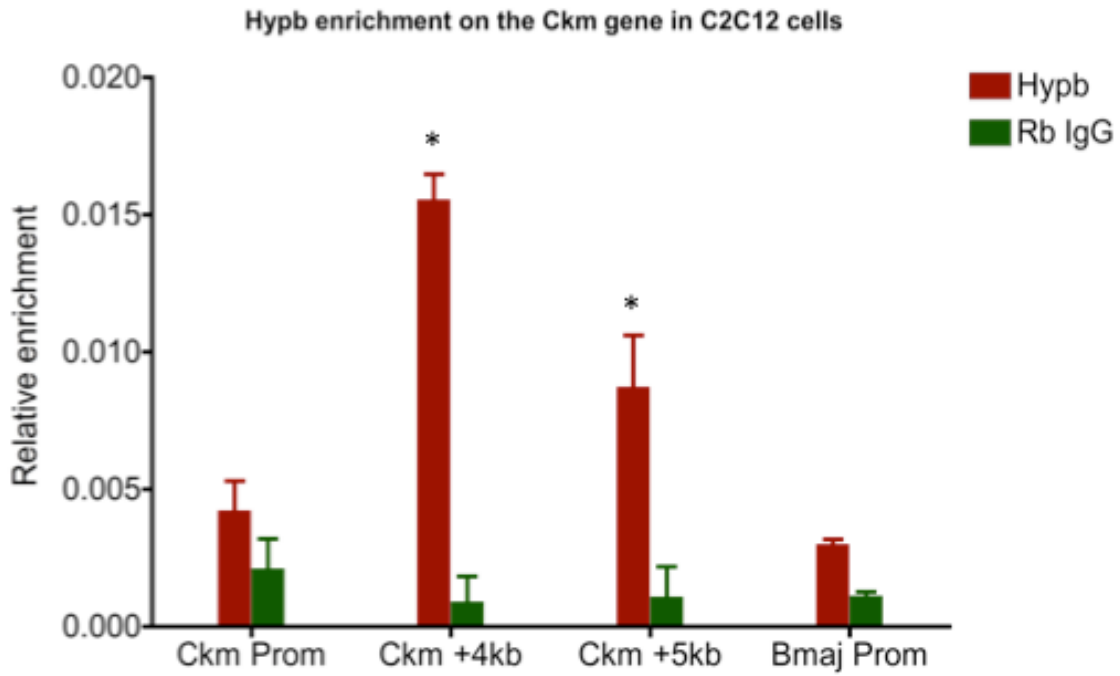
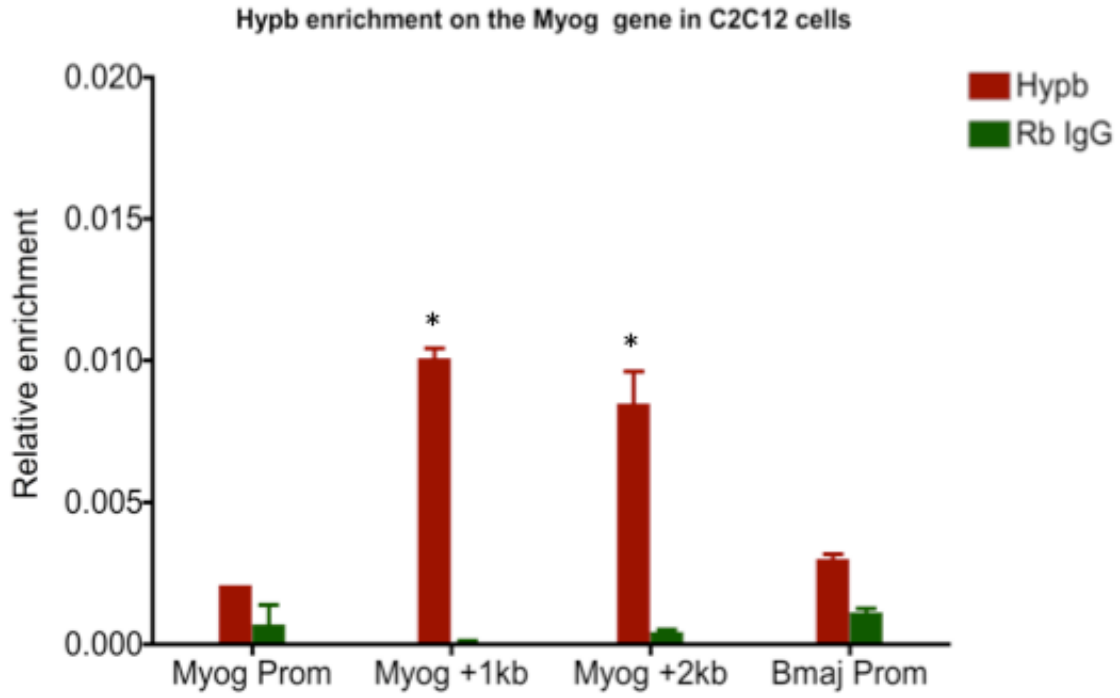


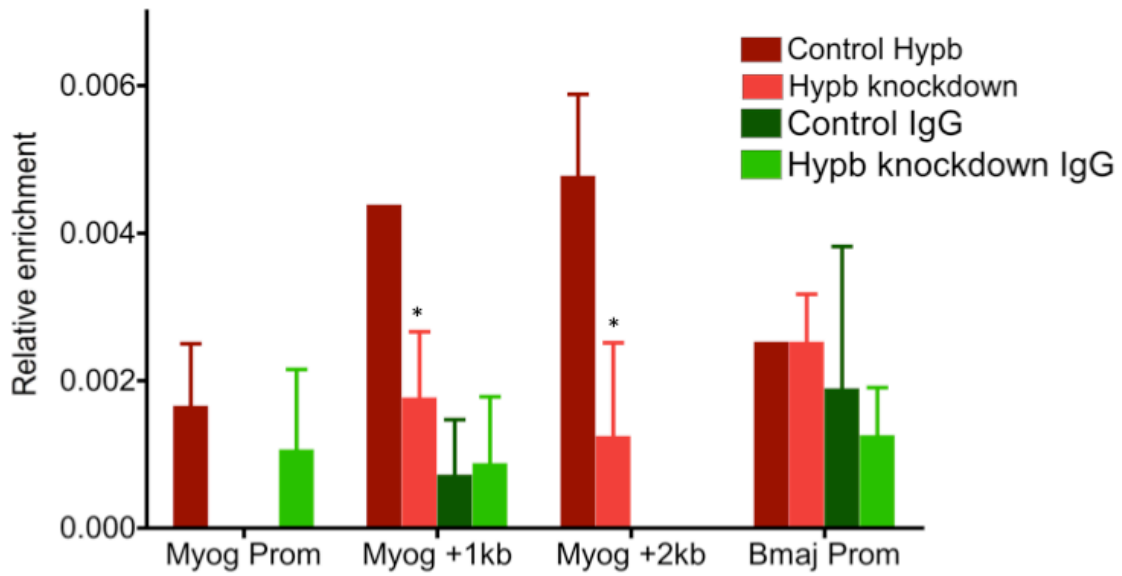
Figure 3.6B. Hypb enrichment on the Myog gene in wild-type differentiated (day 3) C2C12 cells by X-ChIP. ChIP-qPCR graph displaying relative enrichment levels of Hypb at three positions on the Myog gene (**Top**) and Ckm gene (**Bottom**). Rb IgG was used as a negative control for antibody-binding. Beta major promoter is a negative control binding region. Error bars represent standard error. Changes that are significant ($p < 0.05$) compared to the control are denoted by an asterisk.

3.7 Enrichment of Hypb at the coding regions of Myog and Ckm genes is lost upon its knockdown

In order to confirm the enrichment of Hypb on the coding regions of muscle-specific genes, X-ChIP on differentiated control TRC line and the stable knockdown line (shHypb) was performed.

As seen previously, Hypb enrichment was observed on +1kb and +2kb regions of the Myog gene and +4kb and +5kb regions of the Ckm gene (**Fig.3.6B**). The subsequent loss of Hypb enrichment is observed on these regions in the stable knockdown line, suggesting Hypb binding at the coding regions of these muscle-specific genes (**Fig.3.7**). As expected, Hypb does not enrich at the promoter regions of both genes as well as Bmaj promoter, the negative control binding region.

Hypb enrichment on the Myog gene in TRC and shHypb stable cell lines



Hypb enrichment on the Ckm gene in TRC and shHypb stable cell lines

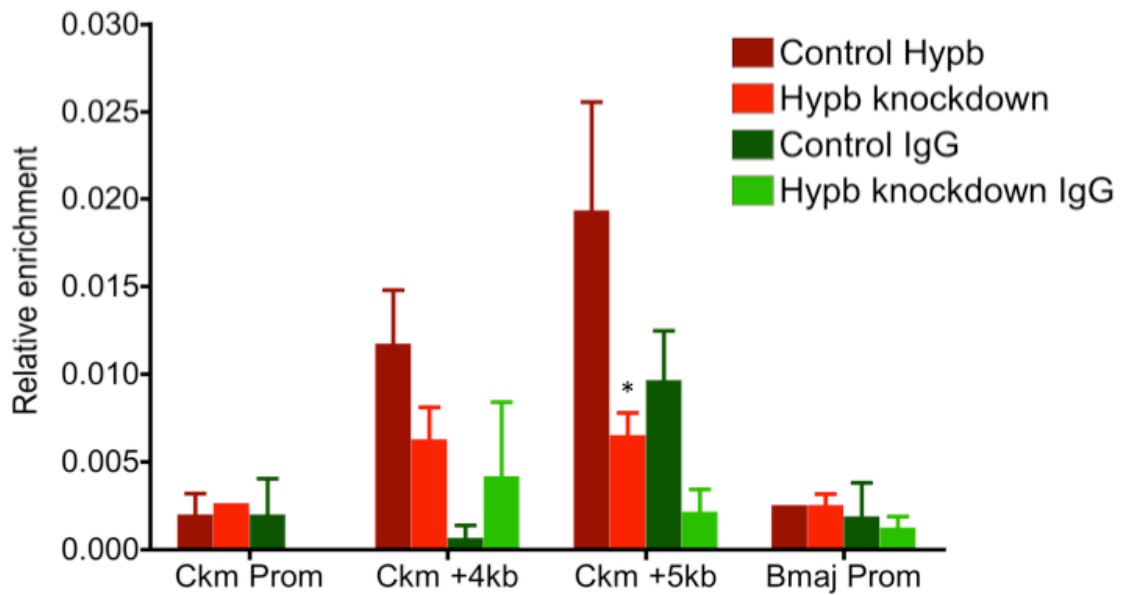


Figure 3.7. Hypb enrichment on the Myog and Ckm genes in 72h differentiated, stable TRC and shHypb knockdown lines by X-ChIP. ChIP-qPCR graph displaying relative enrichment levels of Hypb at three positions along the Myog gene (**Top**) and Ckm gene (**Bottom**). Rb IgG was used as a negative control for antibody-binding. Beta major promoter is a negative control binding region. Error bars represent standard error. Changes that are significant ($p < 0.05$) compared to the control are denoted by an asterisk.

CHAPTER 4

DISCUSSION

4.1 Summary of findings

In this study, we have examined role of the histone mark H3K36me3 in muscle differentiation. In order to establish the importance of this mark to muscle differentiation, the knowledge of histone methyltransferase responsible for this mark was pivotal. Therefore, the H3K36-specific HMT Hypb was knocked down. Interestingly, the Hypb knockdown resulted in enhanced myogenesis.

Next, it was important to compare H3K36me3 levels in myoblasts and myotubes. Therefore, a differentiation time course experiment was undertaken where upon differentiation, H3K36me3 enrichment levels went up on coding regions of two muscle-specific genes Myog and Ckm.

N-ChIP experiments were then performed on the TRC and Hypb knockdown (shHypb) lines on Myog and Ckm genes to look for H3K36me3 targets. Enrichment for this methylation mark was markedly higher in the coding regions of these genes (away from the 5' end and towards the 3' end). X-ChIP followed these experiments in order to see if Hypb is being enriched at these loci and has direct effects. The results showed, in general, it enriches on the coding regions of Myog and Ckm. In order to confirm these results, X-ChIP experiments on TRC and the knockdown lines

were undertaken. Enrichment of Hypb in the coding regions and its subsequent loss upon Hypb knockdown suggests possible occupancy of this transcription factor in the coding regions of muscle-specific genes.

In the end, a model for enhanced myogenesis is proposed which could explain the phenotype observed upon Hypb knockdown in C2C12 cells.

All these experiments have helped us identify the role of H3K36me3 in muscle differentiation and how epigenetic regulation of muscle-specific genes could be undertaken by H3K36-specific histone methyltransferases.

4.2 The coding regions of muscle-specific genes are enriched with the H3K36me3 mark and the H3K36 HMT Hypb upon muscle differentiation

Consistent with previous reports, we observed enrichment of H3K36me3 mark away from the promoter region and on the coding region of the genes of interest **(44)**. As shown previously the H3K36me3 mark plays a role in transcriptional elongation where Set2 is recruited to the coding region of the gene, rather than initiation of transcription, which is initiated by the H3K4me3 mark catalyzed by Set1 at the promoter region **(75)**. The presence of H3K36me3 mark on the promoter would mean inability for RNA Pol II to interact with Set1 and initiate transcription at the promoter **(75)**. Our results are also consistent with the observations of Asp et al (2011) where they provided ChIP-Seq genome-wide

dataset of H3K36me3 in myoblasts as well as myotubes and showed that muscle-specific genes such as various myosin heavy chain isoforms, namely, Myh1, Myh3, Myh4, and Myh7 show H3K36me3 enrichment upon muscle differentiation **(76)**, whereas, in our study we show Myog and Ckm as two additional genes that are enriched by this mark. In general, all these genes showed H3K36me3 enrichment in their respective coding regions.

Although the role of Hypb in muscle is not very well understood, de Almeida et al (2011) have shown that Hypb enrichment and its subsequent loss follows H3K36me3 loss on the CD44 gene in HeLa cell lysates **(77)**, similar to what we see on the coding regions of the Myog and Ckm genes in C2C12 cells.

4.3 Hypb is a major H3K36 trimethyltransferase in C2C12 cells

As has been reported previously in yeast, Set2 is the only HMT that is responsible for all three methylation events that is, mono-, di-, and trimethylation of H3K36 **(78)**. Upon knockdown of six HMTs --NSD1, NSD2, NSD3, Smyd2 (Data not shown), Ash11 (Refer to Appendix A1-A8), and Hypb at the offset of the project, only two HMTs showed visible changes in cellular phenotype upon their respective knockdown—these were Ash11 and Hypb. Histones extracted from the whole cell nuclear extract for control cell line and the shHypb line showed differences in H3K36me3 levels, whereby a reduction in trimethyl levels was seen in the shHypb line. This is consistent with what Edmunds et al **(73)** observed in their study that

Hypb is responsible for majority of H3K36me3 levels. Therefore, it is reasonable to conclude that Hypb is a major regulator of H3K36me3 mark in our cell model.

4.4 Hypb could prevent formation of cryptic transcripts by occupying the coding regions of muscle-specific genes - The 'enhanced myogenesis' model

It is well established in literature that H3K36me3 mark plays a role in transcriptional elongation in yeast **(79)**. In yeast, Set2 (yeast homolog of human Hypb) associates with phosphorylated RNA Pol II at Serine 2 in its C-terminal domain and marks the gene with H3K36me3. This mark sends a signal for the recruitment of Rpd3S, an HDAC complex. This is required for the suppression of cryptic initiation of transcription. This is achieved through histone deacetylation by the HDAC complex that makes the chromatin inaccessible to RNA Pol II and the transcription machinery.

In 2011, however, a study published by *Lin et al*, showed that this model could also exist in humans where they showed that a histone demethylase KDM5B is critical for embryonic stem cell (ESC) self-renewal **(80)**. Under normal conditions, H3K4me3 mark is restricted to the promoter-region and histone deacetylation takes place as usual, resulting in repression of cryptic initiation of transcription. However upon KDM5B knockdown, intragenic H3K4me3 marks bring the transcription initiation machinery into the coding regions of the gene that interfere with normal transcription causing inhibition of transcription elongation of KDM5B target genes.

In proliferating myoblasts or early differentiating myoblasts (Panel A), Hypb marks the muscle-specific genes involved in proliferation namely Myf5, Pax7, or MyoD, with H3K36me3 mark. This signals recruitment of the HDAC complex causing deacetylation of the genes, hence suppressing cryptic transcription from occurring. However, when we knockdown Hypb (Panel B), no H3K36me3 mark is present to signal deacetylation of the genes, hence our proposition is that cryptic transcripts are formed. The formation of these unspliced cryptic transcripts reduce the expression level of the aforementioned genes involved in proliferation and we would see precocious differentiation. Therefore, at an early stage of differentiation, Hypb could be acting as a positive regulator of proliferation. At a later stage in differentiation, however, Hypb binds the coding regions of differentiation marker genes such as Myog, Ckm (Data shown), and MHC and still prevents initiation of cryptic transcription. But upon Hypb knockdown, another HMT, for example, Ash1l, which is an H3K36me1 and H3K36me2 methyltransferase **(81)** takes Hypb place and marks genes preferentially with H3K36me1. It is known in the literature that H3K36me2 is a minimum requirement and is sufficient to prevent cryptic transcription from taking place **(82)**. Therefore, presence of Ash1l still results in cryptic transcription, but the differentiation phenotype could be owed to this HMT. Our data suggests that Ash1l is a positive regulator of differentiation (Fig. A1) as upon its knockdown, we observe inhibition of differentiation. Therefore, our proposition is that Ash1l binds the coding regions (Fig.A7) upon Hypb loss, causing cells to undergo accelerated differentiation, while cryptic transcription takes place

due to hyperacetylated histones.

Due to the dynamic nature of epigenetic marks, cryptic transcription could also be seen as an opportunity to regulate the expression of Myog and Ckm genes during different stages of muscle differentiation. At the initial stages of differentiation, Myog gene is highly expressed which could be attributed to the formation of normal gene products. However, during the late stages of differentiation, cryptic transcription might become active resulting in the formation of cryptic Myog transcripts that cause reduction in Myog gene expression. Similarly, for the Ckm gene, cryptic transcription could be active during early stages of differentiation when the Ckm gene is not highly expressed.

An alternative hypothesis is that Hypb associates with different factors on different genes. It has been shown in the literature that KDM5B acts as an activator of transcription in ESCs, whereas, in other cell types it acts to repress transcription **(83-85)**. Similarly, a demethylase belonging to the same family, KDM5A, has been shown to act an activator or repressor depending on the gene context **(86,87)**. This hypothesis overlaps with the previous hypothesis in a way that on early proliferation genes, Hypb acts to maintain their proliferation, while on differentiation genes, it acts as a negative regulator of differentiation. Therefore, Hypb might be interacting with different cofactors on two different genes resulting in the 'enhanced myogenesis' phenotype.

FUTURE DIRECTIONS

(a) Rescuing Hypb expression in C2C12 cells

Based on the results of this project, Hypb is identified as an important player in the establishment of H3K36me3 mark and muscle differentiation. One of the most important experiments in the future to do would be trying to rescue Hypb expression and observing the effect on muscle differentiation phenotype. This could be done by overexpressing the plasmid that contains a full-length Hypb transcript. Subsequent formation of long and thin myotubes could point to a potential link Hypb loss and the 'enhanced myogenesis' phenotype that was earlier observed. Also, the rescue experiment would confirm that the effect of the lentivirus is primary and that the enhanced differentiation we observe is not due to any secondary effects of the virus.

(b) Hypb and Ash1l ChIP-sequencing to look for binding sites and genes involved

Once we know that observed phenotype is actually due to the loss of Hypb, we would like to perform differentiation time course ChIP-Seq analysis which compares Hypb binding at day zero (proliferation) to 72h of differentiation in both wild-type C2C12 cells and shHypb cells. This would help identify possible Hypb binding sites on a genome-wide basis in C2C12 cells by which potential target genes could be analyzed and shortlisted. Similarly, differentiation time course ChIP-Seq

analysis of Ash1l could also be undertaken to know its binding sites, especially when Hypb expression is lost. This way we could potentially look at the spatial and temporal binding of both Hypb and Ash1l in C2C12 cells. To confirm the place and timing of binding, ChIP-Seq on Hypb/Ash1l overexpressed cells could be subsequently performed. These datasets could be correlated to existing H3K36me, H3K36me2, and H3K36me3 datasets or could be performed as ChIP-Seq experiments simultaneously. These experiments would help validate the 'enhanced myogenesis' model.

(c) Testing the cryptic transcription model

Our next step would be to test for cryptic transcription and see whether our proposed model is possible. For this, Northern blot analysis needs to be undertaken on total RNA samples from the TRC and shHypb knockdown lines during proliferation as well as after three days of differentiation. The blots will be probed by primer sets from various regions on Myf5, MyoD, Pax7, Myog, Ckm, and MHC genes. Cryptic transcripts always appear as shorter transcripts as compared to the full-length transcript on the Northern blot. Once we identify formation of cryptic transcripts, we can look for alternate transcription start sites other than the promoter region of the gene. For example, promoter deletion constructs of proliferation (Myf5, MyoD, and Pax7) and differentiation (Myog, Ckm, and MHC) genes could be undertaken followed by luciferase reporter assays. The activity of the putative promoter regions could be compared to that of the control SV40 promoter.

Any increase in the activity of these alternative regions might suggest presence of alternative transcription start sites and hence, generation of multiple unspliced cryptic transcripts.

(d) Testing for the alternative hypothesis

In order to look for potential interacting partners of Hypb, yeast two hybrid screening could be performed on wild-type and shHypb C2C12 cells at day zero and three days of differentiation. The interactions could then be validated using Hypb endogenous or overexpressed co-immunoprecipitation experiments. Once we have performed the yeast-two hybrid screening, we should look for muscle-specific factors that could be potential interactors of Hypb. The next step would be performing stable knockdown of the factors using shRNA and observing the differentiation phenotype. If the phenotype observed is similar to what we observe for Hypb knockdown cells, we could speculate that they could be potential cofactors that associate with Hypb either during proliferation or differentiation, depending on the nature of the cofactor and the gene of which it could be regulating the expression.

(e) Generation of a tissue-specific Hypb knockout mouse

Lastly, in the longer run, it would be useful to generate a tamoxifen inducible muscle-specific Hypb knockout mouse by using the Cre recombinase system to

study the effects of its loss on muscle regeneration *in vivo*. Muscle regeneration could be analyzed seven or 14 days post-injury (with a cardiotoxin injection). Based on the results we observed in our cell model, we speculate that the myotubes would undergo hypertrophy and it is possible that the muscle fibres would regenerate faster than their wild-type littermates.

References

1. Waddington, C.H.1957. *The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology*, London: Allen & Unwin.
2. Goldberg AD., Allis CD, and Bernstein E. 2007. Epigenetics: A landscape takes shape. *Cell*: 128: 635-8. (Licence Number: 2983720972993)
3. Fischle W, Wang Y, Allis CD. 2003. Histone and chromatin cross-talk. *Curr Opin Cell Biol*. 15:172–183.
4. Kim JR, Kee HJ., Kim YJ, Joung H, Nam H., Eom KI., Choe G, Kim N., Kim HS., Kook C., Kook H. 2009. Enhancer of polycomb1 acts on serum response factor to regulate skeletal muscle differentiation. *JBC*. 284: 16308-16.
5. Rudnicki MA, Le Grand F, McKinnell I, and Kuang S. 2009. The Molecular Regulation of Muscle Stem Cell Function. *CSH Symp Quant Biol*.
6. Bennett, A. M. and Tonks, N.K. 1997. Regulation of Distinct Stages of Skeletal Muscle Differentiation by Mitogen-Activated Protein Kinases. *Science*. 278: 1288-1291.
7. Zetser A, Gredinger E, Bengal E. 1999. p38 mitogen-activated protein kinase pathway promotes skeletal muscle differentiation. Participation of the Mef2c transcription factor. *JBC*. 274(8): 5193-200.
8. Rampalli S., Li L., Mak E., Ge K., Brand M., Tapscott S.J., and Dilworth F.J. 2007. p38 MAPK signaling pathway regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. *Nat. Struct. and Mol. Biol*. 14: 1150-1156.
9. Felsenfeld, G., and Groudine, M. 2003. Controlling the double helix. *Nature* 421:448-453. (Licence number: 2983741092826)
10. Kornberg RD, Lorch Y. 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*. 98:285–294.
11. Berger SL. 2002. Histone modifications in transcriptional regulation. *Curr Opin Genet Dev*. 12:142–148.
12. Fischle W, Wang Y, Allis CD. 2003. Histone and chromatin cross-talk. *Curr Opin Cell Biol*. 15:172–183.
13. Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell*. 128:707–719.
14. Henikoff, S. 2008. Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat. Rev. Genet*. 9: 15–26.
15. Hake, S.B., and Allis, C.D. 2006. Histone H3 variants and their potential role in indexing mammalian genomes: the “H3 barcode hypothesis”. *Proc. Natl. Acad. Sci. USA*. 103:6428–6435.
16. Hake, S.B., Garcia, B.A., Duncan, E.M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D.P., Allis, C.D., and Hunt, D.F. 2006. Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J. Biol. Chem*. 281:559–568.
17. Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. 2009.

- H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat. Genet.* 41:941–945.
18. Mito, Y., Henikoff, J.G., and Henikoff, S. 2005. Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* 37:1090–1097.
 19. Mito, Y., Henikoff, J.G., and Henikoff, S. 2007. Histone replacement marks the boundaries of cis-regulatory domains. *Science.* 315:1408–1411.
 20. Ray-Gallet, D., Quivy, J.P., Scamps, C., Martini, E.M., Lipinski, M., and Almouzni, G. 2002. HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. *Mol. Cell.* 9:1091–1100.
 21. Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. 2004. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell.* 116:51–61.
 22. Smith, S., and Stillman, B. 1989. Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. *Cell.* 58:15–25.
 23. Ahmad, K., and Henikoff, S. 2002. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell.* 9: 1191–1200.
 24. Kouzarides T. 2007. Chromatin modifications and their functions. *Cell.* 128(4):693-705.
 25. Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell.* 128:707–719.
 26. Li Y et al. 2009. The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *JBC.* 284:34283-295.
 27. Allis, C.D., Jenuwein, T., Danny Reinberg, and Caparros, M.L. 2006. *Epigenetics.* New York, CSH lab.press.
 28. Brown, C.E., Lechner, T., Howe, L., and Workman, J.L. 2000. The many HATs of transcription coactivators. *Trends Biochem. Sci.* 25:15–19.
 29. Kurdistani, S.K., and Grunstein, M. 2003. Histone acetylation and deacetylation in yeast. *Nat Rev Mol Cell Biol.* 4:276-284.
 30. Li, B., Carey, M., and Workman, J.L. 2007. The role of chromatin during transcription. *Cell* 128, 707-719.
 31. Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. *Genes and Dev.* 25(10): 1010-22.
 32. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science.* 324(5929):930-5.
 33. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature.* 466(7310):1129-33.
 34. Serandour AA, Avner S, Oger F, Bizot M, Percevault F, and Salbert G., 2012. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucl. Acid. Res.* 1-11.

35. Akam M. 1987. The molecular basis for metameric pattern in the *Drosophila* embryo. *Development*. 101:1-22.
36. Ingham PW. 1998. Trithorax and the regulation of homeotic gene expression in *Drosophila*: a historical perspective. *Int. J. Dev. Biol.* 42:423-29.
37. Ingham PW. 1983. Differential expression of bithorax complex genes in absence of extra sex combs and trithorax genes. *Nature*. 306:591-93.
38. Schuettengruber B. Martinez AM., Iovino N., and Cavalli G. 2011. Trithorax group proteins: switching genes on and keeping them active. *Nat. Rev. Mol. Cell. Biol.* 12:799-814.
39. Schuettengruber B. and Cavalli G. 2009. Recruitment of Polycomb group complexes and their role in dynamic regulation of cell fate choice. *Development*. 136:3531-42.
40. Kim J., Guermah M., McGinty R.K., Lee J.S., Tang Z., Milne T.A., Shilatifard A., Muir T.W., Roeder R.G. 2009. RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell*. 137:459-471.
41. Wysocka J., Swigut T., Milne T.A., Dou Y., Zhang X., Burlingame A.L., Roeder R.G., Brivanlou A.H., Allis C.D. 2005. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*. 121:859-872.
42. Shilatifard A. 2008. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr. Opin. Cell Biol.* 20:341-348.
43. Landry J, Sutton A, Hesman T, Min J, Xu R, Johnston M, and Sternglanz R. 2003. Set2-catalyzed methylation of histone H3 represses basal expression of GAL4 in *Saccharomyces cerevisiae*. *MCB*. 23: 5972-5978.
44. Dmitry K. Pokholok, Christopher T. Harbison, Stuart Levine, Megan Cole, Nancy M. Hannett, Tong Ihn Lee, George W. Bell, Kimberly Walker, P. Alex Rolfe, Elizabeth Herbolsheimer, Julia Zeitlinger, Fran Lewitter, David K. Gifford, and Richard A. Young. 2005. Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast. *Cell*. 122: 521-27.
45. Wagner EJ. and Carpenter PB. 2012. Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* 13:115-26. (Licence Number: 2966831473701)
46. Rayasam. GV et al. 2003. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* 22:3153-63.
47. Li Y et al. 2009. The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *JBC*. 284:34283-295.
48. Lucio-Eterovic. AK et al. 2010. Role for the NSD1 methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA Pol II function. *PNAS*. 107:16952-957.
49. Qiao Q et al. 2011. The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *JBC*. 286:8361-68.
50. Martinez-Garcia et al. 2011. The MMSET HMT switches global histone methylation and alters gene expression in multiple myeloma cells. *Blood* 117:211-220.

51. Rahman S et al. 2011. The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Molecular Cell Biology*. 31:2641-52.
52. Edmunds JW., Mahadevan LC, and Clayton AL. 2008. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J*. 27:406-420.
53. Hu M et al. 2010. Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *PNAS*. 107:2956-61.
54. Yuan W. et al. 2009. Heterogeneous nuclear ribonuclear protein L is a subunit of human KMT3a/Set2 complex required for H3 Lys 36 trimethylation activity in vivo. *JBC*. 284:15701-07.
55. Tanaka Y. Katagiri Z., Kawahashi K, Kioussis D., and Kitajima S. 2007. Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene*. 397:161-168.
56. An S., Yeo KJ., Jeon YH, and Song JJ. 2011. Crystal structure of the human HMT ASH1L catalytic domain and its implications for the regulatory mechanism. *JBC*. 286:8369-74.
57. Gregory GD et al. 2007. Mammalian ASH1L is a HMT that occupies the transcribed region of active genes. *MCB*. 27:8466-79.
58. Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. 2010. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *JBC* 285(34): 26114-20.
59. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science*. 327(5968): 996-1000.
60. Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., et al. 2002. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *MCB*. 22, 1298-1306.
61. Adhvaryu, K.K., Morris, S.A., Strahl, B.D., and Selker, E.U. 2005. Methylation of histone H3 lysine 36 is required for normal development in *Neurospora crassa*. *Eukaryotic Cell* 4, 1455- 1464.
62. Morris, S.A., Shibata, Y., Noma, K., Tsukamoto, Y., Warren, E., Temple, B., Grewal, S.I., and Strahl, B.D. 2005. Histone H3 K36 methylation is associated with transcription elongation in *Schizosaccharomyces pombe*. *Eukaryotic Cell* 4, 1446-1454.
63. Hu M et al. 2010. Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *PNAS*. 107:2956-61.
64. Yuan W. et al. 2009. Heterogeneous nuclear ribonuclear protein L is a subunit of human KMT3a/Set2 complex required for H3 Lys 36 trimethylation activity in vivo. *JBC*. 284:15701-07.
65. Michael J. Carrozza, Bing Li, Laurence Florens, Tamaki Suganuma, Selene K. Swanson, Kenneth K. Lee, Wei-Jong Shia, Scott Anderson, John Yates, Michael P. Washburn, Jerry L. Workman. 2005. Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell*. 123: 581-92.

66. Eom GH, Kim KB, Kim JH, Kim JY, Kim JR, Kee HJ, Kim DW, Choe N, Park HJ, Son HJ, Choi SY, Kook H, Seo SB. 2011. Histone methyltransferase SETD3 regulates muscle differentiation. *JBC*. 286(40):34733-42.
67. Hu M et al. 2010. Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *PNAS*. 107:2956-61.
68. Chomczynski P. and Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem*. 162:156-59.
69. Kedzierski W. and Porter J. 1991. A novel non-enzymatic procedure for removing DNA template from RNA transcription mixtures. *BioTechniques*. 10:210-14.
70. E.W. Johns. 1967. A method for the selective extraction of histone fractions f2(a)1 and f2(a)2 from calf thymus deoxyribonucleoprotein at pH7. *Biochem J*. 105(2): 611-614.
71. Brand M., Rampalli S., Chaturvedi CP., and Dilworth FJ. 2008. Analysis of epigenetic modifications of chromatin at specific gene loci by native chromatin immunoprecipitation of nucleosomes isolated using hydroxyapatite chromatography. *Nature Protocols*. 3:398-409.
72. Yaffe D and Saxel O. 1977. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature Letters*. 270:725-27.
73. Edmunds JW., Mahadevan LC, and Clayton AL. 2008. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J*. 27:406-420.
74. Kim JR, Kee HJ., Kim YJ, Joung H, Nam H., Eom KI., Choe G, Kim N., Kim HS., Kook C., Kook H. 2009. Enhancer of polycomb1 acts on serum response factor to regulate skeletal muscle differentiation. *JBC*. 284: 16308-16.
75. Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F., and Allis, C.D. 2001. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev*. 15, 3286-3295.
76. Asp P., Blum R., Vethantham V., Parisi F., Micsinai M., Cheng J., and Dynlacht D.B. 2011. Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *PNAS*. 1-10.
77. De Almeida S.F., Grosso A.R., Koch F., Fenouil R., Carvalho S., Andrade J., Levezinho H., Gut M., Eick D., Gut I., Andrau J.C., Ferrier P., and Carmo-Fonseca M. 2011. Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struc. And Mol Biol*. 18: 977-83.
78. Chu, Y., Sutton, A., Sternglanz, R., and Prelich, G. 2006. The BUR1 cyclin-dependent protein kinase is required for the normal pattern of histone methylation by SET2. *MCB*. 26, 3029-3038.
79. Kim T. and Buratowski S. 2010. The role of cotranscriptional histone methylations. *CSH*. 75:95-102
80. Lin X., Pelz C., Wang. W., and Impey S. 2011. KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *EMBO J*. 30(8): 1473-84.

81. Tanaka, Y., Z. Katagiri, K. Kawahashi, D. Kioussis, and S. Kitajima. 2007. Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene*. 397: 161–168.
82. Li B, Jackson J, Simon MD, Fleharty B, Gogol M, Seidel C, Workman JL, Shilatifard A. 2009. Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. *JBC*. 284(12):7970-6.
83. Christensen J, Agger K, Cloos PA, Pasini D, Rose S, Sennels L, Rappsilber J, Hansen KH, Salcini AE, Helin K. 2007. RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. *Cell* 128: 1063–1076.
84. Klose RJ, Yan Q, Tothova Z, Yamane K, Erdjument-Bromage H, Tempst P, Gilliland DG, Zhang Y, Kaelin Jr WG. 2007. The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell*. 128: 889–900.
85. Yamane K, Tateishi K, Klose RJ, Fang J, Fabrizio LA, Erdjument-Bromage H, Taylor-Papadimitriou J, Tempst P, Zhang Y. 2007. PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol Cell*. 25: 801–812.
86. Chan SW, Hong W. 2001. Retinoblastoma-binding protein 2 (Rbp2) potentiates nuclear hormone receptor-mediated transcription. *JBC*. 276: 28402–28412.
87. Benevolenskaya EV, Murray HL, Branton P, Young RA, Kaelin Jr WG. 2005. Binding of pRB to the PHD protein RBP2 promotes cellular differentiation. *Mol Cell*. 18: 623–635.

APPENDIX

Phenotype characterization

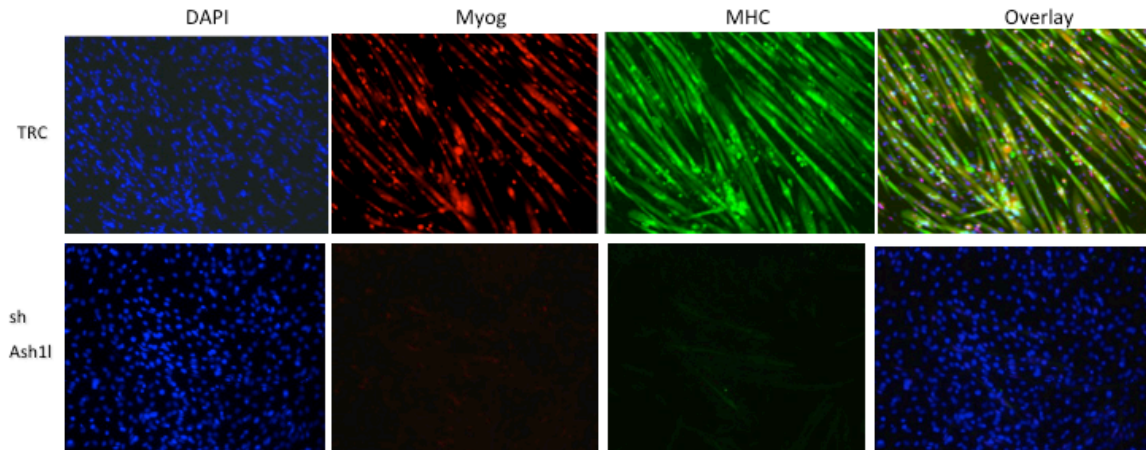


Figure A1- Lentiviral-mediated stable knockdown of Ash1l comparing them to the control (TRC). All lines were differentiated for 72 hours before being fixed for immunofluorescence and stained for myogenin (Red), myosin heavy chain(Green) and nucleus (Blue).

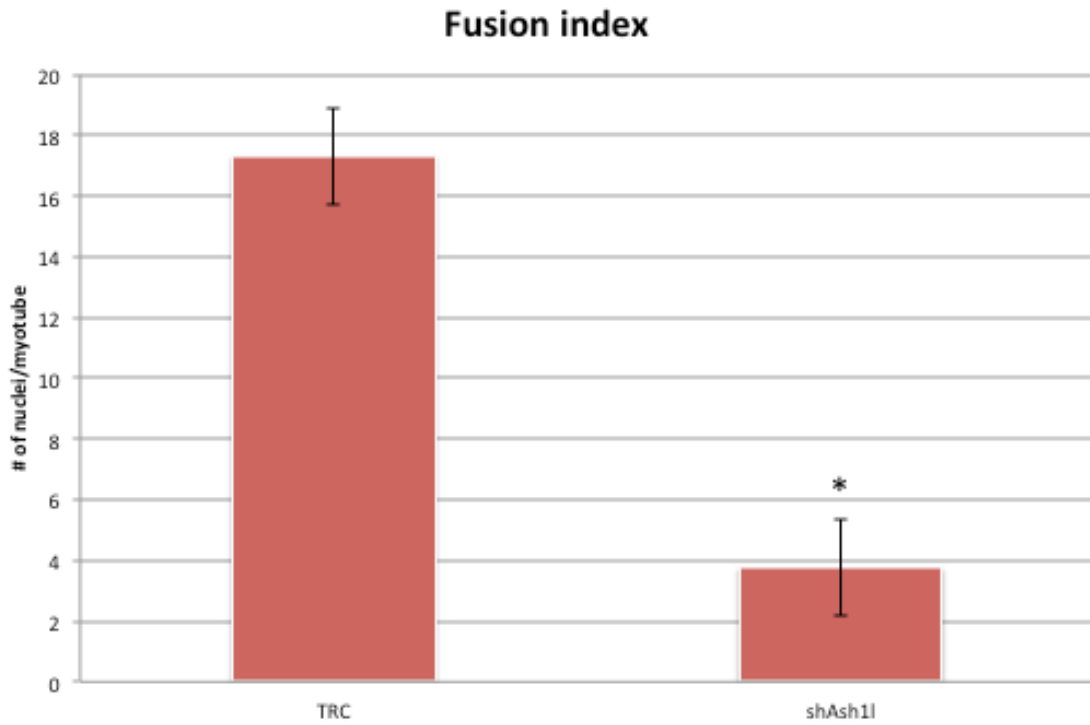


Figure A2. Fusion index graph showing the average number of nuclei per myotube in TRC and shAsh1l knockdown lines (n=50, p<0.05). The significance is denoted by an asterisk.

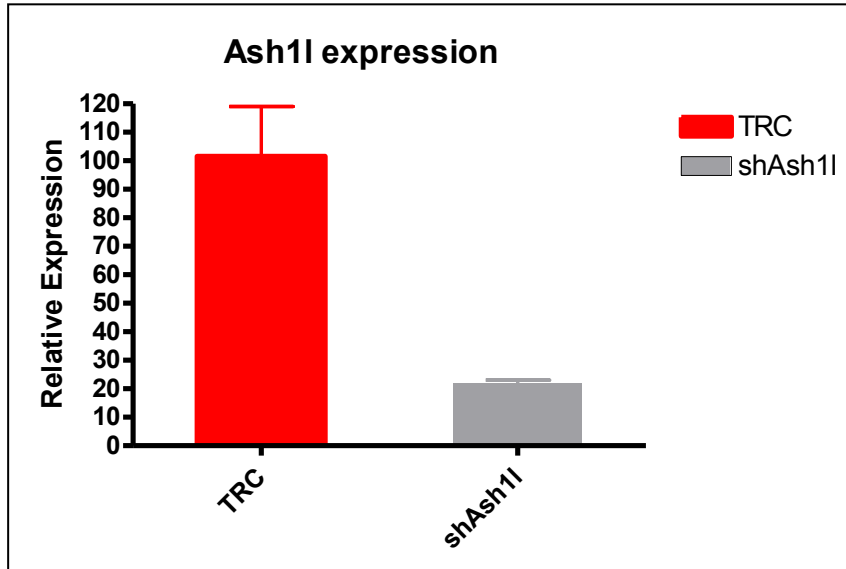
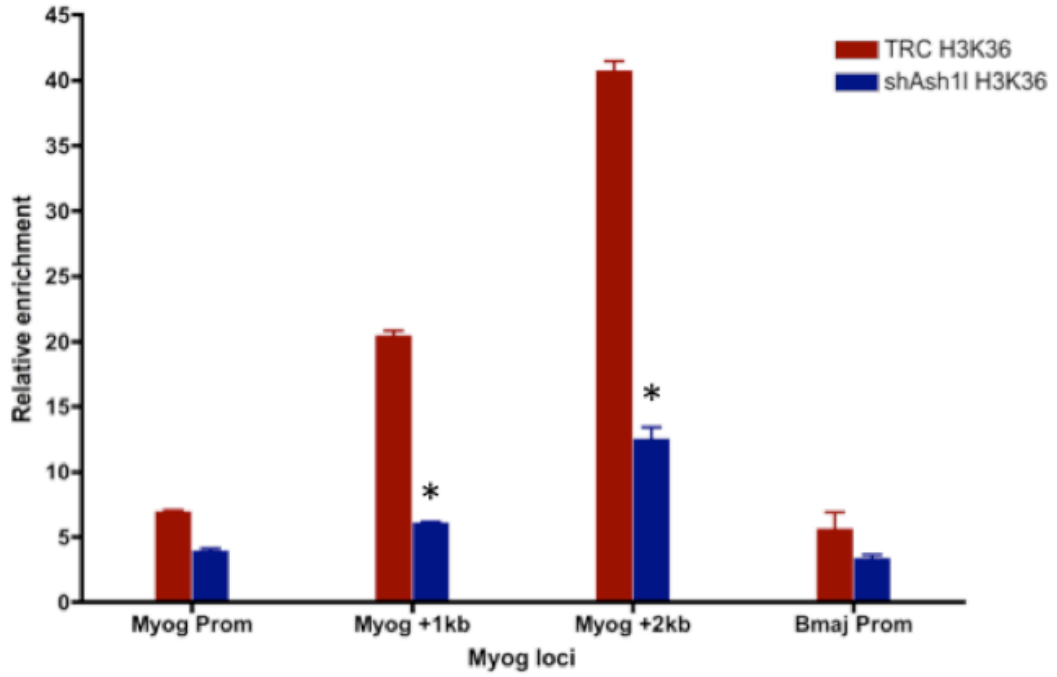


Fig.A3. Lentiviral-mediated shRNA knockdown of histonemethyltransferases Ash1l in C2C12 cells. Expression of the respective genes in the stable knockdown lines was compared to the TRC (control) line and plotted in a bar graph as relative expression. RNA from 100,000 cells/condition was isolated, reverse transcription performed. The cDNA was then used as a template for RT-qPCR. Samples were run in triplicates. Error bars represent standard error.

H3K36me3 enrichment on Myog gene



H3K36me3 enrichment on the Ckm gene

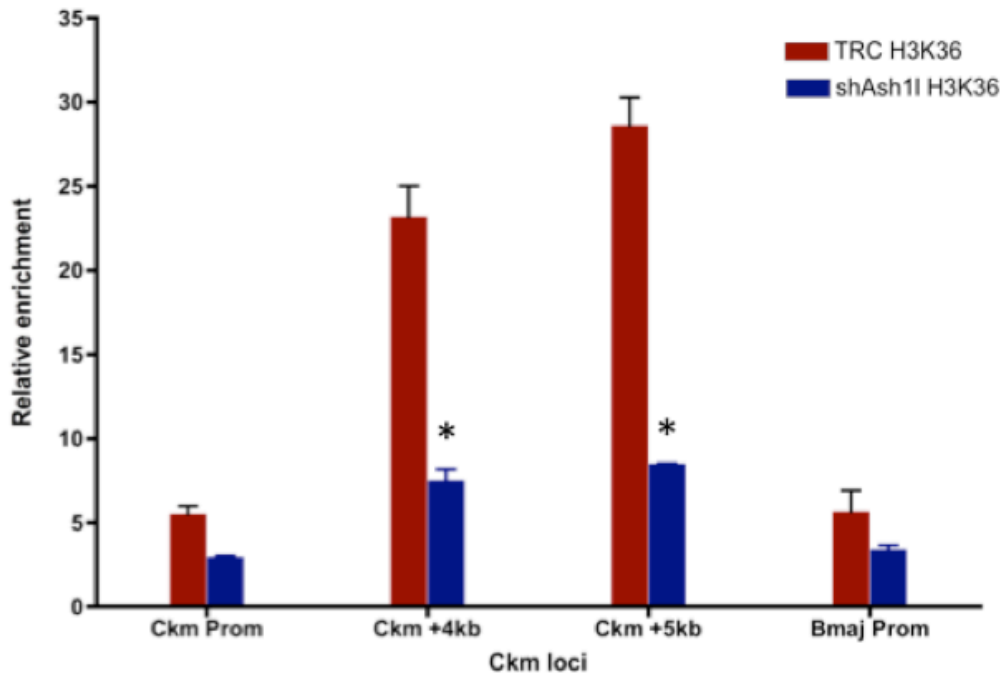
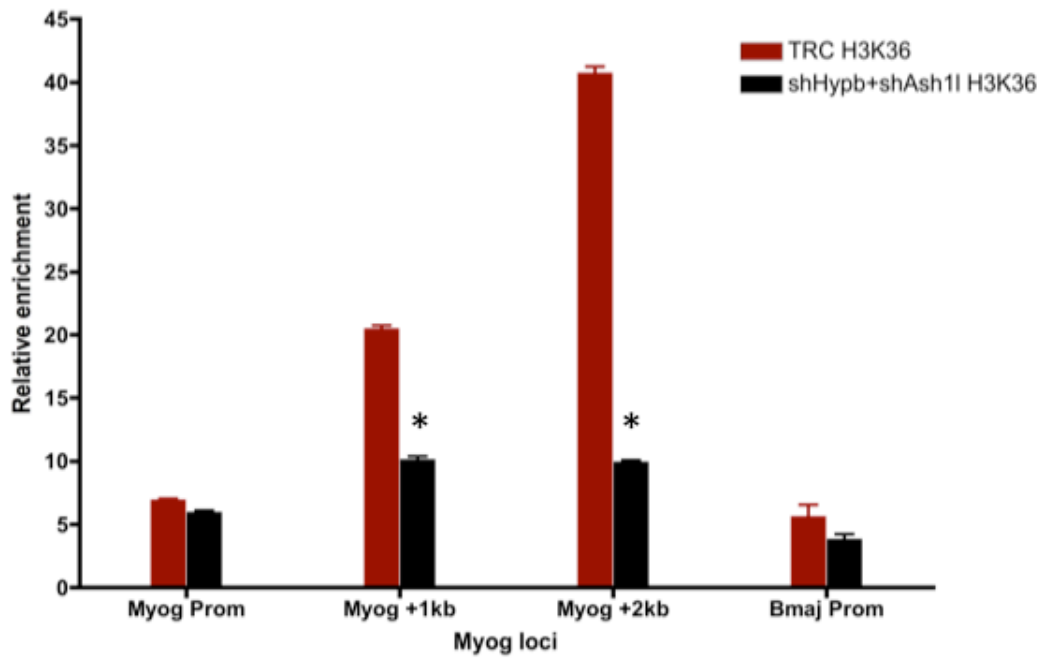


Figure A4. H3K36me3 enrichment on the Myog and Ckm genes by N-ChIP. **(Top)** - ChIP-qPCR graph displaying relative enrichment levels of H3K36me3 on three Myog gene loci and three Ckm gene loci **(Bottom)** in differentiated (day 3), stable TRC and shAsh1l knockdown cell lines. Rb IgG was used as a negative control for antibody-binding and the levels have been accounted for in relative enrichment calculations. Beta major promoter is a negative control binding region. Error bars represent standard error. The experiment was reproduced three times with similar results ($p < 0.05$). The significance is denoted by an asterisk.

H3K36me3 enrichment on Myog gene



H3K36me3 enrichment on the Ckm gene

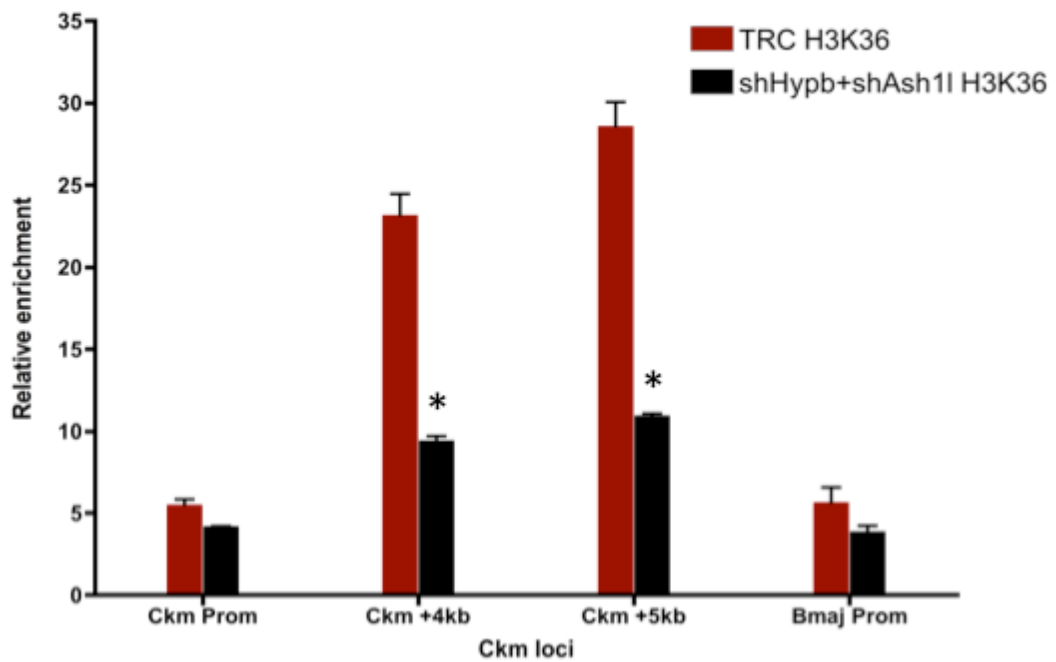


Figure A5. H3K36me3 enrichment on the Myog and Ckm genes by N-ChIP in the stable double knockdown line (shHypb+shAsh1l). **(Top)** - ChIP-qPCR graph displaying relative enrichment levels of H3K36me3 on three Myog gene loci and three Ckm gene loci **(Bottom)** in differentiated (day 3), stable TRC and double knockdown lines (shHypb+shAsh1l). Rb IgG was used as a negative control for antibody-binding and the levels have been accounted for in relative enrichment calculations. Beta major promoter is a negative control binding region. Error bars represent standard error. The experiment was reproduced three times with similar results ($p < 0.05$). The significance is denoted by an asterisk.

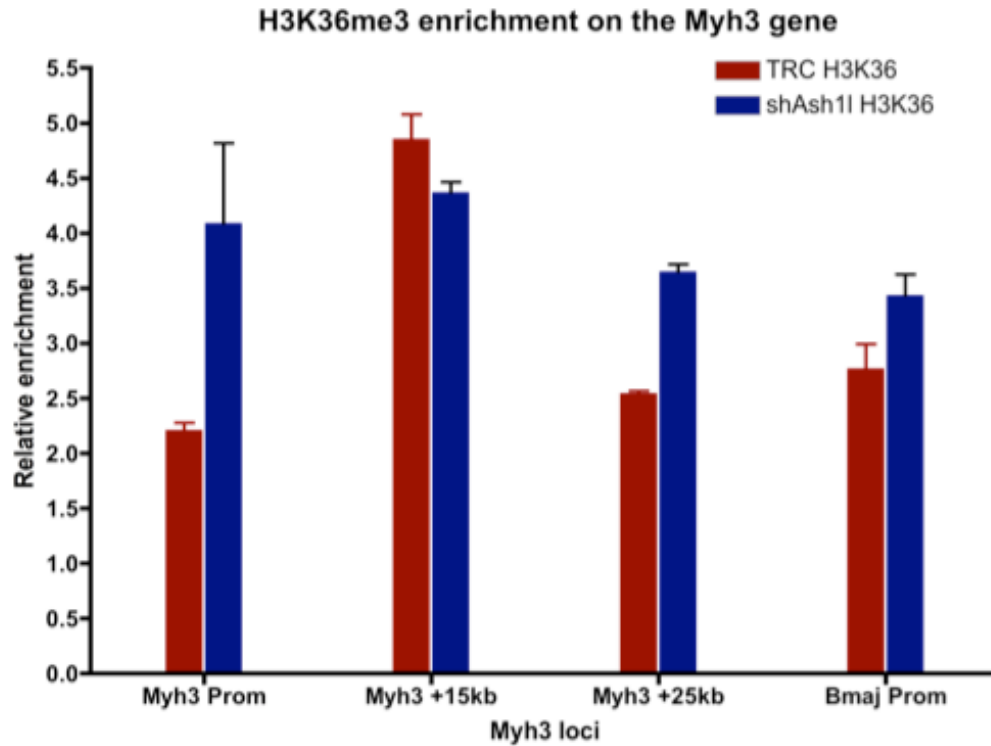
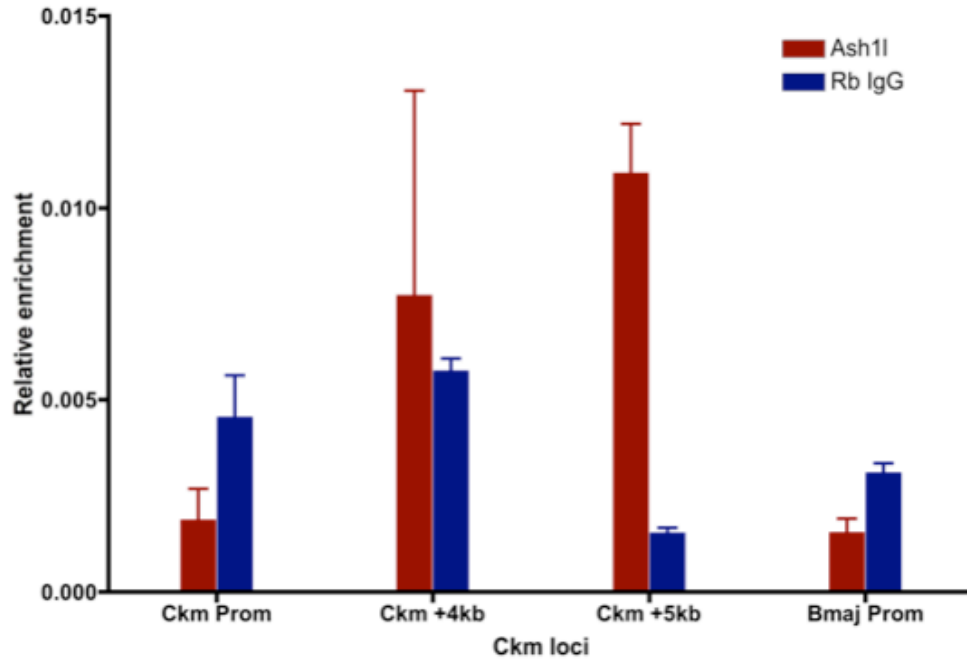


Figure A6. H3K36me3 enrichment on the Myh3 gene by N-ChIP. ChIP-qPCR graph displaying relative enrichment levels of H3K36me3 on three Myh3 gene loci in differentiated (day 3), stable TRC and shAsh1l knockdown cell lines. Rb IgG was used as a negative control for antibody-binding and the levels have been accounted for in relative enrichment calculations. Beta major promoter is a negative control binding region. Error bars represent standard error.

Ash1l and Hypb enrichment on the Ckm gene



Ash1l and Hypb enrichment on Myog gene

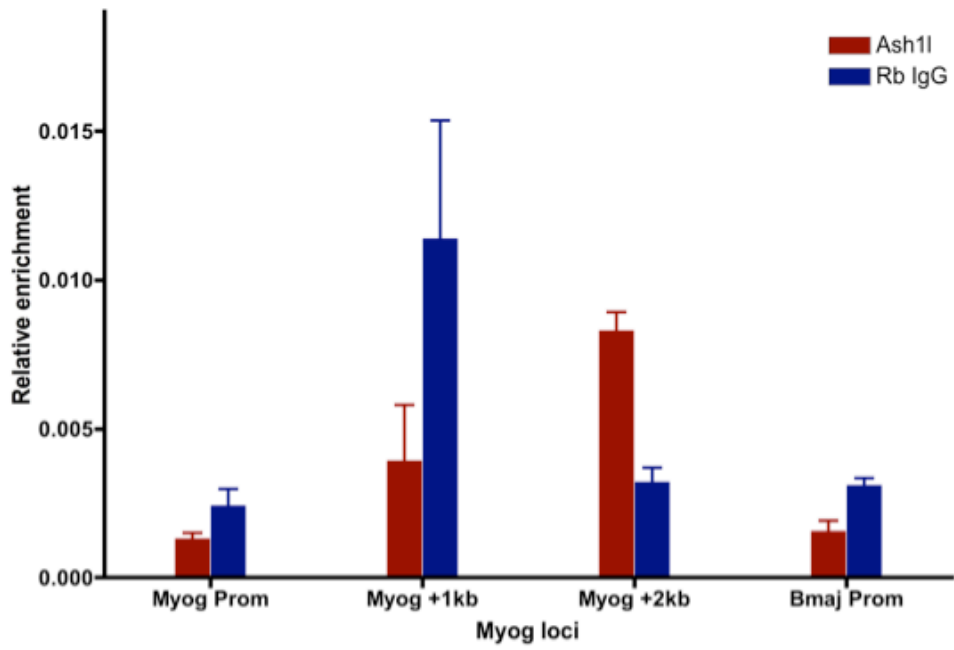
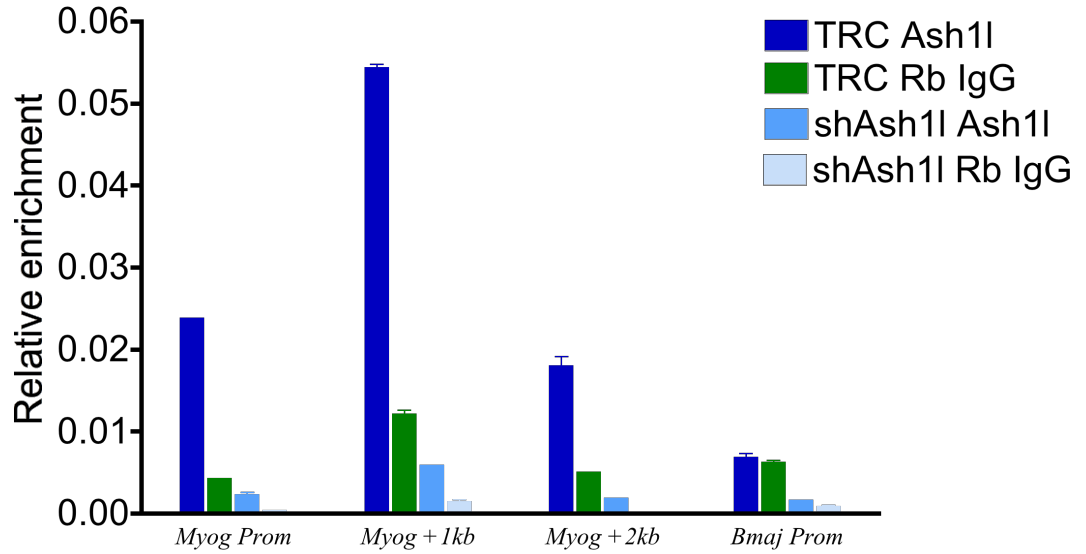


Figure A7. Ash1l enrichment on the Ckm gene (**Top**) and Myog gene (**Bottom**) in wild-type differentiated (day 3) C2C12 cells by X-ChIP. ChIP-qPCR graph displaying relative enrichment levels of Ash1l on three Myog gene loci. Rb IgG was used as a negative control for antibody-binding. Beta major promoter is a negative control binding region. Error bars represent standard error.

**Ash1l enrichment
on Myog gene in
TRC and shAsh1l lines**



**Ash1l enrichment
on Ckm gene in
TRC and shAsh1l lines**

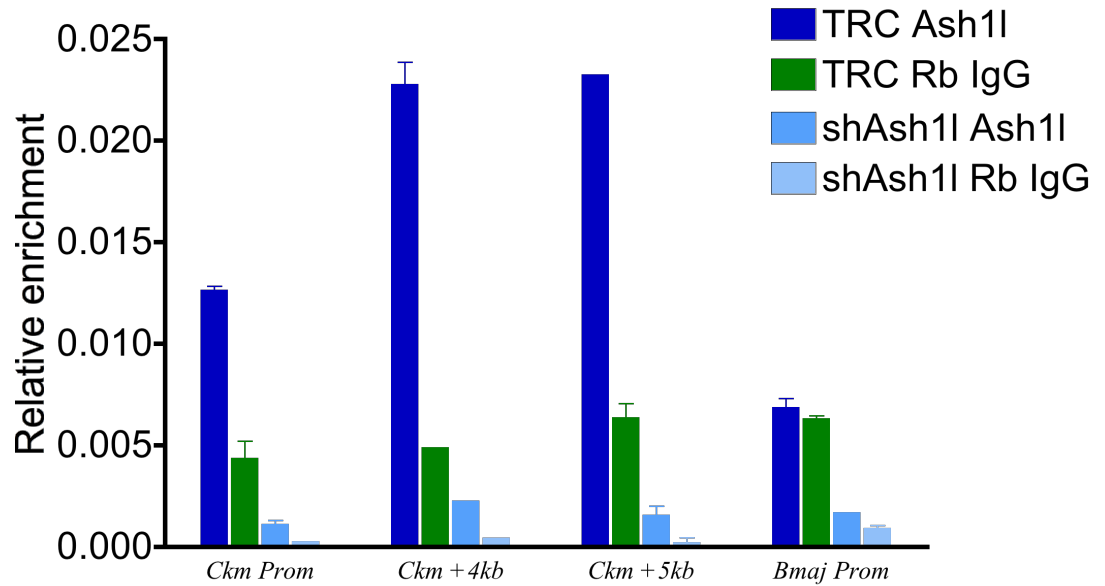


Figure A8. Ash1l enrichment on the Myog gene (**Top**) and Ckm gene (**Bottom**) in 72h differentiated, stable TRC and shAsh1l knockdown lines by X-ChIP. ChIP-qPCR graph displaying relative enrichment levels of Hypb on three Myog gene loci and three Ckm loci. Rb IgG was used as a negative control for antibody-binding. Beta major promoter is a negative control binding region. Error bars represent standard error.