CHARACTERIZATION OF THAP10 AND THAP11 AS TRANSCRIPTIONAL REPRESSORS IN DNA DAMAGE AND COLON CANCER PROGRESSION

James B. Parker

University of Pennsylvania, jbparker@mail.med.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/edissertations

Part of the Molecular Biology Commons

Recommended Citation

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/462
For more information, please contact repository@pobox.upenn.edu.
CHARACTERIZATION OF THAP10 AND THAP11 AS TRANSCRIPTIONAL REPRESSORS IN DNA DAMAGE AND COLON CANCER PROGRESSION

Abstract
The THAP (Thanatos associated protein) domain is an evolutionarily conserved zinc-finger motif highly similar to the sequence specific DNA binding domain of Drosophila P element transposase. Emerging data suggest THAP proteins may function in DNA and chromatin dependent processes, including transcription. However, the transcriptional regulatory function, mechanisms of action, and role of most THAP proteins in normal and aberrant cellular processes remain largely unknown.

In this thesis, we demonstrate that several human THAP proteins contain transcriptional repressor activity and specifically identify THAP10 and THAP11 as differentially expressed in human DNA damage and colon cancer progression, respectively. THAP10 and THAP11 repressed basal and VP16 activator driven transcription when tethered to promoters as heterologous Gal4-DNA binding domain fusion proteins and physically associated with histone deacetylases in vitro and in vivo. THAP11 was found to be differentially expressed in the SW480/SW620 cell culture model of human colon cancer progression and immunohistochemical analysis of tissue microarrays similarly revealed increased THAP11 expression concomitant with disease progression. The increase in THAP11 expression in colon cancer tumors and cell lines suggests that THAP11 dependent transcriptional repression may contribute to disease progression. Consistent with this hypothesis we find that knockdown of THAP11 in metastatic SW620 colon cancer cells results in a modest but significant decrease in cell proliferation. Gene expression profiling in THAP11 depleted SW620 cells identified 80 differentially expressed genes, 70% of which were de-repressed by THAP11 knockdown. Directly repressed THAP11 gene targets were found to contain chromatin bound THAP11 near their transcription start sites. THAP11 mediated repression requires the multi-functional transcriptional regulator HCF-1 (Host cell factor-1). THAP11 physically associates with and recruits HCF-1 to repressed promoters and knockdown of HCF-1 is sufficient to de-repress THAP11 target genes.

Collectively, this data provides the first characterization of a directly regulated, THAP11 dependent gene expression program in human cells and suggests THAP11 may be an important transcriptional regulator in human colon cancer. These results, in conjunction with previous findings from our laboratory and others, suggest THAP proteins likely function as biologically relevant transcriptional regulators.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Pharmacology

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/462
First Advisor
Debabrata Chakravartí

Subject Categories
Molecular Biology

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/462
CHARACTERIZATION OF THAP10 AND THAP11 AS TRANSCRIPTIONAL REPRESSORS IN DNA DAMAGE AND COLON CANCER PROGRESSION

James B. Parker

A DISSERTATION

in

Pharmacology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2010

Supervisor of Dissertation
Signature______________________
Debabrata Chakravarti, Associate Professor, Northwestern University

Graduate Group Chairperson
Signature______________________
Vladimir R. Muzykantov, Associate Professor of Pharmacology and Medicine

Dissertation Committee
Marcelo Kazanietz, Professor of Pharmacology
Richard Assoian, Professor of Pharmacology
Judy Meinkoth, Associate Professor of Pharmacology
Randall Pittman, Professor of Pharmacology
Acknowledgements

Foremost, I would like to thank Debu Chakravarti for providing an enthusiastic research environment and the material and intellectual support necessary to complete this work. I am grateful to Debu not only for his scientific guidance but also his aid in my professional development as well.

I am indebted to Marcelo Kazanietz, Rick Assoian, Judy Meinkoth, and Randy Pittman for their willingness to participate in a long distance thesis committee. I gratefully acknowledge the time and efforts invested by the committee members on my behalf and thank them for their support.

I would like to thank current and former members of the Chakravarti laboratory, especially Todd Macfarlan and Santanu Pal Chaudhuri, for providing reagents and preliminary observations central to this work. I am especially grateful to Manop Buranapramest for providing consistently thoughtful commentary and helpful suggestions.

I would also like to acknowledge our collaborators at Northwestern University and elsewhere who have assisted in our efforts to characterize THAP11 function in cancer cells. In this capacity, I am especially thankful to Jian-Jun Wei for providing expert histopathological analysis and Jennifer Koblinski for help in establishing tumor xenograft assays.

Finally, I would like to thank my wife, Rebecca. This work would not have been possible without her continued love, support, and patience.
ABSTRACT

CHARACTERIZATION OF THAP10 AND THAP11 AS TRANSCRIPTIONAL REPRESSORS IN DNA DAMAGE AND COLON CANCER PROGRESSION

James B. Parker
Debabrata Chakravarti

The THAP (Thanatos associated protein) domain is an evolutionarily conserved zinc-finger motif highly similar to the sequence specific DNA binding domain of Drosophila P element transposase. Emerging data suggest THAP proteins may function in DNA and chromatin dependent processes, including transcription. However, the transcriptional regulatory function, mechanisms of action, and role of most THAP proteins in normal and aberrant cellular processes remain largely unknown.

In this thesis, we demonstrate that several human THAP proteins contain transcriptional repressor activity and specifically identify THAP10 and THAP11 as differentially expressed in human DNA damage and colon cancer progression, respectively. THAP10 and THAP11 repressed basal and VP16 activator driven transcription when tethered to promoters as heterologous Gal4-DNA binding domain fusion proteins and physically associated with histone deacetylases in vitro and in vivo. THAP11 was found to be differentially expressed in the SW480/SW620 cell culture model of human colon cancer progression and immunohistochemical analysis of tissue microarrays similarly revealed increased THAP11 expression concomitant with disease progression. The increase in THAP11 expression in colon cancer tumors and cell lines suggests that THAP11...
dependent transcriptional repression may contribute to disease progression. Consistent with this hypothesis we find that knockdown of THAP11 in metastatic SW620 colon cancer cells results in a modest but significant decrease in cell proliferation. Gene expression profiling in THAP11 depleted SW620 cells identified 80 differentially expressed genes, 70% of which were de-repressed by THAP11 knockdown. Directly repressed THAP11 gene targets were found to contain chromatin bound THAP11 near their transcription start sites. THAP11 mediated repression requires the multi-functional transcriptional regulator HCF-1 (Host cell factor-1). THAP11 physically associates with and recruits HCF-1 to repressed promoters and knockdown of HCF-1 is sufficient to de-repress THAP11 target genes.

Collectively, this data provides the first characterization of a directly regulated, THAP11 dependent gene expression program in human cells and suggests THAP11 may be an important transcriptional regulator in human colon cancer. These results, in conjunction with previous findings from our laboratory and others, suggest THAP proteins likely function as biologically relevant transcriptional regulators.
Table of Contents

CHAPTER 1: REGULATION OF TRANSCRIPTION IN EUKARYOTES ............1

Overview ............................................................................................................. 1

Mechanisms of Regulation ................................................................................. 2
  Transcription Initiation and Elongation ........................................................... 2
  Chromatin and Transcriptional Regulation ....................................................... 8
  Histone Modification ......................................................................................... 9
  Chromatin Remodeling .................................................................................. 13

CHAPTER 2: CHARACTERIZATION OF THAP10 AND THAP11 AS
TRANSCRIPTIONAL REPRESSORS EXPRESSED IN HUMAN DNA DAMAGE
AND COLON CANCER PROGRESSION ..........................................................17

Summary ............................................................................................................. 17

Introduction ....................................................................................................... 18
  Introduction to the THAP Domain .................................................................... 18
  Structure-Function Analysis of THAP domains ................................................. 19
  THAP Proteins in Biological Processes .............................................................. 25

Experimental Procedures .................................................................................. 31
  Plasmids and Cloning ...................................................................................... 31
  Bacterial Protein Expression and Purification .................................................... 32
  Generation and Purification of THAP11 Antibody ........................................... 33
  Cell Culture and Treatment ............................................................................. 34
  Immunoblotting and Immunofluorescence ....................................................... 34
  Immunohistochemistry and Tissue Microarrays ............................................... 36
GST Pulldown Assays ............................................................................................................... 37
Co-immunoprecipitation ............................................................................................................ 37
RNA Isolation and Quantitative RT-PCR .................................................................................. 38
Gal4 Repression Assay ............................................................................................................. 38

Results .......................................................................................................................................... 39
THAP10 Expression is Induced by DNA Damage .................................................................... 39
THAP11 Expression in a Cell Culture Model of Colon Cancer Progression............................. 40
THAP11 Expression is Correlated with Colon Cancer Progression ......................................... 42
THAP10 and THAP11 Repress Transcription ........................................................................... 43
THAP10 and THAP11 Associate with HDACs .......................................................................... 44

Discussion .................................................................................................................................... 45

CHAPTER 3: IDENTIFICATION OF ENDOGENOUS THAP11 GENE
TARGETS AND A POTENTIAL ROLE FOR THAP11 IN COLON CANCER

CELL PROLIFERATION .......................................................................................... 66

Summary ...................................................................................................................................... 66

Introduction .................................................................................................................................. 67

Experimental Procedures ......................................................................................................... 68
Plasmids and Cloning .................................................................................................................. 68
Immunoblotting ........................................................................................................................... 70
Cell Culture ................................................................................................................................. 70
Retrovirus Production ................................................................................................................ 71
Retroviral Transduction and Stable Cell Production ................................................................. 71
RNA Isolation and Quantitative RT-PCR .................................................................................. 72
Microarray Gene Expression Analysis ...................................................................................... 73
Results ........................................................................................................................................... 79

Gene Expression Profiling in SW620-THAP11 Knockdown cells ............................................ 79
Identification of Direct THAP11 Gene Targets in SW620 cells .............................................. 84
RNAPII and Acetylated-Histone Occupancy at THAP11 Repressed Genes ................................. 86
THAP11 associates with HCF-1 .................................................................................................... 87
THAP11 recruits HCF-1 to promoters of THAP11 repressed genes .......................................... 89
HCF-1 is required for THAP11 mediated repression .................................................................. 90
Affinity purification of THAP11/HCF-1 protein complex .......................................................... 90
Annexin A1 is de-repressed by THAP11 knockdown ................................................................. 92
Cell growth suppression in THAP11 knockdown SW620 cells ................................................. 93

Discussion .................................................................................................................................... 95

CHAPTER 4: GENERAL CONCLUSIONS AND FUTURE DIRECTIONS ......131

APPENDIX ..................................................................................................................................140

BIBLIOGRAPHY .........................................................................................................................144
List of Tables

Table 2.1: THAP proteins in model organisms..............................................................................55
Table 3.1: Summary of genes re-repressed by THAP11 rescue .........................................................111
Table A.1: shRNA sequences .....................................................................................................140
Table A.2: shRNA cloning oligonucleotides ................................................................................140
Table A.3: Non-silenceable THAP11 site-directed mutagenesis primers ...................................141
Table A.4: ChIP primer sequences ..............................................................................................141
List of Figures

Figure 1.1: Mechanisms of transcriptional regulation ............................................................. 16
Figure 2.1: Human THAP domain containing proteins ............................................................ 54
Figure 2.2: THAP10 mRNA is DNA damage inducible .............................................................. 56
Figure 2.3: THAP10 mRNA induction by DNA damage in HepG2 and Saos-2 cells .................... 57
Figure 2.4: Ecotopic p53 expression induces THAP10 mRNA ............................................... 58
Figure 2.5: THAP11 expression in SW480 and SW620 colon cancer cell lines ....................... 59
Figure 2.6: THAP11 expression in colon cancer cell lines ....................................................... 60
Figure 2.7: Immunohistochemical analysis of THAP11 expression in human colon cancer specimens ............................................................................................................................. 61
Figure 2.8: THAP10 and THAP11 repress basal and activated transcription ....................... 62
Figure 2.9: Transcriptional regulatory properties of THAP4, THAP6, and THAP8 .................. 63
Figure 2.10: THAP11 associates with HDACs ............................................................................. 64
Figure 2.11: THAP10 associates with HDACs ............................................................................. 65
Figure 3.1: THAP11 knockdown by retrovirally delivered shRNA in SW620 cells ............... 107
Figure 3.2: Gene expression analysis in THAP11 knockdown SW620 cells ....................... 108
Figure 3.3: Nuclear run-on assay in THAP11 knockdown SW620 cells ............................. 109
Figure 3.4: Expression of non-silenceable THAP11 reverses gene expression induced by THAP11 knockdown .................................................................................................................. 110
Figure 3.5: THAP11 knockdown de-represses putative gene targets in Colo320HSR cells ....... 112
Figure 3.6: Chromatin immunoprecipitation (ChIP) analysis of THAP11-Rescue-3xFLAG binding at Lsmd1 and ncRNA00095 ............................................................................................................. 113
Figure 3.7: Chromatin immunoprecipitation (ChIP) analysis of THAP11-Rescue-3xFLAG binding at AA862256 and Praf2 .............................................................................................................. 114
Figure 3.8: Chromatin immunoprecipitation (ChIP) analysis of THAP11-Rescue-3xFLAG binding at C1orf83 and ZSCAN20 ................................................................................................. 115
Figure 3.9: Chromatin immunoprecipitation (ChIP) analysis of endogenous THAP11 .......... 116
Figure 3.10: Chromatin immunoprecipitation (ChIP) analysis of RNAPII at Lsmd1 and ncRNA00095 in THAP11 knockdown SW620 cells

Figure 3.11: Chromatin immunoprecipitation (ChIP) analysis of RNAPII at Lsmd1 and β-actin in THAP11 knockdown and THAP11 rescue SW620 cells

Figure 3.12: Chromatin immunoprecipitation (ChIP) analysis of Acetyl-K9/K18 histone H3 at Lsmd1 and ncRNA00095 in THAP11 knockdown SW620 cells

Figure 3.13: THAP11 interacts with HCF-1 in colon cancer cells

Figure 3.14: THAP11 and HCF-1 co-occupancy on chromatin

Figure 3.15: Knockdown of THAP11 decreases HCF-1 occupancy on THAP11 target genes Lsmd1 and ncRNA00095

Figure 3.16: Knockdown of HCF-1 de-represses THAP11 target genes

Figure 3.17: Knockdown of HCF-1 increases RNA Polymerase II occupancy at Lsmd1

Figure 3.18: Affinity purification of THAP11/HCF-1/OGT/HDAC-1 complex from SW620 cells

Figure 3.19: OGT occupancy at THAP11 repressed gene promoters

Figure 3.20: Knockdown of THAP11 restores Annexin A1 expression in SW620 cells

Figure 3.21: Annexin A1 and THAP11 expression inversely correlate in colon cancer cell lines

Figure 3.22: Knockdown of THAP11 decreases cell proliferation

Figure 3.23: Annexin A1 de-repression is dispensable for decreased cell proliferation in THAP11 knockdown cells

Figure 4.1: Proposed model of THAP11 mediated transcriptional repression

Figure 4.2: Creation of GFP and luciferase expressing SW620 THAP11 knockdown cells
Abbreviations

bp (base pair);
BRE (TFIIB recognition element);
BrU (5-bromouridine);
BrUTP (5-bromouridine 5'-triphosphate);
BSA (Bovine serum albumin);
CHD (Chromodomain, helicase, DNA binding);
ChIP (Chromatin immunoprecipitation);
CTD (C-terminal domain);
DAB (3,3'-Diaminobenzidine);
DAPI (4',6-Diamidino-2-phenylindole);
DBD (DNA binding domain);
DCE (Downstream core element);
DIC (Differential interference contrast);
DMSO (Dimethyl sulfoxide);
DmTHAP (Drosophila THAP);
DPE (Downstream promoter element);
DSIF (DRB sensitivity inducing factor);
DTT (dithiothreitol);
EDTA (Ethylenediaminetetraacetic acid);
EGTA (Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid);
GlcNAc (N-Acetyl-D-glucosamine);
GST (Glutathione-S-transferase);
HAT (Histone acetyltransferase);
HCF-1 (Host cell factor-1);
HDAC (Histone deacetylase);
INHAT (Inhibitor of acetyltransferases);
INO80 (Inositol requiring 80);
Inr (Initiator);
ISWI (Imitation switch);
MTE (Motif-ten element);
NELF (Negative elongation factor);
NMR (Nuclear magnetic resonance);
NuRD (Nucleosomes remodeling and deacetylase);
OGT (O-linked N-Acetyl-D-glucosamine transferase);
PBS (Phosphate buffered saline);
PML (promyelocytic leukemia);
pSRNG (pSuper.Retro.Neo+GFP);
pSRP (pSuper.Retro.Puro);
P-TEFb (Positive transcription elongation factor b);
qRT-PCR (Quantitative reverse transcription polymerase chain reaction);
RIPA (Radio-Immunoprecipitation assay);
RNAi (RNA interference);
RNAPI (RNA polymerase I);
RNAPII (RNA polymerase II);
RNAPIII (RNA polymerase III);
SDS-PAGE (Sodium dodecyl sulfate polyacrylamide gel electrophoresis);
shRNA (Short-hairpin RNA);
SWI/SNF (Switching defective/sucrose nonfermenting);
TAFs (TFIID associated factors);
THABS (THAP1 binding sequence);
THAP (Thanatos associated protein);
TMA (Tissue microarray);
TSS (Transcription start site);
WGA (Wheat germ agglutinin);
XCPE (X-core promoter element);
Chapter 1: Regulation of Transcription in Eukaryotes

Overview

Gene expression in eukaryotes is controlled by the dynamic interplay between regulatory mechanisms governing transcriptional activation and repression. Dysregulation of either transcriptional activation or repression can result in aberrant spatiotemporal gene expression and corresponding defects in cellular and organismal physiology. Numerous genetic studies utilizing gain-of-function or loss-of-function mutations in various transcriptional regulators have revealed alterations in cell growth, differentiation, and apoptosis, thereby emphasizing the importance of proper transcriptional regulation in biological processes. In addition, elaborate and complex gene expression patterns and the requisite increase in transcriptional control mechanisms have been suggested to account for the increased organismal complexity observed in metazoans versus lower eukaryotes (Levine and Tjian 2003). Consistent with this idea is the correlation between absolute number and ratio of transcription factors per genome and apparent organismal complexity; yeast, fly and human genomes have been estimated to encode ~300, 1000, and 3000 transcription factors, respectively (Levine and Tjian 2003). Thus, the identification and characterization of novel transcriptional regulators and their mechanisms of action is a necessary component of cellular and molecular biology.

Eukaryotes utilize several distinct DNA-dependent RNA polymerases to transcribe their full complement of cellular RNAs. RNA polymerase I (RNAPI)
transcribes the 45S precursor ribosomal RNA (rRNA) which is processed into mature 18S, 5.8S and 28S rRNAs (Russell and Zomerdijk 2005). RNA polymerase III (RNAPIII) also transcribes untranslated RNAs including 5S rRNA, tRNAs, U6 RNAs, and some microRNAs (Dieci, Fiorino et al. 2007). In contrast, RNA polymerase II (RNAPII) transcribes all protein coding mRNAs as well as long non-coding RNAs, and microRNAs (Lee and Young 2000; Lee, Kim et al. 2004; Guttman, Amit et al. 2009). Dysregulation of transcription catalyzed by each RNA polymerase has been implicated in the pathogenesis of human diseases, including cancer (White 2005; Marshall and White 2008). However, given the breadth of DNA templates transcribed by RNAPII and the relevance of these gene products in establishing normal and disease processes, mechanisms governing RNAPII dependent transcription have been widely studied and will be briefly summarized here.

**Mechanisms of Regulation**

**Transcription Initiation and Elongation**

Transcription by RNAPII initiates at core promoters following recruitment of the basal transcription machinery, operationally defined as the minimum set of proteins necessary to drive in vitro transcription from an isolated core promoter (Smale and Kadonaga 2003). Core promoters can be classified into two types, focused or dispersed, which reflects their relative number and location of transcription start sites (Juven-Gershon, Hsu et al. 2008). Focused promoters contain either a single transcription start site or a distinct cluster of start sites.
over several nucleotides while dispersed promoters characteristically contain several start sites spanning 50-100 nucleotides. Focused transcription initiation occurs throughout all eukaryotes and is believed to be the primary mechanism of transcription initiation in lower eukaryotes (Juven-Gershon and Kadonaga 2010). Conversely, dispersed promoters are typically restricted to vertebrates, owing to their frequent location in CpG islands which are present in approximately 70% of vertebrate promoters (Juven-Gershon and Kadonaga 2010).

Despite this discrepancy the vast majority of studies to date have focused on mechanisms regulating transcription initiation from focused core promoters. Eight core promoter elements, characterized by DNA sequence, have been identified at focused promoters of RNAPII transcribed genes and include: TATA, initiator (Inr), downstream promoter element (DPE), upstream and downstream TFIIB recognition elements (BRE<sup>u</sup> and BRE<sup>d</sup>), motif ten element (MTE), X core promoter element 1 (XCPE1) and downstream core element (DCE) (Smale and Kadonaga 2003; Juven-Gershon and Kadonaga 2010). None of these core promoter elements is universal and each element is found in only a fraction of core promoters.

Core promoter elements, with the exception of the appropriately named TFIIB recognition elements, are recognized by distinct subunits of general transcription factor TFIID, a multi-subunit complex comprised of TATA-binding protein (TBP) and approximately 13 or 14 TFIID associated factors (TAFs) (Smale and Kadonaga 2003). TFIID binding at core promoters is stabilized by TFIIA, followed by recruitment of TFIIB which associates through both DNA
recognition at BREs and direct contact with TFIID. Binding of TFIID and TFIIB results in additional general transcription factor (TFIIE, TFIIF, TFIIH) and RNAPII binding, either sequentially or pre-associated as an RNAPII holoenzyme, to form the transcription preinitiation complex (PIC) (Thomas and Chiang 2006). Because preinitiation complexes are competent to initiate transcription, cells utilize numerous control mechanisms to regulate their assembly and activity.

Core promoters delineate the site of transcription initiation but additional regulatory cues provided by cis-acting DNA sequences are necessary to achieve the range of dynamic gene expression observed in vivo. Sequence specific DNA binding factors present at cis-acting DNA sequences, including proximal promoter elements and distal enhancers, can positively regulate transcription initiation by stimulating the assembly and activity of the basal transcription machinery at the core promoter (Lee and Young 2000). Numerous sequence specific transcription factors have been shown to facilitate PIC assembly through direct interaction with TFIID and TFIIB (Deng and Roberts 2007; Cler, Papai et al. 2009). The glutamine-rich transactivation domains of both Sp1 and CREB directly interact with the TFIID subunit TAF4 and mutation of the TAF4 residues responsible for Sp1 binding significantly diminish Sp1 mediated transcriptional activation (Gill, Pascal et al. 1994; Rojo-Niersbach, Furukawa et al. 1999; Asahara, Santoso et al. 2001). Electron microscopy and biochemical studies using purified native TFIID in complex with either p53, c-Jun or Sp1 has revealed that these activators bind distinct surfaces on TFIID, including specific TAFs (Liu, Coleman et al. 2009). This finding suggests that certain combinations of
promoter bound transcription factors may synergistically activate transcription by contacting unique, non-overlapping surfaces on TFIID (Liu, Coleman et al. 2009). Activator dependent recruitment of TFIID is thought to be especially important at dispersed, CpG island promoters which typically lack consensus TATA, Inr, and DPE elements (Smale and Kadonaga 2003). Rather, CpG islands frequently contain multiple Sp1 binding GC-boxes and transcription start sites are often located 40-80bp downstream of these sites, suggesting Sp1 may stimulate transcription by recruiting basal transcription factors at these promoters (Smale and Kadonaga 2003).

Recent genome-wide studies have identified basal transcription factor and RNAPII occupancy at transcriptionally inactive genes, indicating an essential role for post-recruitment regulatory mechanisms in controlling gene expression (Muse, Gilchrist et al. 2007; Zeitlinger, Stark et al. 2007; Core, Waterfall et al. 2008). Multiple events occur following preinitiation complex assembly and prior to productive transcript elongation, including transcription bubble formation, promoter escape, and promoter-proximal pausing (Fuda, Ardehali et al. 2009). Each step may potentially be regulated but numerous studies have identified promoter-proximal pausing as a key mechanism in regulated gene expression. Initially identified at Drosophila heat shock genes, promoter-proximal pausing is characterized by the stalling of RNAPII following transcription of 20-40bp of nascent mRNA (Margaritis and Holstege 2008). Promoter-proximal pausing is governed by the association of pausing factors such as DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) with RNAPII (Margaritis and
holstege 2008). the conversion from promoter-proximal pausing to productive elongation is regulated by positive transcription elongation factor b (p-tefb); a heterodimer composed of cyclin dependent kinase 9 (cdk9) and one of the c-type cyclins t1, t2a, t2b or k (kohoutek 2009). recruitment of p-tefb to stalled rnapii results in phosphorylation of both the pausing factors as well as serine 2 of the ysptspS heptapeptide repeat present in the c-terminal domain (ctd) of rbp1 subunit of rnapii. these phosphorylations relieve pause factor induced stalling and promote the transition to productive elongation.

p-tefb activity is negatively regulated by its reversible association with the 7sk small nuclear ribonucleoprotein complex (7sk snrnp) composed of 7sk snrna and RNA binding proteins HEXIM1/2, LARP7, and BCDIN3 (Peterlin and Price 2006; Barboric, Lenasi et al. 2009). a substantial fraction of active p-tefb associates with the dual bromodomain-containing protein Brd4 suggesting a mechanistic link between histone acetylation and transcriptional elongation (Jang, Mochizuki et al. 2005; Yang, Yik et al. 2005; Chiang 2009). in addition, sequence specific transcription factors including c-myc, NF-κB and the HIV-1 transactivator Tat, also recruit P-TEFb to their respective target gene promoters indicating some activators function to regulate transcription at the level of elongation as well as initiation (Barboric, Nissen et al. 2001; Luecke and Yamamoto 2005; Barboric, Yik et al. 2007; Rahl, Lin et al. 2010).

sequence specific DNA binding factors also function as platforms for the assembly of non-DNA binding transcriptional co-regulators. these recruited co-regulators, frequently in the form of macromolecular protein complexes, modulate
transcription by altering chromatin structure or stimulating the recruitment and activity of the basal transcription machinery. These mechanisms, either alone or in combination, are applicable to a variety transcription factors. For example, activation of wingless (Wg/Wnt) target genes in *Drosophila* requires the metazoan-specific ETO-TAFH domain present near the amino-terminus of TAF4. The ETO-TAFH domain physically interacts with Pygopus, a critical component of the tri-partite Armadillo/Legless/Pygopus bridge, thereby enabling DNA bound T-cell factor (TCF) to interact with TFIID (Wright and Tjian 2009). Interestingly, TFIID and RNAPII were found to be pre-loaded at Wg/Wnt target genes prior to activation suggesting the Pygopus/TAF4 interaction may function to stimulate transcription elongation by relieving promoter proximal stalling (Wright and Tjian 2009). The integration of chromatin modifying activities with transcriptional output is best exemplified by nuclear hormone receptors which positively and negatively regulate transcription through reversible association with co-activator and co-repressor complexes in a hormone dependent manner. For example, liganded thyroid hormone receptor exchanges histone deacetylase (HDAC) containing co-repressor complexes with histone acetyltransferase containing (HAT) co-activator complexes, thereby altering chromatin structure through post-translational modification of amino-terminal histone tails (Sharma and Fondell 2002). Following HAT recruitment, liganded receptor associates with the Mediator complex which in turn facilitates basal transcription factor and RNAPII recruitment and activation (Sharma and Fondell 2002; Belakavadi and Fondell 2010).
Chromatin and Transcriptional Regulation

DNA dependent processes, including transcription, occur not on naked DNA but in the context of chromatin; a highly-ordered, condensed, polymeric structure necessary to fit eukaryotic DNA into the physical constraints of the nucleus. The basic repeating unit of chromatin is the nucleosome; approximately 146bp of DNA wrapped 1.65 turns around an octamer of core histones (Lee and Young 2000). The histone octamer is composed of a two heterodimers of histones H3 and H4 which are flanked by two heterodimers of H2A and H2B. The solved crystal structure of the nucleosome revealed a roughly cylindrical histone octamer particle with tightly wrapped DNA making multiple contacts with core histones along its phosphate backbone (Luger, Mader et al. 1997). Individual nucleosomes are separated by ~10-60bp of linker DNA resulting in a nucleosomal array approximately 10nm in diameter with a topology frequently described as ‘beads on a string’. Binding of histone H1 to linker DNA stimulates further compaction of the nucleosomal array into a fiber of ~30nm in diameter. This chromatin fiber can further compact into 100-400nm interphase fibers or the highly condensed chromosomes observed in metaphase cells (Woodcock and Ghosh 2010).

The assembly of DNA into nucleosomes and higher order chromatin structures is generally regarded as inhibitory towards transcription. The highly condensed structure of heterochromatin suppresses transcription at these repetitive elements while the locally compacted chromatin at Hox loci ensures their expression in a developmentally regulated manner (Grewal and Jia 2007;
Eskeland, Leeb et al. 2010). However, even the fully decondensed nucleosomal arrays found at transcriptionally permissible chromatin may occlude regulatory DNA sequences and inhibit transcription by preventing the association of sequence specific or general transcription factors with their cognate binding sites. Therefore, dynamic alteration of chromatin structure and function is an essential mechanism governing regulated gene expression. The two primary chromatin modifying activities implicated in transcriptional regulation are post-translational modification of histones and ATP-dependent chromatin remodeling.

**Histone Modification**

Core histones provide the protein scaffold for nucleosome formation. The globular domains of core histones mediate histone-histone and histone-DNA interactions while the flexible amino-terminal tails extend away from the histone octamer where they can be reversibly modified by a variety of post-translational modifications including: acetylation, methylation, phosphorylation, ubiquitination, sumoylation and ADP-ribosylation. Lysine residues in histone tails are frequent acceptor sites for these modifications and have been shown to be acetylated, ubiquitinated, sumoylated, ADP-ribosylated and mono-, di-, and tri-methylated (Turner 2005; Messner, Altmeyer et al. 2010). Arginine residues can be mono- or di-methylated; serine and threonine residues can be phosphorylated (Turner 2005).

Genome-wide studies in various model systems have revealed that specific histone modification patterns correlate with transcriptional activity and demarcate functional regions of chromatin (Schneider, Bannister et al. 2004;
Bernstein, Kamal et al. 2005; Pokholok, Harbison et al. 2005; Kolasinska-Zwierz, Down et al. 2009; Karlic, Chung et al. 2010). Hyperacetylation of histones H3 and H4 in conjunction with tri-methylation at H3 lysine 4 (H3K4me3) are observed at promoters of actively transcribed genes. Tri-methylation at H3K36 also denotes actively transcribed genes but this modification is enriched in the body of genes, preferentially at exons, suggesting a role for this modification in regulating co-transcriptional splicing (Kolasinska-Zwierz, Down et al. 2009; Fox-Walsh and Fu 2010; Luco, Pan et al. 2010). The robustness of these histone modifications as hallmarks of actively transcribed chromatin was recently used to identify numerous long intergenic non-coding RNAs (lincRNA) in genomic regions previously considered to be ‘gene deserts’ (Guttman, Amit et al. 2009). Other histone modifications such as methylation of H3K9 and H3K27 are enriched at transcriptionally silent regions including pericentromeric heterochromatin (H3K9me2/3), telomeres (H3K9me2/3, H3K27me2/3) and developmentally regulated genes (H3K9me2/3, H3K27me2/3). In addition, acquisition of these repressive histone modifications is believed to play a critical role in epigenetic silencing of tumor and metastasis suppressor genes (Kondo, Shen et al. 2003; Herranz, Pasini et al. 2008).

One mechanism histone tails and their associated modifications regulate transcription is through modulation of higher order chromatin structure (Kouzarides 2007). Of particular importance is a short stretch of basic amino acids in the H4 tail which contacts a patch of acidic residues formed at the interface of H2A/H2B in an adjacent nucleosome. This internucleosomal
association is believed to contribute to 30nm fiber formation as well as higher order fiber-fiber interactions (Dorigo, Schalch et al. 2003; Dorigo, Schalch et al. 2004; Sinha and Shogren-Knaak 2010). Indeed, deletion of H4 tail residues 14-19, or acetylation of lysine 16 is sufficient to disrupt salt-dependent 30nm chromatin fiber formation in vitro (Dorigo, Schalch et al. 2003; Shogren-Knaak, Ishii et al. 2006). These findings suggest that H4K16 acetylation likely plays an important role in maintaining an open chromatin environment amenable to transcription.

In addition to regulating chromatin structure, a preponderance of evidence suggests that histone tail modifications also act as recognition motifs for a variety of effector proteins which in turn provide unique functionalities that modulate key rate limiting steps of the transcription cycle. These effectors employ specialized domains to recognize histone tails in a modification sensitive manner. The ability to discriminate between specific modifications in a context sensitive manner allows these effectors to participate in disparate transcriptional outcomes. For example, bromodomains recognize acetylated lysines and bromodomain-containing effectors frequently link histone acetylation with transcriptional activation. The aforementioned dual bromodomain-containing Brd4 binds to primary response genes following signal dependent histone acetylation and recruits P-TEFb to stimulate transcription elongation and co-transcriptional mRNA processing (Hargreaves, Horng et al. 2009; Zippo, Serafini et al. 2009). Bromodomain-containing effectors also recognize non-histone substrates. For instance, the bromodomains of TAF1 bind acetylated p53 facilitating TFIID
recruitment to DNA damage inducible promoters (Li, Piluso et al. 2007).

Methylated histone tails are recognized by chromodomain-containing proteins such as Cbx5, Cbx1, and Cbx3 (formerly HP-1α, HP1-β, HP-1γ) which bind di and tri-methylated H3K9 (H3K9me2/3). Binding of Cbx5 and Cbx1 at H3K9me2/3 represses transcription at retinoblastoma regulated genes and promotes chromatin fiber compaction characteristic of heterochromatin (Nielsen, Schneider et al. 2001; Fan, Rangasamy et al. 2004). Likewise, mammalian Polycomb homologues (Cbx2, Cbx4, Cbx6, Cbx7, Cbx8) utilize chromodomains to tether the Polycomb repressive complex PRC1 to chromatin containing H3K27me2/3 resulting in chromatin compaction and gene silencing (Simon and Kingston 2009).

In addition to amino-terminal tail modifications, recent mass spectrometry based experiments have identified over 30 modifications in the globular domain of core histones (Mersfelder and Parthun 2006). Comparison of these modifications with the nucleosome crystal structure reveals these modifications are concentrated at the solvent accessible histone surface, histone-histone interfaces, and the histone lateral surface including residues that directly contact DNA (Mersfelder and Parthun 2006). Globular domain modifications likely regulate transcription and other DNA dependent processes by altering chromatin structure as well as regulating effector protein recruitment.

The reversible nature of post-translational histone modifications ensures that dynamic recruitment of histone modifying enzymes is a primary mechanism of transcriptional regulation. Indeed, histone modifying activities found in co-
regulator complexes are frequently indispensable to co-regulator function. The yeast transcriptional co-activator Gcn5 contains intrinsic histone acetyltransferase (HAT) activity and ablation of HAT function disrupts Gcn5 mediated transcriptional activation in vivo (Kuo, Zhou et al. 1998). In an analogous manner, histone deacetylase (HDAC) activity is critical for co-repressor complex function. Small molecule HDAC inhibitors alleviate transcriptional repression concomitant with histone hyperacetylation and RNAi-mediated depletion of HDAC3 disrupts repression by unliganded nuclear hormone receptors (Ishizuka and Lazar 2003). Similar findings are observed with respect to histone methylation. The histone demethylase Lsd1 functions in multiple co-repressor complexes that repress transcription by removing H3K4 methylation (Wysocka, Milne et al. 2005). In contrast, H3K4 methyltransferases associated with DNA binding or co-activator complexes stimulate transcription (Ruthenburg, Allis et al. 2007). Synergism between demethylases and methyltransferases has also been reported to influence transcriptional output. For example, complex formation between H3K27 demethylases and H3K4 methyltransferases allows the removal of repressive histone methylation with simultaneous deposition of activating methyl marks (Swigut and Wysocka 2007).

**Chromatin Remodeling**

Nucleosome dynamics are controlled by chromatin remodeling complexes which utilize energy derived from ATP hydrolysis to slide, evict or restructure nucleosomes. Nucleosome sliding and/or eviction exposes the underlying DNA template to sequence specific or general transcription factors while nucleosome
restructuring allows incorporation of histone variants (H3.3) associated with active transcription. Four distinct families of chromatin remodelers have been characterized: SWI/SNF (switching defective/sucrose nonfermenting), ISWI (imitation switch), CHD (chromodomain, helicase, DNA binding), and INO80 (inositol requiring 80) (Clapier and Cairns 2009). Each remodeler family contains a catalytic subunit harboring a conserved ATPase domain and at least one unique flanking domain which defines family membership. These unique domains in conjunction with non-catalytic accessory subunits impart individual family members with specialized functions in distinct chromatin dependent processes, including transcription (Clapier and Cairns 2009). For example, the SWI/SNF remodeler Brg1 has been implicated in glucocorticoid receptor mediated transcriptional activation concomitant with chromatin remodeling (Trotter, Fan et al. 2008). Similar findings for SWI/SNF family member Brm have been observed in Drosophila where genetic studies revealed a global decrease in RNAPII binding and transcription following attenuation of Brm function (Armstrong, Papoulas et al. 2002). Other remodeling complexes repress transcription as exemplified by the Mi-2/NuRD (nucleosomes remodeling and deacetylase) complex which couples chromatin remodeling and histone deacetylase activities in the same complex (Denslow and Wade 2007).

Gene transcription is controlled by the combinatorial action of multiple regulatory factors and mechanisms (Figure 1.1). DNA binding factors and their associated co-regulators interact with components of the transcription apparatus including general transcription factors and RNAPII. Co-regulator complexes also
contain histone modifying enzymes and nucleosome remodelers which alter the chromatin template to either activate or repress transcription. The importance of transcriptional regulatory mechanisms in maintaining normal cellular processes is underscored by the observation that approximately 10 percent of metazoan protein coding genes encode putative transcription factors. However, the transcriptional regulatory mechanisms and endogenous gene targets associated with most of these factors remain incompletely understood.
General transcription factors (GTFs) bind to specific sequence elements in the core promoter. The relative position of these core promoter elements with respect to the transcription start site (black arrow) are indicated. Sequence specific transcriptional regulators (orange oval and yellow diamond), including activators and repressors, bind to DNA sequences located near the core promoter (proximal sites) and/or at distant enhancer regions. These sequence specific regulators can interact (green arrows) with GTFs, including TFIID (blue rectangle), TATA-binding protein (TBP, blue horseshoe), and RNAPII (red 'rocket') to activate or repress transcription. They also interact (green arrows) with co-regulators (green hexagon) that can interact (blue arrows) with the general transcription machinery or chromatin-modifying factors, such as histone modifying enzymes and nucleosome remodelers. Transcriptional activators can recruit, stabilize or stimulate these factors while repressors can disrupt or inhibit these factors. Adapted from (Fuda, Ardehali et al. 2009).
Chapter 2: Characterization of THAP10 and THAP11 as Transcriptional Repressors Expressed in Human DNA Damage and Colon Cancer Progression

Summary

The Thanatos associated protein, or THAP, domain is an evolutionarily conserved zinc-finger motif restricted to metazoans. THAP domains show significant homology to the sequence-specific DNA binding domain of Drosophila P element transposase and have been suggested to function in DNA and chromatin dependent processes. However, the transcriptional regulatory capacity of most THAP proteins and their function in normal or aberrant physiological processes are largely unknown. In this chapter, we identify human THAP10 and THAP11 as differentially expressed during DNA damage and colon cancer progression, respectively. THAP10 mRNA was found to be induced by genotoxic chemotherapeutics in HeLa and HepG2 but not p53-null Saos-2 cells. Ecotopic overexpression of p53 was sufficient to markedly induce THAP10 mRNA suggesting THAP10 may be a novel p53-regulated, DNA damage inducible gene. THAP11 mRNA and protein was found to be upregulated in metastatic SW620 colon cancer cells relative to their isogenic, primary tumor derived SW480 counterparts. Immunohistochemical analysis of human colon cancers revealed increased THAP11 staining frequency and intensity correlated with disease progression. Both THAP10 and THAP11 were found to potently repress basal and activated transcription when targeted to promoters as
heterologous Gal4-DNA binding domain fusion proteins. Consistent with
transcriptional repression, both THAP10 and THAP11 were shown to associate
with histone deacetylases in vivo and in vitro. Taken together, these results
suggest that THAP10 and THAP11 dependent transcriptional repression may
contribute to the gene expression programs associated with DNA damage and
colon cancer progression, respectively.

Introduction

Introduction to the THAP Domain

The Thanatos associated protein (THAP) domain is a recently identified,
evolutionarily conserved protein motif containing a high degree of similarity to the
sequence-specific DNA binding domain of *Drosophila* P element transposase
(Roussigne, Kossida et al. 2003). THAP domains are approximately 90 amino
acids in length and characterized by an N-terminally located C2-CH (Cys-Xaa$_{2-4}$-
Cys-Xaa$_{35-50}$-Cys-Xaa$_{2}$-His) zinc-finger signature, several invariant residues
including Pro26, Trp36, Phe58, and Pro78 (numbering with respect to THAP1)
and a C-terminal AVPTIF box (Figure 2.1) (Roussigne, Kossida et al. 2003).

Genome sequencing and bioinformatic analyses have identified over 300
THAP proteins making this domain the second most prevalent DNA binding, zinc-
coordinating motif after C2-H2 containing zinc-fingers (Sabogal, Lyubimov et al.
2010). However, unlike C2-H2 zinc-finger proteins, THAP proteins are restricted
exclusively to animals; none have been identified in the genomes of plants,
yeast, fungi, or bacteria. Therefore, it appears that THAP proteins have evolved to serve specific functions relevant to animal physiology.

Orthologs of human THAP genes have been found in genomes of vertebrates including mouse, rat, dog, cow, chicken, pig, frog and fish (Table 2.1) (Clouaire, Roussigne et al. 2005). However, human orthologs have not been identified in genomes of flies or worms. Twelve distinct THAP domain containing proteins have been identified in humans, each of which contains a single N-terminally located THAP domain (Figure 2.1A) (Roussigne, Kossida et al. 2003). Only a subset of these (THAP0, 1, 2, 4, 7, and 11) has orthologs in rodents. Additionally, only THAP1, 4, 7, and 11 are also conserved in *X. laevis* while THAP0, 1, 7, 9, and 11 are conserved in *D. rerio*.

**Structure-Function Analysis of THAP domains**

Much of our current understanding regarding THAP domains as DNA binding modules comes from studies characterizing THAP domain function in *Drosophila P element* transposase and human THAP1. *Drosophila P element* DNA encodes the mobile DNA element as well as transposase and transpositional repressor proteins (Kaufman, Doll et al. 1989). The 207 amino acid KP repressor protein, arising from a naturally occurring internal deletion in P element DNA, is ostensibly a truncated P element transposase; both contain identical N-terminal 199 amino acids including THAP domain (Lee, Mul et al. 1996). While investigating the molecular mechanisms by which repressor proteins inhibit transposition, Lee et al. discovered that KP repressor contains site specific DNA binding activity indistinguishable from the transposase itself.
(Lee, Beall et al. 1998). Point mutation of the first two zinc coordinating cysteines abolished in vitro binding to P element DNA sequences as determined by DNase I footprinting, thus providing important initial evidence that THAP domains could function as modular, site specific DNA binding zinc fingers (Lee, Beall et al. 1998). More recently, Clouaire and co-workers have used an in vitro selection approach to determine an 11-nucleotide high affinity DNA binding sequence for human THAP1 termed THABS (THAP1 Binding Sequence) (Clouaire, Roussigne et al. 2005). In vitro THAP1/THABS interaction studies revealed that, like the Drosophila KP repressor, THAP1 sequence specific DNA binding was also strictly zinc dependent. Mutation of the conserved zinc coordinating cysteine and histidine residues completely abolished DNA binding as did metal chelation performed either with EDTA or the zinc specific chelator 1,10-o-phenanthroline (Clouaire, Roussigne et al. 2005). Further underscoring the role of zinc in proper THAP domain function, add-back experiments performed in the presence of chelator demonstrated that supplemental zinc but not iron, calcium or magnesium was capable of rescuing chelator mediated inhibition of DNA binding (Clouaire, Roussigne et al. 2005). In addition to the conserved zinc coordinating residues, THAP1, like all THAP domains, also contains several invariant residues (Pro26, Trp36, Phe58, and Pro78). Importantly, site directed mutagenesis of these residues also abrogated the THAP1/THABS sequence specific DNA association (Clouaire, Roussigne et al. 2005). Taken together, these experiments strongly suggested that THAP domains possess zinc dependent, sequence specific DNA binding activity.
Furthermore, the defining amino acids of THAP zinc fingers were also demonstrated to be indispensable for their function as DNA binding domains. THAP domains, like other zinc fingers, display relatively low primary sequence identity. However, recent structural analyses have revealed that diverse THAP domains adopt similar secondary and tertiary structures. Nuclear magnetic resonance determined solution structures of THAP domains from human THAP1 and C. elegans C-terminal binding protein (CtBP), in addition to the x-ray crystal structure of Drosophila P element transposase, have each independently revealed a similar three-dimensional structure characterized as a \(\beta\)-\(\alpha\)-\(\beta\) fold nucleated by tetrahedral coordination of a single zinc ion (Liew, Crossley et al. 2007; Bessiere, Lacroix et al. 2008; Sabogal, Lyubimov et al. 2010). In addition to the \(\beta\)-\(\alpha\)-\(\beta\) fold, the remainder of THAP domains is comprised primarily of loops, with loop 4 the most variable in length and sequence (Sabogal, Lyubimov et al. 2010).

The first two zinc coordinating cysteines are positioned in a loop structure preceding the initial \(\beta\)-strand while the remaining zinc binding cysteine and histidine reside immediately C-terminal to the second \(\beta\)-strand. This flanking of the \(\beta\)-\(\alpha\)-\(\beta\) motifs with the zinc binding amino acids necessarily results in the formation of a two-stranded anti-parallel \(\beta\)-sheet upon zinc coordination (Liew, Crossley et al. 2007; Bessiere, Lacroix et al. 2008; Sabogal, Lyubimov et al. 2010).

The functional significance of this \(\beta\)-sheet with respect to sequence specific DNA binding was recently demonstrated by Sabogal et al. upon solving
the x-ray crystal structure of *Drosophila* P element transposase THAP domain (DmTHAP) bound to its cognate DNA element. In DmTHAP, the β-sheet in conjunction with the N-terminal methionine was shown to contribute several direct and water mediated nucleotide specific contacts to both strands of major groove DNA (Sabogal, Lyubimov et al. 2010). Furthermore, site directed mutagenesis of DNA binding β-strand residues His18 and Gln42 significantly weakened the binding affinity of DmTHAP for its DNA ligand as determined by electrophoretic mobility shift assay (Sabogal, Lyubimov et al. 2010). The contribution of His18 towards sequence specific DNA binding also provides an explanation for the previous observation that KP repressor protein harboring an H18A mutation lost site specific but retained high affinity, nonspecific DNA binding activity (Lee, Beall et al. 1998). Mutation of Lys24 in human THAP1, which occupies a β-strand position synonymous with His18 in DmTHAP, similarly disrupts in vitro DNA binding suggesting that Lys24 may contribute to sequence specific THAP1 β-strand/DNA contact (Bessiere, Lacroix et al. 2008). However, because the NMR solution structure of human THAP1 was performed in the absence of DNA ligand, this possibility remains to be tested (Bessiere, Lacroix et al. 2008).

Comparison of human THAP sequences reveals a high degree of amino acid variability in the predicted THAP β-sheets suggesting diverse THAP proteins likely recognize different DNA major groove sequences (Sabogal, Lyubimov et al. 2010). Indeed, the THAP zinc fingers of human THAP2 and THAP3 were shown to be incapable of binding the THAP1 specific THABS DNA sequence in
electrophoretic mobility shift assays despite sharing up to 50% sequence identity throughout their respective THAP domains (Bessiere, Lacroix et al. 2008). The relative contribution of the individual divergent β-sheet residues to this apparent sequence specificity remains to be determined.

In addition to the sequence specificity imparted by β-sheet contacts with the DNA major groove, Sabogal and co-workers also demonstrated that DmTHAP additionally interacts with minor groove DNA. Sequence specific minor groove DNA contacts are mediated through loop 4 residues Arg65 and Arg67 while Arg66 provides nonspecific DNA interaction by contacting the phosphate backbone of minor groove DNA (Sabogal, Lyubimov et al. 2010). Interestingly, the loop 4 regions in THAP domains are the most variable in terms of composition and length but each contains at least one basic residue (Liew, Crossley et al. 2007; Sabogal, Lyubimov et al. 2010). Molecular modeling studies have indicated that even the dramatically shortened loop 4 regions of human THAP11 and C. elegans CtBP likely make contact with minor groove DNA (Sabogal, Lyubimov et al. 2010). However, it has been suggested that the potential for diminished loop 4/minor groove contacts in THAP11 may necessitate a modified DNA binding mechanism, involving perhaps homo- and hetero-dimerization (Sabogal, Lyubimov et al. 2010). In support of this hypothesis, mouse THAP11 (termed Ronin) has recently been shown to associate both with itself and THAP7; however the relevance of these associations in the context of sequence specific DNA binding has not yet been reported (Dejosez, Krumenacker et al. 2008).
In addition to the defining C2-CH zinc finger, all THAP domains also contain a characteristic C-terminal AVPTIF motif (amino acids) and four highly conserved residues Pro26, Trp36, Phe58, and Pro78 (numbering with respect to THAP1) each essential for in vitro DNA binding (Figure 2.1B) (Roussigne, Cayrol et al. 2003; Clouaire, Roussigne et al. 2005). The aforementioned structural studies have independently determined that the four invariant residues form a hydrophobic core, anchored by the unique tryptophan (Trp36) positioned within the conserved α-helix. In THAP1, Trp36 makes hydrophobic contacts with several surrounding residues including the conserved Phe58, Pro26 and Pro78 which lies within the AVPTIF motif (Bessiere, Lacroix et al. 2008). These contacts contribute to the overall THAP protein fold by bringing the C-terminal AVPTIF motif in contact with the conserved α-helix (Bessiere, Lacroix et al. 2008). Accordingly, perturbation of the invariant residues or the AVPTIF box likely inhibits DNA binding by disrupting overall THAP domain structure.

Collectively, these data demonstrate that the β-sheet and flexible loop 4 structure of THAP domains target DNA via a bipartite recognition of adjacent major and minor groove DNA (Sabogal, Lyubimov et al. 2010). The reliance on the two-stranded β-sheet as a sequence specific DNA binding structure is atypical amongst zinc-finger proteins (Sabogal, Lyubimov et al. 2010). Classical C2-H2 type zinc-fingers typically utilize α-helices to form nucleotide specific contacts. In contrast, the THAP domain α-helix appears oriented away from the DNA-protein interface where it functions in overall domain structure and may provide a scaffold for protein-protein interactions. While not directly assessing
the function of the conserved $\alpha$-helix, recent work in our laboratory demonstrated that the THAP domain of THAP7 was necessary and sufficient for in vitro association with HDAC3 demonstrating that THAP domains may also function as protein-protein interaction modules (Macfarlan, Kutney et al. 2005).

**THAP Proteins in Biological Processes**

THAP proteins have been implicated in a diverse array of physiological processes including cell proliferation, apoptosis, and maintenance of pluripotency (Roussigne, Cayrol et al. 2003; Cayrol, Lacroix et al. 2007; Dejosez, Krumenacker et al. 2008; Balakrishnan, Cilenti et al. 2009; Zhu, Li et al. 2009). The DNA binding properties of THAP domains naturally suggests that THAP proteins may regulate these cellular processes in a DNA and chromatin dependent manner but this has been directly demonstrated in only a few instances (Cayrol, Lacroix et al. 2007; Dejosez, Krumenacker et al. 2008; Zhu, Li et al. 2009).

THAP1, the first characterized human THAP protein, was initially identified as a pro-apoptotic nuclear protein that localized in PML nuclear bodies and physically associated with the pro-apoptotic Par-4 (prostate and apoptosis-4) protein (Roussigne, Cayrol et al. 2003). Roussigne et al. found that over-expression of THAP1 could potentiate both serum withdrawal and cytokine induced apoptosis. Importantly, this pro-apoptotic function was shown to be dependent on the presence of the THAP domain since over-expression of a THAP deleted construct failed to affect apoptosis by the aforementioned stimuli (Roussigne, Cayrol et al. 2003). Subsequent work from the same laboratory
revealed that THAP1, through modulation of pRb/E2F target genes, could also regulate the proliferation and cell cycle progression of vascular endothelial cells (Cayrol, Lacroix et al. 2007). Knockdown of THAP1 using RNA interference reduced expression of several pRB/E2F target genes including *RRM1* (ribonucleotide reductase 1); a gene previously known to be required for S-phase DNA synthesis. Subsequent chromatin immunoprecipitation and electrophoretic mobility shift assays demonstrated THAP1 bound to the *RRM1* promoter at a DNA sequence highly similar to the previously characterized THABS consensus sequence, thereby providing the first report of a directly regulated endogenous THAP target gene (Cayrol, Lacroix et al. 2007).

THAP1 has recently been identified as a genetic determinant for mixed-onset primary torsion dystonia (DYT6 dystonia), a neurological disorder characterized by involuntary twisting movements and abnormal postures (Fuchs, Gavarini et al. 2009). Various types of dystonia can be distinguished genetically with most inherited in an autosomal dominant manner with incomplete penetrance. To identify the *DYT6* gene, Fuchs et al. systematically sequenced genes in the DYT6 disease haplotype, a 23 cM region on chromosome 8 previously linked to disease progression in affected Amish-Mennonite families. This approach identified two independent mutations within the *THAP1* gene that co-segregated with affected individuals and obligate carriers. The first mutation identified is an insertion/deletion mutation resulting in a frameshift at amino acid 44 that further generates a premature stop codon at residue 73 (Fuchs, Gavarini et al. 2009). This frameshift mutation likely results in a non-functional protein.
owing to deletion of important DNA binding determinants within the THAP domain including zinc coordinating residues (Cys54, His57) and the AVPTIF box. The second isolated mutation encodes an F81L missense mutation within the conserved AVPTIF motif (Fuchs, Gavarini et al. 2009). Importantly, Fuchs et al. demonstrated that the F81L mutant possessed substantially reduced affinity for the THAP1 consensus DNA element as determined by electrophoretic mobility shift assay. These mutations suggest that DYT6 dystonia may manifest as a result of perturbation in a THAP1 dependent gene expression program. Identification and characterization of downstream THAP1 gene targets may provide both novel DYT6 therapeutic opportunities and important insights into the mechanism of THAP1 mediated transcriptional regulation.

While the mechanism of THAP1 dependent transcriptional regulation is incompletely understood, work in our laboratory has described a detailed molecular mechanism by which THAP7 functions as a transcriptional repressor (Macfarlan, Kutney et al. 2005; Macfarlan, Parker et al. 2006). THAP7 was isolated in a yeast two-hybrid screen while searching for interacting partners of TAF-1β (template activating factor-1β), a subunit of the INHAT (inhibitor of acetyltransferase) complex which represses activated transcription by precluding histone acetyltransferases access to their histone tail substrates (Seo, McNamara et al. 2001; Seo, Macfarlan et al. 2002; Macfarlan, Kutney et al. 2005). THAP7 was shown to associate with chromatin in intact cells through a THAP domain independent interaction with amino-terminal histone tails (Macfarlan, Kutney et al. 2005). The histone tail binding activity of THAP7 was
found to reside within its C-terminal 77 amino acids and, similarly to TAF-1β, sensitive to the post-translational modifications of the histone tails (Macfarlan, Kutney et al. 2005).

The preferential association of THAP7 with hypoacetylated histone H4 tails suggested that THAP7 may transduce and/or establish a histone code amenable to transcriptional repression (Macfarlan, Kutney et al. 2005). Indeed, when tethered to a transcriptionally active promoter as a Gal4-DNA binding domain fusion protein, THAP7 was found to potently repress transcription and concomitantly decrease histone acetylation at the targeted promoter (Macfarlan, Kutney et al. 2005; Macfarlan, Parker et al. 2006). THAP7 dependent transcriptional repression was found to be partially sensitive to HDAC (histone deacetylase) inhibition and THAP7 was subsequently shown to interact directly with HDAC3 and the nuclear receptor co-repressor NCoR (Macfarlan, Kutney et al. 2005). Domain mapping studies surprisingly revealed that the THAP domain of THAP7 was necessary and sufficient for HDAC3 association indicating that THAP domains can mediate both DNA and protein interactions (Macfarlan, Kutney et al. 2005). While this initial characterization of THAP7 did not identify an endogenous target gene it nevertheless provided an important first realization that THAP proteins could function as transcriptional regulators. Additionally, this work expanded the repertoire of THAP protein functionality to include protein-protein as well as protein-DNA interactions.

In addition to THAP1 and THAP7, recent work by our laboratory (presented herein) and others has demonstrated that THAP11 possesses
transcriptional repressor activity in several distinct biological contexts (Dejosez, Krumenacker et al. 2008; Zhu, Li et al. 2009). Dejosez and co-workers identified THAP11 (termed Ronin) in a yeast two-hybrid assay designed to identify novel caspase-3 targets which may play a role in maintaining embryonic stem cell pluripotency. Mice bearing homozygous deletion of THAP11/Ronin were found to be embryonically lethal. Further characterization of this phenotype revealed that the inner cell mass of THAP11 null blastocysts failed to proliferate when cultured in vitro while forced overexpression of THAP11 in embryonic stem cells prevented spontaneous differentiation upon culture in the absence of leukemia inhibitory factor (LIF), suggesting THAP11 functions in maintaining embryonic stem cell pluripotency and viability (Dejosez, Krumenacker et al. 2008). Dejosez et al. hypothesized that the role of THAP11 in maintaining embryonic stem cell pluripotency may involve a transcriptional regulatory function similar to that observed with THAP1 and THAP7. Gene expression profiling of embryonic stem cells transiently transfected with a THAP11 expression construct revealed a remarkable trend towards global transcriptional repression (Dejosez, Krumenacker et al. 2008). Similarly, inducible expression of THAP11 under control of a doxycycline regulated promoter resulted in a sharp decrease in global nascent mRNA synthesis and a corresponding increase in the levels of histone H3 dimethyl-lysine 9 (H3K9me2), a histone modification correlated with transcriptional repression (Dejosez, Krumenacker et al. 2008). Dejosez et al. reported specific increases in H3K9me2 in genomic regions upstream of GATA4 and GATA6 coincident with THAP11 occupancy in undifferentiated embryonic
stem cells. Since neither GATA4 nor GATA6 are expressed in undifferentiated embryonic stem cells, these results indicated that THAP11 may maintain embryonic stem cell pluripotency by transcriptional repression of differentiation associated genes through chromatin and epigenetic modifications (Dejosez, Krumenacker et al. 2008).

Emerging data suggests THAP proteins may function in DNA and chromatin dependent processes. However, the transcriptional regulatory properties of most human THAP proteins and their role in physiological processes are largely unknown. The overall goal of this thesis work is to understand and characterize members of this recently identified protein family in regards to transcriptional regulation, cell function, and their potential roles in human disease. To begin to elucidate the physiological role of THAP domain containing proteins, we performed a systematic survey of publicly available gene expression data to identify conditions of differential THAP protein expression. Using this approach, two datasets were identified which revealed increased THAP10 and THAP11 mRNA expression in human DNA damage and colon cancer progression, respectively. THAP11 protein level was subsequently shown to correlate with disease progression in a panel of primary and metastatic human colon cancer specimens and cell lines. In addition, both THAP10 and THAP11 were found to potently repress transcription as heterologous Gal4 DNA-binding domain fusion proteins and directly associate with histone deacetylases (HDACs). Furthermore, preliminary experiments with previously uncharacterized THAP4 and THAP8 also suggest these proteins repress transcription.
Experimental Procedures

Plasmids and Cloning

Mammalian Gene Collection (MGC) verified full-length cDNA clones for human THAP4, 6, 8, 10 and 11 were purchased from Open Biosystems. THAP protein coding sequences were PCR amplified from their respective MGC clone and inserted into mammalian expression vectors pCMX-PL1, pcDNA3.1(-) myc/his (Invitrogen) or p3xFLAG-CMV14 (Sigma) using standard molecular cloning procedures. THAP proteins were also cloned into pCMX-Gal4_{1-147} for use in Gal4 repression assays and pGEX-4T1 (GE Lifesciences) for use in bacterial protein expression. Plasmids encoding pcDNA3.1(-)myc/his-HDAC1, -HDAC3, and -HDAC4 were described previously (Macfarlan, Kutney et al. 2005). pcDNA3.1(+) plasmids encoding FLAG epitope tagged HDAC1 (Addgene plasmid 13820), HDAC3 (Addgene plasmid 13819), HDAC4 (Addgene plasmid 13821), and HDAC5 (Addgene plasmid 13822) generated in the laboratory of Eric Verdin were purchased from Addgene. Reporter plasmids pRL-SV40 and p4xMH100-TKLuc were described previously (Macfarlan, Kutney et al. 2005) and pL8G5-Luc was provided by the laboratory of Roland Schüle with the permission of Saadi Khochbin. Plasmids pCMV-Lex-HA (Addgene plasmid 14591) and pCMV-LexVP16-HA (Addgene plasmid 14593) generated in the laboratory of Matija Peterlin were purchased from Addgene. Sequences of all recombinant constructs were verified by automated DNA sequencing.
Bacterial Protein Expression and Purification

Recombinant glutathione-S-transferase (GST), GST-THAP10, GST-THAP11 or the indicated THAP11 truncation mutants were produced in *Escherichia coli* strain Rosetta-2 BL21(DE3) (Novagen) as previously described (Macfarlan, Kutney et al. 2005). Induced GST fusion proteins were isolated from bacteria and purified according to the method of Frangioni and Neel with minor modifications (Frangioni and Neel 1993). Briefly, bacterial cell pellets were washed once in ice-cold STE buffer (10mM Tris-HCl pH 8.0, 150mM NaCl, 1mM EDTA), resuspended in STE containing 0.2mg/ml lysozyme and incubated on ice for 15 minutes. Protease inhibitors (Roche) were added as well as dithiothreitol to a final concentration of 5mM. Sarkosyl was added to a final concentration of 1.5% and the bacteria were lysed by sonication using a probe tip microsonicator (Misonix). Lysates were clarified by centrifugation at 20,000 x g for 15 minutes at 4°C. Clarified lysates were adjusted to 2% Triton X-100 and incubated with glutathione sepharose for one hour at 4°C with gentle inversion. Following incubation, glutathione sepharose was extensively washed with ice-cold PBS, resuspended in PBS supplemented with 1mM DTT and stored at 4°C short term. For long term storage GST fusion proteins were eluted with PBS containing 20mM reduced glutathione and dialyzed at 4°C overnight against PBS containing 1mM DTT. Dialyzed proteins were adjusted to 20% glycerol, aliquoted and stored at -80°C. Protein purity was assessed by SDS-PAGE and Coomassie Blue staining.
Generation and Purification of THAP11 Antibody

A custom rabbit polyclonal antibody was generated against the carboxy-terminus of human THAP11 (amino acids 132-313). Recombinant GST-THAP11 (132-313) was produced in E. coli and purified using glutathione sepharose as described above. Protein was subjected to preparative scale SDS-PAGE and gel bands corresponding to GST-THAP11 (132-313) were excised and used as immunogen. Animal immunizations and serum collection were performed by a commercial facility (Covance). GST specific antibodies were depleted by passing crude serum over a crosslinked GST-glutathione sepharose column. Further affinity purification of anti-THAP11 antibodies was performed using immunogen immobilized on nitrocellulose as described elsewhere (Sambrook, Fritsch et al. 1989). Briefly, one milligram of GST-THAP11 (132-313) was electrophoresed in a single-well SDS-PAGE gel and transferred to nitrocellulose. The nitrocellulose bound immunogen was identified by Ponceau S staining and excised. The nitrocellulose strip was blocked for one hour with 3% bovine serum albumin (BSA) in PBS and then incubated overnight at 4°C with one milliliter of anti-THAP11 antisera diluted (1:10) in PBS with 3% BSA. The antibody solution was discarded and the nitrocellulose strip washed with 150mM sodium chloride for 20 minutes at room temperature and then with PBS for 20 minutes at room temperature. The THAP11 antibody was eluted by incubating the nitrocellulose with 0.2M glycine pH 2.5, 1mM EGTA for 20 minutes at room temperature and immediately neutralized with 0.1 volume of 1M Tris-HCl pH 8. One-tenth volume of 10x PBS was added to the affinity purified anti-THAP11 antibody which was
then concentrated by filtration using Microcon centrifugal filter devices according to manufacturer’s instructions (Millipore).

**Cell Culture and Treatment**

Human cell lines HeLa (CCL-2), HepG2 (HB-8065), 293T/17 (CRL-11268), Saos-2 (HTB-85), HT-29 (HTB-38), HCT-116 (CCL-247), SW480 (CCL-228), SW620 (CCL-227) and Colo-320HSR (CCL-220.1) were purchased from the American Type Culture Collection (ATCC). HT-29 and HCT-116 were maintained in McCoy’s 5A with 10% fetal bovine serum. Colo-320HSR were grown in RPMI-1640 with 10% fetal bovine serum. All other cells were maintained in Dulbecco’s modified eagle medium (high glucose) containing 10% fetal bovine serum. Cells were maintained in a 37°C humidified incubator at 5% CO₂. To induce DNA damage, exponentially growing HeLa, HepG2, and Saos-2 cells were treated with the indicated dose of camptothecin, etoposide or cisplatin (all purchased from Sigma unless otherwise noted) for 8 hours. DNA damaging agents were dissolved immediately prior to use in DMSO. Vehicle control and all serial dilutions contained equivalent amounts of DMSO (0.1%).

**Immunoblotting and Immunofluorescence**

Whole cell extracts were prepared from subconfluent cells using modified RIPA buffer (20mM Tris-HCl pH 7.6, 150mM NaCl, 1mM EDTA, 1mM EGTA, 1% IGEPAL CA-630, 1% sodium deoxycholate, 0.25% SDS). Nuclear and cytoplasmic extracts were prepared by the method described by Dignam et al (Dignam, Lebovitz et al. 1983). Extracts were clarified by centrifugation at
20,000 x g for 15 minutes at 4°C and protein concentrations determined by BCA assay (Pierce).

For Immunoblots, 30μg of whole cell or nuclear extracts were separated on precast 8-16% polyacrylamide gels (Invitrogen). Protein was transferred to nitrocellulose membrane and stained with Ponceau S to confirm equal protein loading. Membranes were then blocked in PBS containing 0.05% Tween-20 (PBST) and 5% non-fat dry milk for one hour at room temperature. Membranes were incubated with primary antibody overnight at 4°C in PBST with 5% non-fat dry milk. Membranes were washed three times with PBST and incubated with the indicated secondary antibodies in PBST-5% non-fat dry milk for one hour at room temperature. Membranes were again washed three times with PBST and developed with ECL plus chemiluminescence detection reagent (GE Lifescience). Commercially available primary antibodies used for immunoblotting include: β-actin (Sigma), FLAG M2 (Sigma) myc 9B11 (Cell Signaling), and Histone H3 (Abcam). Horseradish peroxidase conjugated anti-mouse and anti-rabbit secondary antibodies were purchased from Sigma.

For indirect immunofluorescence, cells grown on glass coverslips were fixed with 4% paraformaldehyde in PBS for 10 minutes at room temperature. Fixed cells were rinsed three times with PBS and permeabilized with 0.2% Triton X-100 in PBS for 5 minutes at room temperature. Coverslips were washed twice with PBS, blocked with PBS-10% BSA for 30 minutes at 37°C, and incubated with affinity purified anti-THAP11 (1:100) in PBS-3% BSA for one hour at 37°C. Coverslips were then washed three times with PBS and incubated with Alexa-
488 conjugated anti-rabbit antibody (1:1000, Invitrogen) in PBS-3% BSA for one hour at 37°C. Coverslips were again washed three times in PBS, counterstained with DAPI and mounted in Prolong antifade reagent (Invitrogen). Samples were analyzed using a Zeiss LSM 510 META laser scanning confocal microscope.

**Immunohistochemistry and Tissue Microarrays**

Tissue microarray slides prepared from formalin fixed, paraffin embedded samples were obtained from the Cooperative Human Tissue Network (National Cancer Institute). Tissue microarrays were deparafinized in two changes of xylene and rehydrated in a graded alcohol series using standard procedures. Slides were subjected to heat-induced antigen retrieval by microwaving in citrate buffer (10mM Sodium Citrate, 0.05% Tween-20, pH 6.0) followed by blocking endogenous peroxidases with 3% hydrogen peroxide. Immunohistochemical staining for THAP11 was performed using affinity purified anti-THAP11 antibody (1:100), Vectastain Elite ABC detection reagents, and DAB (3,3’-diaminobenzidine tetrahydrochloride) substrate (Vector Laboratories). Hematoxylin and eosin counterstained slides were evaluated by an independent pathologist using a semi-quantitative dual-scoring system as described elsewhere (Wei, Chiriboga et al. 2006). Briefly, the intensity of THAP11 immunoreactivity was scored numerically (0=negative, 1=weak, 2=moderate, 3=strong) as was the percentage of THAP11 immunopositive cells (0=0%, 1=1-10%, 2=11-50%, 3=51-100%). The scores for intensity and percent immunopositivity were added and samples categorized into low/weak expression (combined score ≤ 3) or high/strong expression (combined score > 3) groups.
GST Pulldown Assays

Equivalent amounts of GST or the indicated GST-THAP proteins were bound to glutathione sepharose and equilibrated in 1ml of pulldown buffer (20mM Tris-HCl pH 7.6, 150mM KCl, 5mM MgCl$_2$, 1mM DTT, 10μg/ml BSA, and 0.5% IGEPAL CA-630). In vitro translated radiolabeled histone deacetylases (HDACs) were produced using the TNT-T7 coupled in vitro transcription/translation system (Promega) with $^{35}$S-methionine according to manufacturer’s instructions. $^{35}$S-labeled HDAC was added to the GST proteins in pulldown buffer and incubated for one hour at room temperature with gentle inversion. Beads were then washed three times in pulldown buffer and once in pulldown buffer without BSA. Bound proteins were eluted by boiling in 2x Laemmli sample buffer and resolved by SDS-PAGE. Gels were fixed and dried, and radiolabeled HDAC was detected by autoradiography.

Co-immunoprecipitation

293T/17 cells in 10cm$^2$ tissue culture dishes were co-transfected with 2μg each of the indicated pcDNA3.1-HDAC (FLAG or myc-his tagged) construct and either pCMX-THAP11 or p3xFLAG-THAP10 using Lipofectamine 2000 according to manufacturer’s instructions. Twenty-four hours after transfection, cell monolayers were rinsed three times with ice-cold PBS, scraped into PBS and pelleted by centrifugation at 500 x g for five minutes at 4°C. Cells were then lysed in 1ml of EBC buffer (50mM Tris-HCl pH 7.6, 125mM NaCl, 1mM EDTA, 0.5% IGEPAL CA-630) for 15 minutes at 4°C with gentle inversion. Whole cell extracts were clarified by centrifugation at 20,000 x g for 15 minutes at 4°C.
Immunoprecipitations were performed with 1μg of either anti-FLAG M2 (Sigma) or anti-myc 9B11 (Cell Signaling) monoclonal antibodies for 4 hours at 4°C with inversion. Protein G agarose (20μl packed beads) was added and immunoprecipitations allowed to proceed for an additional two hours. Beads were then washed three times with EBC buffer and bound proteins eluted by boiling in 2x Laemmli buffer. Proteins were resolved by SDS-PAGE, transferred to nitrocellulose membrane and immunoblotted as described above with indicated antibodies.

**RNA Isolation and Quantitative RT-PCR**

Total RNA was extracted using Qiagen RNeasy Mini or 96-well kits. Isolated RNA was then subjected to quantitative real-time RT-PCR using an ABI PRISM 7900HT 384-well real time PCR machine, TaqMan one-step RT-PCR kit and pre-developed primers/TaqMan probe mixes for β-actin, p21, p53, GADD45A, THAP1, THAP4, THAP7, THAP10, and THAP11 (Applied Biosystems).

**Gal4 Repression Assay**

For repression of basal transcription, 293T/17 cells in 48-well tissue culture plates were co-transfected with p4xMH100-TKLuc (100ng), pRL-SV40 (5ng), and the indicated Gal4-DBD fusion protein at 1, 5, or 25ng plasmid DNA using Lipofectamine 2000 according to manufacturer's instructions. Repression of VP16 activator driven transcription was performed similarly except pL8G5-Luc was substituted for p4xMH100-TKLuc and pCMV-Lex-HA or pCMV-LexVP16-HA
(5ng) were included as indicated. Empty pCMX-PL1 plasmid was added where necessary to maintain equal DNA amounts. The reported DNA amounts represent quantity transfected per 6-wells, or one column of the 48-well plate. Luciferase activities were measured 24 hours after transfection using the Dual Luciferase Assay kit (Promega) according to manufacturer’s instructions.

Results

THAP10 Expression is Induced by DNA Damage

Increased \textit{THAP10} gene expression was noted in a microarray dataset characterizing the gene expression profile of HeLa cells treated with the topoisomerase poison camptothecin (Carson, Zhang et al. 2004). To confirm and extend this finding, we treated HeLa cells with camptothecin, etoposide and cisplatin and examined the mRNA levels of \textit{THAP10}, \textit{p21}, and \textit{GADD45A} by quantitative RT-PCR. As shown in Figure 2.2A, \textit{THAP10} mRNA levels were induced in a dose dependent manner by each DNA damaging drug with camptothecin being the most sensitive. The dose dependent induction of \textit{THAP10} mRNA by camptothecin closely correlated with the induction of apoptosis as evidenced by immunoblotting for PARP cleavage (Figure 2.2B), suggesting THAP10 may be involved in DNA damage dependent apoptosis. Importantly, the relative level of \textit{THAP10} mRNA induction by each treatment was comparable to that observed with prototypical DNA damage inducible genes \textit{p21} and \textit{GADD45A}. HeLa cells are considered to have an attenuated p53 dependent response to DNA damage due to the presence of the human papillomavirus E6
oncogene, which accelerates proteasome mediated degradation of p53 protein (Scheffner, Werness et al. 1990). Therefore, we also examined the DNA damage dependent induction of THAP10 mRNA in cells with wild-type p53 (HepG2) or null p53 (Saos-2) status. Treatment of HepG2 cells under identical conditions as HeLa cells yielded similar results for both THAP10 and GADD45A gene expression (Figure 2.3A). However, p21 expression was markedly enhanced in HepG2 despite both GADD45A and p21 being well-characterized p53 target genes. When performed in p53-null Saos-2 cells, treatment with DNA damaging agents failed to appreciably increase the mRNA levels of THAP10, p21, or GADD45A (Figure 2.3B). To determine if the p53-null status of Saos-2 cells was at least partially responsible for the lack of THAP10 mRNA induction, these cells were transiently transfected with wildtype p53 and assayed for THAP10 and p21 mRNA levels. As shown in Figure 2.4A, ectopic expression of p53 alone was sufficient to robustly increase THAP10 (as well as p21) mRNA levels. Taken together, these results suggest that THAP10 mRNA induction by DNA damage may be partially dependent on p53 and suggests that THAP10 may play a role in the transcriptional response to DNA damage.

**THAP11 Expression in a Cell Culture Model of Colon Cancer Progression**

Increased THAP11 mRNA expression was identified in a microarray dataset originally designed to elucidate gene expression differences in the SW480/SW620 cell culture model of colon cancer progression (Provenzani, Fronza et al. 2006). SW480 and SW620 cells are isogenic colon carcinoma derived cell lines isolated from a single patient at different stages of disease
progression (Hewitt, McMarlin et al. 2000). SW480 cells were established from the primary colon adenocarcinoma while SW620 were derived from a subsequent lymph node metastasis (Hewitt, McMarlin et al. 2000). Accordingly, SW480 and SW620 cells are a frequently used model system to study the molecular underpinnings associated with colon cancer progression. We therefore reasoned that SW480/SW620 cells may also provide an ideal cell culture system to further evaluate THAP11 function.

To corroborate the microarray finding, we prepared total RNA and whole cell extracts from exponentially growing SW480 and SW620 cells and determined the relative amount of \textit{THAP11} mRNA and protein by quantitative RT-PCR and immunoblotting, respectively. While both cell types express THAP11, SW620 cells express approximately 4-fold more \textit{THAP11} mRNA than SW480 cells (Figure 2.5A). Consistent with elevated \textit{THAP11} mRNA levels, SW620 cells also demonstrated marked increases in THAP11 protein relative to SW480 cells (Figure 2.5B). Immunofluorescence analysis showed endogenous THAP11 to be located almost exclusively within the nucleus (Figure 2.5C) consistent with a potential function in DNA/chromatin dependent processes including transcriptional regulation. Examination of additional colon cancer cell lines by immunoblotting of nuclear extracts revealed weak but detectable levels of THAP11 in HCT-116 and HT29 cells (Figure 2.6). In contrast, Colo320HSR cells contained nearly as much THAP11 protein as SW620 cells (Figure 2.6). These results suggest that gain of THAP11 expression may play a role in colon cancer cell function.
THAP11 Expression is Correlated with Colon Cancer Progression

We next examined a large number of human colon cancer specimens to determine if the potential link between THAP11 expression and disease progression extends beyond colon cancer cell lines. Immunohistochemical analysis of THAP11 expression in human colon cancer tissue microarrays revealed a close correlation between increased THAP11 immunoreactivity and disease progression (Figure 2.7). The majority of samples from normal colonic epithelium (n=33) and benign adenomas (n=7) stained both weak and infrequently for THAP11 (Figure 2.7, panel N). An increase in THAP11 staining frequency and intensity was progressively observed in low-grade (grade I, well-differentiated) to high-grade (grade III, poorly differentiated) colon adenocarcinomas (n=133) (Figure 2.7 compare panels gl to gII and gIII). Frequent, high-intensity THAP11 staining was observed in both liver (n=3) and lymph node metastases (n=37) (Figure 2.7 panels LM and LNM) consistent with elevated THAP11 in metastatic colon cancer derived SW620 cells. A quantitative assessment of THAP11 immunoreactivity in all tissue microarray samples using a dual-scoring system accounting for both staining frequency and intensity revealed a statistically significant increase in THAP11 expression in primary malignant adenocarcinomas and metastases compared with normal tissues/benign adenomas (Figure 2.7B). Taken together, these data demonstrate that increased THAP11 expression correlates with colon cancer progression and suggests that SW620 cells may represent a tractable model to evaluate the function of THAP11 in colon cancer.
THAP10 and THAP11 Repress Transcription

Previous work in our laboratory identified THAP7 as a novel transcriptional repressor (Macfarlan, Kutney et al. 2005). THAP7 mediated repression was shown to be at least partially mediated through a THAP domain dependent association with HDAC3 (Macfarlan, Kutney et al. 2005). To ascertain whether THAP10 and THAP11 also possess transcriptional repressor function, we determined the effect of Gal4 DNA-binding domain THAP protein fusions on the activity of a luciferase reporter construct harboring four copies of a GAL4 DNA binding site in front of the minimal thymidine kinase promoter (p4xMH100-TK-Luc) (Figure 2.8A). Consistent with our previous THAP7 data, both THAP10 and THAP11 significantly repressed basal transcription as promoter targeted Gal4 fusions (Figure 2.8B) while Gal4 DNA-binding domain alone weakly stimulated transcription (data not shown) as reported elsewhere (Hublitz, Kunowska et al. 2005). Importantly both Gal4-THAP10 and Gal4-THAP11 revealed a dose dependent transcriptional repression activity comparable to the well-characterized nuclear hormone receptor co-repressors NCoR and SMRT suggesting that the extent of THAP10/11 transcriptional repressor function is likely to be biologically relevant. To determine if THAP10 and THAP11 can also repress activator driven transcription, we performed transient transfection assays using a luciferase reporter containing both LexA and Gal4 binding sites (Figure 2.8C). LexA-VP16 mediated transcriptional activation was significantly abrogated by co-transfection with either Gal4-THAP10 or Gal4-THAP11 (Figure
2.8D) but not Gal4 DNA-binding domain alone (data not shown). Finally, we wished to determine if additional THAP domain proteins also possess transcriptional repressor activity. Gal4-DBD fusions of human THAP4, and THAP8 but not THAP6 repressed basal transcription from p4xMH100-TKLuc when transiently transfected in 293T/17 cells (Figure 2.9). Taken together, these data indicate that THAP10 and THAP11, like THAP7, can potently repress both basal and activated transcription. Furthermore, despite limited sequence homology, several THAP domain proteins display transcriptional repressor activity when artificially tethered to promoters as Gal4 DNA binding domain fusions. These studies, together with previous results, strongly suggest that the human THAP family members may exert their physiologic effect by functioning as transcriptional repressors.

**THAP10 and THAP11 Associate with HDACs**

Since THAP10 and THAP11 repress transcription similarly to THAP7, we next addressed whether these THAP proteins can also associate with histone deacetylases, thereby providing a potential mechanism for THAP10/THAP11 mediated transcriptional repression. For this purpose, THAP11 and FLAG-tagged HDACs were transiently transfected in 293T/17 cells and THAP11/HDAC in vivo association assessed by co-immunoprecipitation. The class I histone deacetylases HDAC1 and HDAC3 were found to specifically co-precipitate THAP11 while class II HDACs, HDAC4 and 5 did not (Figure 2.10A). Additionally, HDAC1 and HDAC3 but not HDAC4 were found to directly associate with THAP11 in GST pull down assays using bacterially expressed GST-THAP11
and in vitro transcribed/translated $^{35}$S-labeled HDAC (Figure 2.10B). Aside from the THAP domain, THAP11 also contains poly-glutamine and coiled-coil motifs (Figure 2.10C), either of which may be capable of interacting with histone deacetylases. To determine the HDAC interacting region of THAP11 we repeated the GST pull down assay using either full-length (FL) GST-THAP11 or amino acid truncation mutants 1-90 and 91-314. Relative to full-length protein, GST-THAP11 (1-90) showed slightly enhanced binding to HDAC3 while GST-THAP11 (91-314) showed significantly diminished association (Figure 2.10D) indicating the THAP domain is likely the primary HDAC3 interacting motif of THAP11. Similarly, THAP10 was also found to specifically co-precipitate with myc-tagged HDAC1 and HDAC3, but not HDAC4 in 293T/17 cells under identical transfection and immunoprecipitation conditions (Figure 2.11A). Additionally, THAP10 was also found to associate directly with HDAC1 and HDAC3 by GST pulldown assay (Figure 2.11B). These data indicate that THAP domains may function as HDAC interacting modules in addition to their role as sequence specific DNA binding domains.

**Discussion**

Human THAP proteins have been suggested to function in DNA and chromatin dependent processes, including transcription. However, the transcriptional regulatory properties of most human THAP proteins have not been reported. In this chapter, we provide evidence to suggest that multiple human THAP proteins possess transcriptional repressor activity. In addition, novel biological contexts within which to evaluate THAP protein transcriptional
repression were identified. *THAP10* was found to be DNA damage inducible while *THAP11* expression was shown to correlate with colon cancer progression in human primary tumor specimens and cell lines.

Previous work from our laboratory has shown that *THAP7* is transcriptionally repressive when artificially targeted to a promoter as a heterologous Gal4-DNA binding domain (Gal4-DBD) fusion protein (Macfarlan, Kutney et al. 2005). The work described here extends this approach to demonstrate that Gal4-DBD fusions of *THAP4*, *THAP8*, *THAP10*, and *THAP11* repress basal transcription. In addition, Gal4-THAP10 and Gal4-THAP11 also significantly blunted activator driven transcription resulting from simultaneous targeting of a LexA-DBD VP16 transactivation domain fusion protein. In both instances, the extent of THAP10 and THAP11 mediated transcriptional repression was comparable to that observed with Gal4-DBD fusions of nuclear hormone receptor co-repressors NCoR and SMRT. Since NCoR and SMRT have well characterized roles in transcriptional repression, similar findings with THAP proteins under identical assays conditions suggests that THAP protein mediated repression is likely to be biologically meaningful.

A caveat to this is the relevance of the Gal4-DBD system to assess the transcriptional regulatory properties of THAP proteins. Gal4-DBD fusion proteins have been used extensively to successfully elucidate the transcriptional regulatory mechanisms of various DNA-binding or chromatin associated factors (Chen and Bieker 2001; Ishizuka and Lazar 2003; Hublitz, Kunowska et al. 2005; Sun, Yu et al. 2007; Pasini, Hansen et al. 2008). Importantly, while the Gal4-
THAP protein fusions described here repress transcription, a recent report from Mazars et al. demonstrates that Gal4-THAP3 activates transcription (Mazars, Gonzalez-de-Peredo et al. 2010). This indicates that the Gal4-THAP mediated repression reported here is likely specific to these THAP proteins and not a general phenomenon associated with Gal4-THAP fusion proteins. Nevertheless, definitive proof of THAP protein mediated transcriptional regulation undoubtedly requires additional experiments to identify directly regulated endogenous gene targets, as described for THAP11 in chapter 3 of this thesis.

Consistent with their role as transcriptional repressors, THAP10 and THAP11 were both found to physically associate with histone deacetylases HDAC1 and HDAC3. THAP10 and THAP11 co-precipitated with HDAC1 and HDAC3 from transfected cells and GST-pulldown experiments suggested this interaction was direct. It remains to be determined whether endogenous HDACs are also associated with THAP10 and THAP11. Nonetheless, in vitro domain mapping studies identified the THAP domain of THAP11 as the primary determinant of interaction with HDAC3. These results are strikingly similar to those reported previously for THAP7 and suggest that the THAP zinc-finger, like C2-H2 zinc-fingers, may function in protein-protein interactions in addition to sequence specific DNA binding (Brayer and Segal 2008). THAP domains display low primary sequence identity but, as discussed previously, adopt similar folded structures which may account for the conservation of HDAC interaction amongst THAP7, THAP10, and THAP11. Mutation of conserved zinc-coordinating or invariant residues has previously been shown to disrupt the THAP domain fold.
and abolish sequence specific DNA binding (Clouaire, Roussigne et al. 2005). Use of recombinant THAP proteins harboring these mutations in future interaction studies may reveal a similar defect in HDAC binding.

We performed a meta-analysis of published gene expression datasets to identify conditions of differential THAP protein expression that may also suggest conditions of biological relevance. This approach identified $\text{THAP10}$ as a DNA damage inducible transcript following treatment of HeLa cells with the topoisomerase poison camptothecin (Carson, Zhang et al. 2004). We extended this finding to additional DNA damaging chemotherapeutics (etoposide, cis-platin) and cell lines (HepG2). Induction of $\text{THAP10}$ mRNA by DNA damage was compromised in $p53$-null Saos-2 cells and restoration of $p53$ by transient transfection was sufficient to markedly induce $\text{THAP10}$ mRNA. Bioinformatic analysis of the $\text{THAP10}$ promoter failed to identify any high-probability $p53$ binding sites (data not shown). However, a recent genome-wide screen for chromatin bound $p53$ in 5-fluorouracil treated HCT-116 cells identified $p53$ bound in the first intron of $\text{THAP10}$, approximately 2.8kb downstream of the transcription start site (Wei, Wu et al. 2006). The relevance of this $p53$ bound region in regulating $\text{THAP10}$ DNA damage induction is currently unknown but suggests a potential for direct $p53$ mediated transcriptional activation. Indirect or $p53$ independent mechanisms may also contribute to $\text{THAP10}$ mRNA induction following DNA damage. For instance, ectopic expression of $p53$ in Saos-2 cells has previously been demonstrated to activate NF-$\kappa$B as has camptothecin treatment of HeLa cells (Ryan, Ernst et al. 2000; Carson, Zhang et al. 2004).
Thus, multiple factors may contribute to \textit{THAP10} mRNA induction by DNA damaging agents and future experiments will be required to identify the necessary factor(s).

The status of endogenous THAP10 protein levels during DNA damage remains to be determined. In our attempt to address this question, several polyclonal antibodies were raised against human THAP10. Unfortunately, each custom antibody failed to detect endogenous or transiently transfected, overexpressed THAP10. The insufficient quality of our custom antibodies and the current lack of commercial anti-THAP10 antibodies have significantly hampered our efforts to characterize endogenous THAP10. Nevertheless, the existence of endogenous THAP10 has recently been reported in a proteomic screen searching for nuclear proteins differentially phosphorylated by EGF stimulation (Olsen, Blagoev et al. 2006). Several phosphopeptides corresponding to the C-terminal portion of THAP10 were identified but not differentially phosphorylated by EGF treatment (Olsen, Blagoev et al. 2006).

If, as expected, endogenous THAP10 protein levels do increase during DNA damage then THAP10 dependent transcriptional repression may constitute a previously unrecognized component of the DNA damage response gene expression program. Future studies should determine the gene targets of THAP10 under normal and DNA damage conditions, as well as the contribution of THAP10 mediated transcriptional repression to cell cycle arrest and apoptosis resulting from genotoxic insults.
The *THAP10* gene is conserved in human and non-human primates, present in the marsupial *Monodelphis domestica* (grey short-tailed opossum) but absent in rodents. Placental (eutherian) and marsupial (metatherian) mammals diverged 180 million years ago while rodents 40 million years ago, suggesting *THAP10* was present in a shared ancestor but has since been lost in mouse and rat (Mikkelsen, Wakefield et al. 2007). Consistent with this explanation is a deletion in mouse chromosome 9 between syntenic genes *Lrrc49* and *Larp6* which, in the human genome, contains *THAP10* (data not shown). Accordingly, we speculate that THAP10 may either provide a functionally redundant role in humans or its loss compensated for in rodents.

*THAP11* was identified in a microarray dataset characterizing gene expression differences in the isogenic SW480/SW620 cell culture model of colon cancer progression. *THAP11* mRNA and protein levels were found be increased in metastatic SW620 cells versus primary tumor derived SW480 cells. Elevated THAP11 protein levels were also found in the neuroendocrine colon carcinoma Colo320HSR and the metastasis derived LoVo cell line (data not shown) but other colon cancer cell lines exhibited markedly reduced, albeit detectable, THAP11 protein.

THAP11 expression was also found to positively correlate with disease progression in human primary tumor specimens as determined by immunohistochemical analysis of tissue microarrays. THAP11 immunoreactivity was largely absent in normal colonic epithelia but significantly increased in low-to high-grade carcinomas and metastases. THAP11 immunoreactivity at
metastatic sites was only slightly greater than that observed at primary tumors suggesting the THAP11 gain-of-function event likely occurs within the primary tumor and persists following metastasis dissemination and metastatic site colonization.

The contribution of THAP11 towards promoting primary tumor formation or distant metastasis remains an unanswered but pertinent question. Gain of THAP11 expression in colon cancer may, via THAP11 mediated transcriptional repression, contribute to disease progression. A similar scenario has been proposed for the E-box binding transcriptional repressors Snail and Zeb1. Increased expression of Snail and Zeb1 in colon cancer cells has previously been shown to affect tumorigenic and metastatic potential by repressing gene products associated with differentiated epithelial cells (Peinado, Olmeda et al. 2007). Alternatively, THAP11 may represent a “bystander gene” where increased expression in colon cancer results from, rather than drives, disease progression. Discriminating between these possibilities will likely require multiple lines of further experimentation. Cell based in vitro and nude mouse tumor xenograft assays using both THAP11 overexpression and RNA interference mediated knockdown in colon cancer cell lines should prove valuable in determining if THAP11 plays a functional role in colon cancer progression. Initial results utilizing THAP11 knockdown in SW620 cells are described in chapter 3 of this thesis.

While this work was in progress, two independent reports implicated THAP11 as a biologically relevant transcriptional repressor (Dejosez,
Krumenacker et al. 2008; Zhu, Li et al. 2009). Zhu et al. reported that human THAP11 was ubiquitously expressed and largely downregulated in several human cancers including liver cancer (Zhu, Li et al. 2009). This finding contrasts with our observations and may represent a tissue specific role for THAP11 in human liver versus colon cancers. In addition, their assertion that THAP11 is ubiquitously expressed and repressed in human liver cancer was based solely on mRNA expression data and may not accurately reflect THAP11 protein status. Furthermore, close inspection of the THAP11 expression data from their multiple tissue northern blot array (Fig. 1 in Zhu et al.) suggests elevated \textit{THAP11} mRNA in several tumor versus normal colon tissues, consistent with our immunohistochemistry results. Additional conflicting data regarding the tissue distribution of THAP11 was also recently reported by Dejosez et al. who found detectable THAP11 (termed Ronin) mRNA in mouse ES cells but not adult tissues (Dejosez, Krumenacker et al. 2008). Preliminary data generated by other members of our laboratory and not presented in this thesis demonstrates detectable THAP11 protein in nuclear extracts from human cancer cell lines of diverse origin including breast, prostate, ovary, uterus, and both T- and B-cell leukemia. Whether this expression pattern represents the cancerous origin of these cells or reflects a species specific THAP11 expression pattern, as suggested by Zhu et al., remains to be determined.

In agreement with our work, both recent THAP11 reports further substantiate our finding that THAP11 is a transcriptional repressor (Dejosez, Krumenacker et al. 2008; Zhu, Li et al. 2009). Dejosez et al. used an in vitro
selection procedure to demonstrate that THAP11/Ronin could bind DNA directly and also found that overexpressed THAP11/Ronin repressed global transcription in mouse ES cells. Zhu et al. reported that overexpressed THAP11 directly repressed c-myc transcription in the human HepG2 cells. These findings, in conjunction with our results, provide substantial evidence to suggest that THAP11 likely functions as a transcriptional repressor.

The data presented in this chapter provides important new evidence that further implicates THAP proteins in transcriptional regulation. Several previously uncharacterized THAP proteins, including THAP10 and THAP11, were found to repress transcription. THAP10 and THAP11 were shown to be differentially expressed during DNA damage and colon cancer progression, respectively, suggesting THAP protein relevance to human disease. Future studies aimed at determining the function of THAP10 in DNA damage, THAP11 in colon cancer, and identification of their endogenous gene targets are currently underway in our laboratory and will be important new avenues of THAP protein research.
Figure 2.1: Human THAP domain containing proteins. (A) Schematic of human THAP proteins. The N-terminally located THAP domains are indicated by the gray boxes. Coiled-coil motifs present in THAP1-7 and THAP11 are indicated by black boxes. The location of the poly-glutamine rich region of THAP11 and the extended region of P element transposase homology in THAP9 are indicated by red and blue boxes, respectively. Adapted and re-drawn from (Mazars, Gonzalez-de-Peredo et al. 2010). (B) Multiple sequence alignment of human THAP proteins. Invariant residues are indicated by asterisks and the AVPTIF box is indicated with a bracket.
Table 2.1: THAP proteins in model organisms.

<table>
<thead>
<tr>
<th>Organism</th>
<th>THAP Proteins</th>
<th>Putative Homologs</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Homo sapiens</em></td>
<td>12</td>
<td>THAP0-THAP11</td>
</tr>
<tr>
<td><em>Mus musculus</em></td>
<td>7</td>
<td>THAP0, 1, 2, 3, 4, 7, and 11</td>
</tr>
<tr>
<td><em>Gallus gallus</em></td>
<td>6</td>
<td>THAP0, 1, 4, 5, and 7</td>
</tr>
<tr>
<td><em>Xenopus laevis</em></td>
<td>23</td>
<td>THAP1, 4, 7, and 11</td>
</tr>
<tr>
<td><em>Danio rerio</em></td>
<td>32</td>
<td>THAP0, 1, 7, 9, and 11</td>
</tr>
</tbody>
</table>

Adapted from (Clouaire, Roussigne et al. 2005).
Figure 2.2: THAP10 mRNA is DNA damage inducible. (A) HeLa cells were treated with the indicated DNA damaging agent for 8 hours and THAP10 mRNA levels determined by quantitative RT-PCR. Values represent the mean ± standard deviation of cells treated in quadruplicate. (B) HeLa cells were treated with increasing amounts of camptothecin as in (A) and apoptosis determined by immunoblotting for cleaved PARP.

Camptothecin | Etoposide | Cisplatin

Fold Change mRNA

Camptothecin (μM)

0  0.3  1  3  10

Full-length

Cleaved

PARP

β-actin
Figure 2.3: THAP10 mRNA induction by DNA damage in HepG2 and Saos-2 cells. (A) HepG2 and (B) Saos-2 cells were treated with the indicated DNA damaging agent for 8 hours and THAP10 mRNA levels determined by quantitative RT-PCR. Values represent the mean ± standard deviation of cells treated in quadruplicate.
Figure 2.4: Ecotopic p53 expression induces THAP10 mRNA. (A) Saos-2 cells were transfected with either empty or p53 expression vector (2μg per well in a six-well plate) and analyzed for THAP10 and p21 mRNA levels by quantitative RT-PCR 24 hours later. RNA levels are normalized to β-actin and expressed as fold change relative to empty vector transfected cells. Values represent the mean ± standard deviation from three transfected wells from a representative experiment performed at least three times. (B) Immunoblot from cells transfected as in (A) indicating ectopic p53 expression and resulting p21 upregulation.
Figure 2.5: THAP11 expression in SW480 and SW620 colon cancer cell lines. (A) THAP11 mRNA levels in SW480 and SW620 cells were determined by quantitative RT-PCR. THAP11 mRNA levels were normalized to β-actin and expressed relative to SW480. Values represent the mean ± standard deviation of triplicate quantitative RT-PCR reactions from a representative experiment performed at least three times. (B) THAP11 protein levels were determined by immunoblotting whole cell extracts from SW480 and SW620 cells using a custom polyclonal THAP11 antibody. Immunoblotting for β-actin served as a loading control. (C) Immunofluorescence localization of endogenous THAP11 in SW620 cells. THAP11 was detected in SW620 cells by indirect immunofluorescence using an affinity purified custom polyclonal THAP11 antibody as described in Experimental Procedures. Nuclei were identified by counterstaining with DAPI and cells were visualized by differential interference contrast (DIC) microscopy. Merge represents the overlay of THAP11, DAPI, and DIC images.
Figure 2.6: THAP11 expression in colon cancer cell lines. THAP11 protein levels were determined by immunoblotting nuclear extracts from the indicated colon cancer cell lines using a custom polyclonal THAP11 antibody. Immunoblotting for histone H3 served as a loading control.
Figure 2.7: Immunohistochemical analysis of THAP11 expression in human colon cancer specimens. (A) Representative images of THAP11 immunohistochemical staining in normal colon epithelium (N), grade I (gI), grade II (gII), grade III (gIII) adenocarcinomas and metastases from liver (LM) and lymph node (LNM). (B) Quantitative analysis of THAP11 expression in tissue microarray (TMA) samples. THAP11 immunoreactivity was scored as described in Experimental Procedures and samples were placed into either high/strong (>3) or low/weak (≤3) THAP11 expression groups. Asterisk (*) denotes statistical significance (p<0.05) relative to normal/adenoma as determined by T-test.
Figure 2.8: THAP10 and THAP11 repress basal and activated transcription. (A) Schematic of p4xMH100 luciferase reporter used to determine repression of basal transcription by Gal4DBD-fusion proteins. The luciferase reporter, minimal thymidine kinase promoter (TK) and four copies of the MH100 Gal4 upstream activating sequence (Gal4) are indicated. (B) 293T/17 cells in 48-well plates were co-transfected with 100ng of p4xMH100-TK-Luc, 5ng of pRL-SV40, and the indicated Gal4DBD-fusion expression vector as described in Experimental Procedures. Cells were processed for dual luciferase assay 24 hours later and luciferase activity expressed relative to empty vector transfected cells. (C) Schematic of pL8G5 luciferase reporter used to determine repression of activated transcription by Gal4DBD-fusion proteins. The luciferase reporter, TATA-box (TATA), five copies of Gal4 upstream activating sequence (Gal4), and eight copies of LexA binding sites (LexA) are indicated. 293T/17 cells in 48-well plates were co-transfected with 100ng of pL8G5-Luc, 5ng of pRL-SV40, 5ng of the indicated LexA expression vector and 1, 5, or 25ng of the indicated Gal4-THAP fusion protein as described in Experimental Procedures. Cells were processed for dual luciferase assay 24 hours later and luciferase activity expressed relative to LexA-VP16 transfected cells. Values represent the mean ± standard deviation from 6 wells per condition of a representative experiment performed at least three times with similar results.
Figure 2.9: Transcriptional regulatory properties of THAP4, THAP6, and THAP8. 293T/17 cells in 48-well plates were co-transfected with 100ng of p4xMH100-TK-Luc, 5ng of pRL-SV40, and the indicated amounts of Gal4-THAP4 (A), Gal4-THAP6 (B) or Gal4-THAP8 (C) expression vectors as described in Experimental Procedures. Cells were processed for dual luciferase assay 24 hours later and luciferase activity expressed relative to empty vector transfected cells. Values represent the mean ± standard deviation from 6 wells per condition of a representative experiment performed at least three times with similar results.
Figure 2.10: THAP11 associates with HDACs. (A) THAP11 co-immunoprecipitates with HDAC1 and HDAC3. 293T/17 cells were transfected with expression plasmids for THAP11 and the indicated FLAG-HDAC or empty vector (2 μg each). Whole cell extracts were immunoprecipitated (IP) with anti-FLAG M2 monoclonal antibody and immunoblotted (IB) with anti-THAP11 antibody. Whole cell extracts corresponding to 2% of input material (Input) were analyzed on a parallel blot. Blots were stripped and reprobed with anti-FLAG to reveal HDACs. IgG heavy-chain is indicated by the arrow. (B) THAP11 associates with HDAC1 and HDAC3 in vitro. (Left panel) Coomassie Blue stained SDS-PAGE gel of recombinant GST and GST-THAP11 (GST-T11) used in GST-pulldown assay. (Right panel) GST and GST-THAP11 were incubated with in vitro transcribed and translated 35S-labeled HDAC1, HDAC3, and HDAC4. Bound radiolabeled HDACs were resolved by SDS-PAGE and detected by autoradiography. Input corresponds to 10% of starting material. (C) Schematic of THAP11. THAP, poly-glutamine (Q), and coiled-coil motifs of THAP11. Numbers below each motif correspond to the amino acid residues. (D) THAP domain is necessary and sufficient for HDAC interaction. (Left panel) Coomassie Blue stained SDS-PAGE gel of GST and the indicated GST-THAP11 fusion protein: full-length (FL), amino acids 1-90 (1-90) and amino acids (91-314). (Right panel) GST-pulldown assay using the indicated GST-THAP11 fusion protein was performed using 35S-labeled HDAC3 as described in (B).
Figure 2.11: THAP10 associates with HDACs. (A) THAP10 co-immunoprecipitates with HDAC1 and HDAC3. 293T/17 cells were transfected with expression plasmids for THAP10-3xFLAG and the indicated myc-HDAC or empty vector (2 μg each). Whole cell extracts were immunoprecipitated (IP) with anti-myc 9B11 monoclonal antibody and immunoblotted (IB) with anti-FLAG antibody to detect co-precipitated THAP10. IgG heavy-chain is indicated by the arrow. Whole cell extracts corresponding to 2% of input material (Input) were analyzed on a parallel blot. Blots were stripped and reprobed with anti-myc to reveal HDACs. (B) THAP10 associates with HDAC1 and HDAC3 in vitro. (Left panel) Coomassie Blue stained SDS-PAGE gel of recombinant GST and GST-THAP10 (GST-T10) used in GST-pulldown assay. (Right panel) GST and GST-THAP10 were incubated with in vitro transcribed and translated [35]S-labeled HDAC1, and HDAC3. Bound radiolabeled HDACs were resolved by SDS-PAGE and detected by autoradiography. Input corresponds to 10% of starting material.
Chapter 3: Identification of Endogenous THAP11 Gene Targets and a Potential Role for THAP11 in Colon Cancer Cell Proliferation

Summary

THAP11 expression was previously shown to positively correlate with colon cancer progression in human tumor samples and cell lines. This finding, in conjunction with the observation that promoter targeted THAP11 represses transcription, suggests THAP11 dependent transcriptional repression may contribute to primary tumor or metastasis formation through transcriptional regulation. However, endogenous gene targets of THAP11 and its role in colon cancer cell biology are currently unknown. In this chapter, we describe the use of retrovirus-delivered short-hairpin RNA to establish stable knockdown of THAP11 expression in metastatic colon cancer SW620 cells. Microarray based gene expression profiling in THAP11 knockdown cells identified 80 differentially expressed genes, of which 70% were de-repressed upon THAP11 depletion. De-repressed genes were shown to contain chromatin bound THAP11 near their transcription start sites suggesting direct transcriptional repression by THAP11. THAP11 was subsequently shown to associate with and recruit HCF-1 to repressed promoters. Knockdown of HCF-1 similarly de-repressed THAP11 target genes suggesting HCF-1 is a critical effector of THAP11 mediated repression. Knockdown of THAP11 in SW620 cells resulted in diminished cell...
proliferation. Taken together, these results implicate THAP11 transcriptional repression as a potential regulator of colon cancer cell growth in SW620 cells.

**Introduction**

In the previous chapter, we demonstrated that several THAP proteins possess transcriptional repressor activity when tethered to promoters as heterologous Gal4-DNA binding domain fusion proteins. THAP11 potently repressed basal and activated transcription and physically associated with histone deacetylases. In agreement with these observations, two independent reports published while this work was in progress identified THAP11 dependent transcriptional repression in mouse embryonic stem cells and human liver cancer cell lines (Dejosez, Krumenacker et al. 2008; Zhu, Li et al. 2009). These findings suggest that endogenous THAP11 likely functions as a biologically relevant transcriptional regulatory protein. However, the identity of endogenous THAP11 gene targets and mechanism of transcriptional regulation remain largely unknown.

The work presented in chapter 2 also found THAP11 expression positively correlates with colon cancer progression. Significantly increased THAP11 protein levels were observed in primary and metastatic human tumor specimens. THAP11 mRNA and protein levels were also increased in metastatic SW620 colon cancer cells relative to isogenic, primary tumor derived SW480 cells. Derived from a single patient at different stages of disease progression, SW480 and SW620 cells retain several important phenotypic differences associated with primary and metastatic tumors, respectively. SW620 xenografts in nude mice
show increased growth rates, are less well-differentiated, and invade surrounding normal tissues more readily than SW480 tumors (Hewitt, McMarlin et al. 2000). Also, intrasplenic injections of SW620 cells metastasize to liver whereas SW480 cells do not (Hewitt, McMarlin et al. 2000). Thus, SW620 cells represent a unique model to not only further characterize the transcriptional regulatory properties of endogenous THAP11 but also investigate the role of THAP11 in colon cancer progression. To that end, work in this chapter describes the identification of directly repressed gene targets of endogenous THAP11 in SW620 cells. In addition, we find that stable knockdown of THAP11 results in decreased SW620 cell proliferation suggesting THAP11 dependent transcriptional regulation may contribute to colon cancer progression.

Experimental Procedures

Plasmids and Cloning

Retroviral expression vectors pBABE-puro and pBABE-hygro have been described elsewhere (Morgenstern and Land 1990). Vector pBABE-EGFP was generated from pBABE-puro by replacing the puromycin resistance cassette (HindIII/ClaI excised) with the EGFP coding sequence which was PCR amplified from pEGFP-C2 (Clontech) using the forward primer

AATT\*AAGCTT\*CACCATGGTGAGCAAGGGCGA and reverse primer

AATTATCGATTTTACTTTGTACAGCTCGTCCATGC (restriction endonuclease sites are underlined).
Retroviral short-hairpin RNA expression vectors pSuper.Retro.Puro and pSuper.Retro.Neo+GFP, hereafter abbreviated pSRP and pSRNG, respectively, were purchased from Oligoengine as was the nonspecific shRNA pSRP MAMM-X (designated shNS). THAP11, Annexin A1 and HCF-1 shRNA targeting sequences were designed using Dharmacon’s siDesign Center website. Synthetic oligonucleotides corresponding to shRNA targeting sequences were cloned into BglII/HindIII linearized pSRP and pSRNG according to manufacturer’s instructions. An additional control shRNA (designated shNS2) targeting GFP was also cloned into pSRP. A summary of shRNA targets and the corresponding synthesized oligonucleotide sequences is provided in Table A.1 and Table A.2, respectively.

An expression construct for non-silenceable THAP11 (p3xFLAG-THAP11-Rescue) resistant to shRNAs T11A, T11C and T11E was generated by two successive rounds of site-directed mutagenesis using p3xFLAG-THAP11 and the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Each mutagenesis reaction introduced silent mutations into three consecutive codons within the shRNA targeting sequence using primers listed in Table A.3. The p3xFLAG-THAP11-Rescue construct was further subcloned by PCR amplifying the non-silenceable THAP11 coding sequence including 3xFLAG tag with EcoRI/SalI adapted primers and inserting into the EcoRI/SalI sites of pBABE-puro and pBABE-EGFP. The final retroviral expression constructs were named pBABE-puro-THAP11-Rescue-3xFLAG and pBABE-EGFP-THAP11-Rescue-3xFLAG.
Sequence correctness of all constructs was verified by automated DNA sequencing.

**Immunoblotting**

Extract preparation and immunoblotting procedures were previously described in Chapter 2. Briefly, whole cell extracts were prepared using modified RIPA buffer (20mM Tris-HCl pH 7.6, 150mM NaCl, 1mM EDTA, 1mM EGTA, 1% IGEPAL CA-630, 1% sodium deoxycholate, 0.25% SDS) and protein concentration determined by BCA assay. Proteins (30μg) were separated by SDS-PAGE on 8-16% precast polyacrylamide gels and transferred to nitrocellulose. Blots were blocked in PBS, 0.05% Tween-20 (PBST) containing 5% non-fat dry milk and incubated with primary antibody overnight at 4°C in PBST-5% non-fat dry milk. Membranes were washed with PBST, probed with horseradish peroxidase-conjugated secondary antibodies in PBST-5% non-fat dry milk and developed using enhanced chemiluminescence. Immunoblotting for THAP11 was performed using our custom generated THAP11 antibody. Mouse monoclonal antibodies against β-actin (working dilution 1:10,000) and FLAG-M2 (1:3000) were purchased from Sigma. Monoclonal anti-annexin A1 (1:3000) was purchased from Becton-Dickinson. HCF-1 antibody (1:5000) was purchased from Bethyl Laboratories.

**Cell Culture**

293T/17 (CRL-11268) cells and human colon cancer cell lines SW620 (CCL-227), HCT-116 (CCL-247), SW480 (CCL-228), Colo-320HSR (CCL-220.1)
and LoVo (CCL-229) were purchased from ATCC. 293T/17, SW480 and SW620 were maintained in Dulbecco’s modified eagle medium (high glucose) containing 10% fetal bovine serum. Colo-320HSR, HCT-116 and LoVo were grown in RPMI-1640, McCoy’s 5A and F-12K, respectively, containing 10% fetal bovine serum. All cells were grown without supplemental antibiotics in a humidified 37°C incubator containing 5% CO₂.

**Retrovirus Production**

VSV-G pseudotyped retrovirus was produced in 293T/17 cells (~70% confluent in 10cm² dishes) by co-transfection with pCMV-VSVG (4μg), pMLV-GagPol (8μg), and retroviral construct (12μg) using Lipofectamine 2000 (Invitrogen) according to manufacturer’s instructions. Following overnight transfection, cells were given fresh media (DMEM, 10% FBS) and allowed to equilibrate for several hours in the 37°C cell culture incubator prior to being shifted to a 32°C, 5% CO₂, humidified cell culture incubator for an additional 24-30 hours. Retroviral supernatants were harvested ~48 hours from the start of transfection, cleared of residual 293T/17 cells by centrifugation (2000 x g, 5 minutes) and either used immediately or aliquoted and stored at -80°C for future use.

**Retroviral Transduction and Stable Cell Production**

To generate pools of SW620 and Colo-320HSR cells stably expressing either control or THAP11 shRNA, cells in six-well plates (~20% confluent) were transduced with a 1:1 mixture of viral supernatant and fresh media, adjusted to
8μg/ml polybrene and spin-infected at 500 x g for two hours at 32°C. Following spin-infection cells were returned to the 37°C incubator for 8-16 hours and then given fresh growth media. Two days post-transduction, cells from individual wells of the six-well plate were split into 10cm² dishes containing the appropriate selection medium (2μg/ml puromycin or 1mg/ml Geneticin). Cells were grown under selection for an additional two (puromycin) to seven (Geneticin) days at which point mock transduced cells exhibited complete cell death. For rescue experiments, SW620 cells were first transduced with pSRP shRNA virus and 24 hours later transduced with the indicated pBABE-EGFP virus. Doubly expressing cells were then selected with puromycin and sorted for GFP expression using fluorescence activated cell sorting (FACS). Alternatively, rescue experiments were also performed using SW620 cells transduced with pSRNG shRNA virus, FACS sorted and then transduced with the indicated pBABE-puro rescue virus followed by puromycin selection. To generate double-knockdown cells expressing THAP11 and Annexin A1 shRNAs, SW620 cells were first transduced with pSRP-THAP11 C retrovirus and 24 hours later transduced with pSRNG-Annexin A1 D retrovirus. Twenty-four hours after the second transduction, cells were split into 10cm² dishes and selected with 2μg/ml puromycin and 1mg/ml Geneticin for seven days.

**RNA Isolation and Quantitative RT-PCR**

Total RNA was isolated using Qiagen RNeasy Mini Kit according to manufacturer's instructions. Total RNA (1μg) was reverse transcribed using qScript cDNA synthesis mix (Quanta Biosciences) containing both random
hexamer and oligo dT primers. Quantitative PCR was performed on diluted cDNA using an ABI PRISM 7900HT 384-well real time PCR machine (Applied Biosystems) in a final volume of 20μl using SYBR green PCR master mix (Applied Biosystems) and gene specific primers. Fold change in mRNA levels was determined using the ΔΔCt method normalized to β-actin.

**Microarray Gene Expression Analysis**

Gene expression analysis of SW620 cells stably expressing either control (pSRP-shNS, pSRP-shNS2) or THAP11 (pSRP-T11A, pSRP-T11C) shRNAs were determined using Nimblegen *Homo sapiens* 385K oligonucleotide microarrays. Total RNA from two independent pools of SW620 cells per shRNA were isolated using Qiagen RNeasy Mini Kits as described above. RNA quality was verified using an Agilent 2100 Bioanalyzer and provided to Nimblegen for subsequent cDNA synthesis, labeling and microarray hybridization.

**Nuclear Run-On Assay**

Modified nuclear run-on assays using 5-bromouridine (BrU) labeled nascent RNA were performed as previously described with minor modifications (Core, Waterfall et al. 2008). SW620 cells (~1 x 10⁸) expressing either pSRP-shNS, pSRP-shT11A, or pSRP-shT11C were rapidly cooled by rinsing in ice-cold PBS, scraped into ice-cold PBS and collected by centrifugation (500 x g, 5 minutes, 4°C). Cell pellets were resuspended in hypotonic lysis buffer (10mM Tris pH 7.6, 10mM NaCl, 3mM MgCl₂) and immediately centrifuged as before. Cell pellets were loosened by gentle vortexing (setting 6), resuspended in
hypotonic lysis buffer containing 0.5% IGEPAL CA-630, and incubated on ice for
5 minutes. Cell lysis was routinely >90% as determined by trypan blue staining
and hemacytometer counting. Nuclei were recovered by centrifugation (1000 x
g, 5 minutes, 4°C), washed once in hypotonic lysis buffer containing 0.5%
IGEPAL CA-630 and centrifuged as before. Recovered nuclei were resuspended
in 1ml of freezing buffer (50mM Tris pH 8.0, 40% glycerol, 5mM MgCl2, 0.1mM
EDTA, 40U RnasinPlus (Promega)) and aliquoted into 1.5ml tubes at 200μl (~10^7
nuclei) per tube, snap frozen in liquid nitrogen and stored at -80°C.

For run-on transcription, thawed nuclei were mixed with an equal volume
(200μl) of 2x run-on buffer (10mM Tris pH 8, 5mM MgCl2, 300mM KCl, 1%
Sarkosyl, 160U RNASinPlus, 5mM DTT, and 1mM each ATP, CTP, GTP, BrUTP)
and incubated at 30°C for 30 minutes. Following run-on transcription, samples
were DNaseI and proteinase K digested. Total RNA was extracted using acid
phenol:chloroform:iso-amyl alcohol (25:24:1), ethanol precipitated and
resuspended in nuclease-free water. Total RNA was then further purified using
Qiagen RNeasy mini spin columns according to the RNA cleanup procedure
described by the manufacturer. Anti-BrdU monoclonal antibody (Sigma, 5μl per
IP, 15μl total) was pre-incubated with 90μl of Protein G Dynabeads in 1ml
binding buffer (10mM Tris pH 7.6, 100mM NaCl, 1mM EDTA, 0.05% Tween-20,
0.1% IGEPAL CA-630, 15μg yeast tRNA) for 1 hour at 4°C. Dynabeads were
washed three times in binding buffer without supplemental tRNA and
resuspended in 90μl of binding buffer. To immunoprecipitate BrU labeled
nascent RNA, 25μg of nuclear run-on RNA, 2.5μg yeast tRNA and 40U
RNasinPlus were added to 500μl of binding buffer on ice. Anti-BrdU bound Dynabeads (30μl per IP) were added and immunoprecipitation allowed to proceed for one hour at 4°C with end-over-end rotation. Beads were subsequently washed three times in binding buffer and eluted by addition of 300μl of Buffer RLT (Qiagen RNeasy Kit). Immunoprecipitated RNAs were purified using Qiagen RNeasy Mini Kit and eluted in 30μl of water. Isolated RNA was reverse transcribed and analyzed by quantitative RT-PCR as described above.

**Chromatin Immunoprecipitation**

Chromatin immunoprecipitation was performed essentially as described with minor modifications (Lee, Johnstone et al. 2006). SW620 cells (~ 5 x 10^7) or those stably expressing the indicated retroviral constructs were cross-linked by addition of 1/10th volume of freshly prepared formaldehyde cross-linking buffer (11% formaldehyde, 10mM HEPES-KOH pH 7.6, 100mM NaCl, 1mM EDTA, 0.5mM EGTA) to tissue culture dishes and incubated for 10 minutes at room temperature. Cross-linking was terminated by addition of 1/20th volume 2.5M glycine. Cross-linked cells were washed three times with ice-cold PBS, scraped into PBS and recovered by centrifugation (1500 x g, 10 minutes, 4°C). Cell pellets were resuspended in 10mls of lysis buffer 1 (50mM HEPES-KOH pH 7.6, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% IGEPAL-CA630, 0.25% Triton X-100, 1x protease inhibitors) and incubated for 10 minutes at 4°C with gentle rocking. Nuclei were recovered by centrifugation, resuspended in 10mls of lysis buffer 2 (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA, 1x
protease inhibitors) and extracted for 10 min at room temperature with gentle inversion. Nuclei were again recovered by centrifugation, resuspended in 3.5 mls of lysis buffer 3 (10 mM Tris- HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na- Deoxycholate, 0.5% Sarkosyl, 1 x protease inhibitors) and sonicated in an ice-water bath using a Misonix micro-tip equipped sonicator at setting 6 (~6 W RMS output power) for 12 cycles of 15 seconds sonication followed by a 1 minute cooling interval. The sonicated chromatin was adjusted to 1% Triton X-100 from a 10% stock solution and debris removed by centrifugation at 20,000 x g at 4°C for 20 minutes. The protein concentration of solubilized chromatin was determined by BCA assay and approximately 1 mg of chromatin was immunoprecipitated overnight at 4°C with the indicated antibodies. Protein G Dynabeads (30 μl) blocked overnight in PBS containing 0.5% BSA and 0.1% Triton X-100 were added and immunoprecipitation allowed to proceed for an additional two hours. Beads were collected using a Dynal MPC magnetic stand (Invitrogen) and washed four times with 1 ml of ChIP-RIPA wash buffer (50 mM HEPES-KOH pH 7.6, 500 mM LiCl, 1 mM EDTA, 1.0% I GEPAL-CA630, 0.7% Na-Deoxycholate) and once with TE containing 50 mM NaCl.

Following the final wash DNA was recovered as described by Nelson et al. (Nelson, Denisenko et al. 2006). Briefly, Dynabeads as well as precipitated input chromatin were resuspended in 100 μl of 10% Chelex resin (BioRad) and incubated for 10 minutes at 100°C. Samples were cooled to room temperature and then digested with proteinase K (0.2 mg/ml) for one hour at 55°C. Samples were again boiled for 10 minutes to inactivate proteinase K and centrifuged at
14,000 x g for 3 minutes to pellet the Chelex/Dynabeads mixture. Supernatants (80 μl) containing the immunoprecipitated DNA were transferred to clean 1.5ml tubes and the Chelex/Dynabead resins were resuspended in an additional 120 μl of water, vortexed, and centrifuged as before. Supernatants were combined yielding 200 μl of immunoprecipitated DNA.

Determination of relative enrichment was performed by qPCR using an ABI PRISM 7900HT 384-well real time PCR machine with SYBR green PCR master mix (Applied Biosystems) and primers designed as described below. Threshold cycle (Ct) values of ChIP-enriched DNA were exponentiated and expressed as percent recovery relative to input DNA analyzed in parallel.

Primer pairs for qPCR were designed using Primer3Plus against human genome sequence (NCBI36/hg18) retrieved using the UCSC Genome Browser (Kent, Sugnet et al. 2002). The primer sequences and relative position of the ChIP amplicons is provided in Table A.4.

**Co-immunoprecipitation and Immunoblot Analysis**

Cells in 15cm² tissue culture dishes were rinsed three times with ice-cold PBS, scraped into PBS and collected by centrifugation at 500 x g for five minutes at 4°C. Cells were then resuspended in 5 pellet cell volumes (PCV) of Buffer A (10mM HEPES-KOH pH 7.6, 10mM KCl, 1.5mM MgCl2) containing Complete protease inhibitors (Roche) and allowed to swell for 10 minutes on ice. Cytoplasmic membranes were lysed by drop-wise addition of IGEPAL CA-630 from a 10% stock solution to a final concentration of 0.5% while gently mixing the cells by vortexing at half-maximum setting. Cells were incubated on ice for 5
minutes and plasma membrane lysis was verified by trypan blue staining. Nuclei were isolated by centrifugation at 2000 x g for 5 minutes at 4°C, washed once with Buffer A, collected by centrifugation and resuspended in 1 PCV of Buffer C (20mM HEPES-KOH pH 7.6, 420mM NaCl, 1.5mM MgCl₂, 0.2mM EDTA, 25% glycerol) supplemented with Complete protease inhibitors (Roche). Nuclei were extracted for 45 minutes at 4°C with gentle inversion. Nuclear extracts were clarified by centrifugation (20,000 x g for 15 minutes at 4°C), diluted with one volume of Buffer C without glycerol or NaCl, and adjusted to 0.5% IGEPAL CA-630. Nuclear extracts were re-clarified by centrifugation to remove precipitates formed by dilution and protein concentration determined by BCA assay.

Immunoprecipitations were performed with 0.5-1mg of nuclear extract and 1-2μg of either affinity purified rabbit polyclonal anti-HCF-1 (Bethyl Laboratories) or normal rabbit IgG (Sigma) overnight at 4°C with inversion. Protein G Dynabeads (20μl) were added and immunoprecipitations allowed to proceed for an additional two hours. Beads were then washed four times with binding buffer (20mM HEPES-KOH pH 7.6, 150mM NaCl, 1.5mM MgCl₂, 0.2mM EDTA, 0.5% IGEPAL CA-630) and bound proteins eluted by boiling in 2x Laemmli buffer. Proteins were resolved by SDS-PAGE, transferred to nitrocellulose membrane and immunoblotted with HCF-1 and THAP11 antibodies. Blots were developed by enhanced chemiluminescence using an anti-rabbit light chain specific HRP-conjugated secondary antibody (Jackson Immunoresearch) to minimize obscuring THAP11 signal by co-migrating IgG heavy chains.
Alamar Blue Cell Proliferation Assay

Proliferation of SW620 knockdown cells was determined using the Alamar Blue cell viability reagent according to manufacturer’s instructions (Invitrogen). Puromycin or puromycin/geneticin selected (as indicated) SW620 knockdown cells were seeded at 5000 cells per well (n = 8 wells per condition) in black-walled 96-well tissue culture plates in 100μl of DMEM, 1% FBS. Five identical 96-well plates were seeded for determination of cell viability on five consecutive days. Twenty-four hours after seeding, 10μl of Alamar Blue reagent was added per well to one 96-well plate and this plate was incubated for two hours at 37°C in the cell culture incubator. Fluorescence measurements were then performed with a BioTek Synergy HT multi-detection microplate reader equipped with 540nm excitation and 590nm emission filters. Average background fluorescence was calculated from wells (n = 16) containing media alone and subtracted from wells containing cells. This process was repeated every 24 hours for 5 consecutive days using one replicate 96-well plate per day.

Results

Gene Expression Profiling in SW620-THAP11 Knockdown cells

To further characterize the transcriptional regulatory properties of THAP11 in a physiologically relevant system, we performed microarray based gene expression profiling using SW620 cells depleted of THAP11 via retrovirally expressed short-hairpin RNAs (shRNA). Central to this approach was the creation of multiple THAP11 specific shRNA constructs. As shown in Figure 3.1,
shRNA constructs shT11A, shT11C, shT11D, and shT11E significantly diminished THAP11 mRNA (Figure 3.1A) and protein (Figure 3.1B) as determined by quantitative RT-PCR and immunoblotting, respectively. Of these four shRNA constructs, shT11D appeared slightly less effective and was not used in subsequent studies. Additionally, constructs shT11C and shT11E were found to have partially overlapping targeting sequences (Figure 3.1C); therefore we selected shRNAs shT11A, and shT11C for use in creating loss-of-function stable cell lines. Two independent pools of SW620 cells were transduced with either control (shNS, shNS2) or THAP11 (shT11A, shT11C) shRNA expressing retrovirus and analyzed for global gene expression changes using Nimblegen oligonucleotide microarrays. Genes displaying a 1.5 fold change and a p-value < 0.01 (t-Test) between control (shNS, shNS2) and THAP11 (shT11A, shT11C) shRNA expressing cells were defined as differentially expressed. This gene set was further processed to remove predicted genes which have been subsequently “discontinued” by the NCBI and additionally lack supporting mRNA or EST sequences. This analysis identified 80 transcripts (excluding THAP11) as differentially expressed between THAP11 and control knockdown groups (Figure 3.2A). Of these differentially expressed RNAs, 70% (56/80) showed increased expression with THAP11 knockdown, consistent with our previous observation suggesting THAP11 possesses transcriptional repressor activity. Quantitative RT-PCR using the same RNA also analyzed by microarray allowed for independent verification of microarray determined gene expression changes. As shown in Figure 3.2B, most of the microarray results were recapitulated by qRT-
PCR suggesting the dataset as a whole likely represents a biologically relevant, THAP11 dependent gene expression program.

Gene expression measurements determined by oligonucleotide microarray and quantitative RT-PCR reflect steady-state mRNA levels and by themselves are incapable of assessing whether regulation occurs at the transcriptional or post-transcriptional level. To determine if the gene expression changes observed in THAP11 knockdown cells were at least partially attributable to increased transcription, we performed a modified nuclear run-on assay in SW620 cells expressing either control or THAP11 shRNAs (Core, Waterfall et al. 2008). As illustrated in Figure 3.3A, run-on transcription from isolated nuclei was performed using BrUTP, in place of $^{32}$P-UTP, to label nascent RNA transcripts in the presence of 0.5% sarkosyl. Inclusion of sarkosyl in run-on reactions prevents re-initiation of transcription thus allowing only the completion of transcripts actively engaged by RNA Polymerase II at cell lysis (Core, Waterfall et al. 2008). BrU labeled nascent transcripts were then specifically immunoprecipitated from total nuclear RNA using a monoclonal anti-BrdU antibody, which recognizes BrU as well as BrdU, and transcript levels determined by quantitative RT-PCR. In THAP11 knockdown cells we found nascent transcript levels for Lsmd1, ncRNA00095, AA862256 and Annexin A1 up-regulated in a manner qualitatively similar to their steady-state mRNAs (Figure 3.3B and data not shown). Importantly, THAP11 steady-state but not nascent transcript levels were depleted in SW620 cells expressing THAP11 shRNAs (Figure 3.3B). This expected discrepancy is in agreement with the proposed mechanism of RNA interference.
as a post-transcriptional gene silencing event and provides an important verification of the specificity of the nuclear run-on assay to detect nascent rather than mature transcripts (Meister and Tuschl 2004). Taken together, we conclude that the increased RNA levels observed in THAP11 knockdown cells for Lsmd1, ncRNA00095, AA862256, Annexin A1 and, perhaps the microarray dataset as a whole, likely reflects increased transcription. However, we cannot exclude the possibility that post-transcriptional regulatory mechanisms may also contribute to the gene expression profile of THAP11 knockdown cells.

Transcripts induced by THAP11 knockdown may represent direct gene targets of THAP11 mediated repression. Alternatively, these gene expression changes may arise from secondary events subsequent to THAP11 knockdown or possibly from nonspecific shRNA events despite our use of multiple control and THAP11 targeted shRNAs. To discriminate between these possibilities, we performed rescue experiments using a THAP11 expression construct rendered non-silenceable by mutation of three consecutive codons in each of the shRNA targeting sequences (Figure 3.4A). SW620 cells were first transduced with either control or THAP11 shRNA, and then with either control or THAP11-Rescue-3xFLAG retroviruses. Cells expressing both constructs were selected by puromycin resistance (shRNA) and fluorescence activated cell sorting (THAP11 rescue). Immunoblotting with THAP11 antibody revealed a robust re-expression of THAP11 in cells expressing THAP11-Rescue-3xFLAG but not control virus (Figure 3.4B, upper panel). The 3xFLAG-tagged THAP11 rescue protein migrates at a slightly higher molecular weight allowing comparison of expression
levels between rescue and endogenous THAP11 (compare bands in lane 4) and verification that endogenous THAP11 remains depleted in THAP11-Rescue-3xFLAG expressing cells (lanes 5 and 6). The identity of the THAP11 bands in rescue expressing cells was further validated by immunoblotting with monoclonal anti-FLAG antibody (Figure 3.4B, lower panel).

If the gene expression profile observed in THAP11 knockdown cells is attributable to specific depletion of THAP11, than restoration of THAP11 status in THAP11-Rescue-3xFLAG cells should reverse this effect. Indeed, quantitative RT-PCR of several putative THAP11 gene targets including Lsmd1, ncRNA00095, AA862256, Praf2 and Alg14 revealed that expression of non-silenceable THAP11 prevented the differential gene expression previously observed in THAP11 knockdown cells (Figure 3.4C and Table 3.1). Importantly, this rescue effect was functional irrespective of the magnitude of putative target gene induction. Modestly induced genes (1.2-1.6 fold) such as Smarca1 and Atg4a and robustly induced genes (3-8 fold) including Lsmd1 and AA862256 were equally rescued by non-silenceable THAP11. The expression of THAP11-Rescue-3xFLAG re-repressed putative THAP11 gene targets below the level observed in cells expressing endogenous amounts of THAP11 (shNS and Empty), likely due to rescue construct over-expression. To determine if THAP11 knockdown de-represses these genes in additional colon cancer cell lines, we expressed THAP11 shRNAs in Colo320HSR cells which also have elevated THAP11 expression (Figure 2.6). As shown in Figure 3.5, knockdown of THAP11 in Colo320HSR cells de-represses putative THAP11 gene targets.
*Alg14, Lsmd1,* and AA862256 similar to knockdown in SW620 cells. Similar de-repression following THAP11 knockdown in Colo320HSR was also observed for *ncRNA00095* and AK021933 (data not shown). Taken together, these results indicate that the majority of gene expression changes observed in THAP11 knockdown cells likely reflects an authentic cellular response to diminished THAP11 protein levels and represents the first characterization of a THAP11 dependent gene expression network in human cells.

**Identification of Direct THAP11 Gene Targets in SW620 cells**

Direct THAP11 gene targets should be readily determined by identifying THAP11 at relevant genomic control regions, such as proximal promoters, by chromatin immunoprecipitation. However, the paucity of known endogenous THAP11 gene targets and the suitability of our custom THAP11 antibody for chromatin immunoprecipitation make this a difficult task. To circumvent these issues, we utilized our SW620 THAP11-Rescue-3xFLAG cells which have stably depleted endogenous THAP11 and over-expressed 3xFLAG tagged, non-silenceable THAP11. Since the non-silenceable THAP11 in these cells functionally rescues endogenous THAP11 knockdown, as assessed by re-repression of THAP11 shRNA induced gene expression, we reasoned that direct THAP11 gene targets should be identified by ChIP assay directed against the 3xFLAG epitope. THAP11-Rescue-3xFLAG cells were processed for ChIP assay using anti-FLAG monoclonal antibody and enrichment of immunoprecipitated chromatin was monitored by quantitative PCR using amplicons spaced approximately 500bp apart and spanning at least 2kb on either
side of the transcriptional start sites of *Lsmd1* and *ncRNA00095* (Figure 3.6A). As shown in Figure 3.6B, this approach identified a peak of THAP11-Rescue-3xFLAG binding in both putative THAP11 gene targets. At *Lsmd1*, THAP11-Rescue-3xFLAG binding peaks at an amplicon centered at +392 relative to the annotated transcriptional start site while at *ncRNA00095* two amplicons centered at -212 and +335 show apparently equivalent enrichment. No enrichment was observed when ChIP was performed using a species and isotype matched control IgG in SW620 rescue cells or with FLAG antibody in SW620 control cells (data not shown) confirming the specificity of the ChIP assay. We subsequently extended this approach to additional putative THAP11 gene targets including *Praf2*, AA862256, *C1orf83*, and *ZSCAN20* (Figures 3.7 and 3.8). Interestingly, each examined target gene displayed a peak of THAP11-3xFLAG binding within 500 base pairs of the annotated transcriptional start site. Because each of the aforementioned genes were effectively re-repressed by and contain chromatin bound THAP11-Rescue-3xFLAG near their transcriptional start sites, we conclude that these genes are likely direct targets of THAP11 mediated transcriptional repression.

Having identified direct THAP11 gene targets using chromatin immunoprecipitation in conjunction with THAP11-Rescue-3xFLAG, we next confirmed that these regions were indeed bound by endogenous THAP11. Normal SW620 cells were analyzed by chromatin immunoprecipitation using anti-THAP11 antibody at the previously identified THAP11 bound or distal genomic regions of *Lsmd1*, *ncRNA00095*, and AA862256. Chromatin bound, endogenous
THAP11 was identified at the same regions previously found to contain THAP11-3xFLAG-Rescue. Amplicons centered at Lsmd1 +392, ncRNA00095 -212, and AA862256 +101 (Figure 3.9A) were enriched relative to normal rabbit IgG while amplicons located at least 1kb upstream were not. To further confirm the specificity of the anti-THAP11 ChIP assay, we repeated the experiment in THAP11 knockdown SW620 cells. As expected, cells expressing THAP11 shRNA exhibited reduced, albeit detectable levels of chromatin bound THAP11 at Lsmd1 +392 (Figure 3.9B). Taken together, these data indicate that an experimental approach consisting of knockdown of endogenous chromatin bound proteins in conjunction with epitope-tagged, non-silenceable rescue expression is a viable strategy for interrogating the chromatin occupancy profile of proteins intractable to conventional chromatin immunoprecipitation. More importantly, these findings provide the necessary orthogonal data to conclude that Lsmd1, ncRNA00095, AA862256, C1orf83, Praf2, and ZSCAN20 are direct targets of THAP11 mediated transcriptional repression.

**RNAPII and Acetylated-Histone Occupancy at THAP11 Repressed Genes**

To begin determining the mechanism associated with THAP11 mediated transcriptional repression, we examined RNAPII occupancy at THAP11 target genes Lsmd1 and ncRNA00095. Chromatin immunoprecipitation using an antibody that recognizes the hypophosphorylated C-terminal domain (CTD) of RNAPII revealed increased RNAPII binding at the TSS of both genes concomitant with THAP11 knockdown (Figure 3.10). The elevated RNAPII occupancy observed at Lsmd1 in THAP11 knockdown cells was reversed in
SW620 rescue cells that simultaneously express non-silenceable THAP11 thus confirming that increased RNAPII occupancy reflects diminished THAP11 levels (Figure 3.11). Taken together, these results suggest that THAP11 may repress transcription by limiting or destabilizing RNAPII at THAP11 target genes.

Data presented in chapter 2 of this thesis demonstrate that THAP11 physically associates with histone deacetylases (HDACs). To determine if recruitment of HDACs by THAP11 may account for THAP11 dependent transcriptional repression, we examined the histone acetylation status of Lsmd1 and ncRNA00095 by ChIP assay in THAP11 knockdown cells. Increased acetyl-K9/K18-H3 (AcH3) was observed in the promoter of Lsmd1 but not ncRNA00095 in THAP11 knockdown SW620 cells (Figure 3.12), suggesting histone deacetylation is not a universal mechanism of THAP11 mediated transcriptional repression. However, it is also possible that THAP11 may recruit HDACs and regulate transcription by deacetylation of non-histone proteins. However, several attempts to chromatin immunoprecipitate HDACs at THAP11 bound chromatin were unsuccessful (data not shown), which may reflect either a technical limitation of the ChIP assay or a limited role for HDACs in repression by THAP11.

**THAP11 associates with HCF-1**

In mouse embryonic stem cells, THAP11 has recently been shown to repress transcription and associate in a large, multimeric complex containing the transcriptional co-regulator host cell factor 1 (HCF-1) (Dejosez, Krumenacker et al. 2008). While the role of HCF-1 in THAP11 mediated transcriptional repression was not directly demonstrated, a recent report by Mazars et al.
showed that HCF-1 is an essential effector of THAP1 mediated transcriptional activation of *RRM1* in human vascular endothelial cells (Mazars, Gonzalez-de-Peredo et al. 2010). Multiple THAP proteins, including THAP11, were also found to interact with HCF-1 by yeast two-hybrid assay (Mazars, Gonzalez-de-Peredo et al. 2010). HCF-1 associating THAP proteins contain the HCF-1 binding motif (HBM), a conserved tetrapeptide sequence (D/EHxY) known to mediate the interaction of various cellular and viral proteins with the N-terminal kelch domain of HCF-1 (Freiman and Herr 1997; Lu, Yang et al. 1997). These observations inspired us to ask whether THAP11 associates with HCF-1 in SW620 cells and if so, if the interaction contributes to THAP11 mediated repression. To address this question, we immunoprecipitated HCF-1 from SW620 nuclear extract and probed the immunoprecipitate for endogenous THAP11 using our custom THAP11 antibody. As shown in Figure 3.13A, endogenous THAP11 was found to specifically co-precipitate with endogenous HCF-1. This observation was also extended to additional colon cancer cell lines where THAP11 was found to co-precipitate with HCF-1 from nuclear extracts prepared from Colo320HSR, SW480, and HCT-116 cells (Figure 3.13B). To determine if the THAP11/HCF-1 interaction was dependent on DNA present in the nuclear extract preparation, we repeated the co-immunoprecipitation experiment in the presence of ethidium bromide, a technique frequently used to assess the DNA dependence of nuclear protein interactions (Lai and Herr 1992). As shown in Figure 3.13C, HCF-1 immunoprecipitation from SW620 nuclear extract in the presence of ethidium
bromide did not diminish the amount of co-precipitated THAP11, suggesting that THAP11 and HCF-1 can stably associate in the absence of DNA binding.

**THAP11 recruits HCF-1 to promoters of THAP11 repressed genes**

Because endogenous THAP11 and HCF-1 physically associate in living cells, we next wished to determine if HCF-1 is also recruited to promoter regions of THAP11 target genes in SW620 cells. Chromatin immunoprecipitation was performed for endogenous HCF-1 and THAP11 in SW620 cells and the occupancy profile of each was determined at the THAP11 target gene *Lsmd1*. As shown in Figure 3.14A-C, both HCF-1 and THAP11 displayed a strikingly similar distribution at *Lsmd1*. ChIP assay also detected HCF-1 at THAP11 target genes *C1orf83, ncRNA00095, AA862256* and *ZSCAN20* in SW620 cells (Figure 3.14D) suggesting HCF-1 is recruited to THAP11 bound and repressed promoters. HCF-1 is not known to possess intrinsic DNA binding activity and its association with chromatin is thought to require interaction with sequence specific DNA binding factors (Wysocka, Reilly et al. 2001). Accordingly, sequence specific binding by THAP11 at a targeted genomic region may result in HCF-1 recruitment by virtue of their physical interaction. To test this directly, we repeated the ChIP assay for HCF-1 in SW620 cells depleted of endogenous THAP11 by shRNA. As shown in Figure 3.15, SW620 THAP11 knockdown cells show a marked reduction in HCF-1 occupancy at *Lsmd1* (Figure 3.15A) and *ncRNA00095* (Figure 3.15B) suggesting THAP11 is the chromatin bound factor that recruits HCF-1 to these target promoters. Importantly, HCF-1 protein levels were unaltered in THAP11 knockdown cells (Figure 3.15C) indicating that the
differential recruitment of HCF-1 observed at THAP11 target promoters does not result from a reduction in total HCF-1 protein.

**HCF-1 is required for THAP11 mediated repression**

Having established that THAP11 physically associates with and recruits HCF-1 to THAP11 repressed promoters, we next asked whether HCF-1 plays a functional role in THAP11 mediated repression. Knockdown of endogenous HCF-1 using retrovirally delivered shRNAs produced a marked depletion of HCF-1 protein while THAP11 protein levels were unchanged (Figure 3.16A). Quantitative RT-PCR analysis of THAP11 target genes *Lsmd1* and *ncRNA00095* in HCF-1 knockdown cells revealed increased steady-state mRNA levels comparable to that observed by THAP11 knockdown (Figure 3.16B). The extent of *Lsmd1* and *ncRNA00095* transcript induction was found to be proportional to level of HCF-1 knockdown. As expected, ChIP analysis of HCF-1 in shHCF-1 expressing cells showed a strong reduction in HCF-1 occupancy at *Lsmd1* (Figure 3.17) and *ncRNA00095* (data not shown) promoters. Importantly, examination of RNAPII by chromatin immunoprecipitation showed increased occupancy at *Lsmd1* (Figure 3.17) and *ncRNA00095* (data not shown) promoters in HCF-1 knockdown cells suggesting HCF-1 is a critical effector of THAP11 mediated repression.

**Affinity purification of THAP11/HCF-1 protein complex**

HCF-1 is believed to function as a molecular scaffold, affecting transcription by linking sequence specific DNA binding factors to chromatin
modifying activities, and has been identified in multiprotein complexes of opposing transcriptional activities (Wysocka, Myers et al. 2003; Liang, Vogel et al. 2009; Chikanishi, Fujiki et al. 2010). As a transcriptional co-activator, HCF-1 has been found associated with Set1/MLL histone H3 lysine 4 methyltransferase while transcriptional co-repressor function has revealed HCF-1 associated with SIN3/HDAC activity (Wysocka, Myers et al. 2003; Tyagi, Chabes et al. 2007). In addition, HCF-1 is also frequently found associated with O-linked N-acetylglucosamine (O-GlcNAc) transferase (OGT) which has been alternately linked to transcriptional activation or repression (Wysocka, Myers et al. 2003; Fujiki, Chikanishi et al. 2009; Ozcan, Andrali et al. 2010). To identify the constituents of the THAP11/HCF-1 repression complex we performed a two-step affinity purification as outlined in Figure 3.18A using SW620 cells stably expressing THAP11-3xFLAG. Since HCF-1 itself is O-GlcNAcylated, HCF-1 and associated proteins can be partially purified from nuclear extracts by lectin affinity chromatography using wheat germ agglutinin (WGA) conjugated agarose (Wysocka, Myers et al. 2003). WGA-agarose bound HCF-1 complexes were eluted with GlcNAc and subsequently immunoprecipitated with anti-FLAG. As shown in Figure 3.18B, the majority of endogenous and 3xFLAG-tagged THAP11 were bound by WGA-agarose (compare Input and WGA-FT lanes). WGA-eluates immunoprecipitated with FLAG-agarose recovered HCF-1 only from THAP11-3xFLAG and not empty vector expressing cells (Figure 3.18B). The relatively small proportion of total cellular HCF-1 recovered (compare FLAG-agarose and Input lanes, Figure 3.18B) is consistent with the idea that HCF-1
functions in many discrete complexes (Wysocka, Reilly et al. 2001). WGA and FLAG immunoprecipitates were then analyzed by immunoblot for the presence of known HCF-1 (OGT, HDAC1, HDAC2, mSIN3A, Lsd1) or THAP11 (HDAC1, HDAC3) interacting proteins consistent with transcriptional repression. The results of the two-step affinity purification revealed HCF-1, OGT, and HDAC1 associated with THAP11-3xFLAG following WGA purification (Figure 3.18C). Lsd1 was not detected in the WGA enriched material suggesting the association of Lsd1 with HCF-1 in SW620 cells does not occur to an appreciable degree or does not survive the purification (data not shown). Next, the co-localization of OGT with THAP11 and HCF-1 on chromatin was determined by ChIP assay. OGT was found at a subset of genomic regions similarly as THAP11 and HCF-1 including C1orf83, Lsmd1, and ncRNA00095 but not AA862256 and ZSCAN20 (Figure 3.19). Taken together, these data suggest that THAP11 recruits a multimeric complex composed of HCF-1/OGT/HDAC1 to mediate transcriptional repression.

**Annexin A1 is de-repressed by THAP11 knockdown**

Many of the THAP11 dependent differentially expressed transcripts we have examined either lack protein coding capacity (ncRNA00095, AA862256, AK021933) or encode uncharacterized proteins without commercially available antibodies (Lsmd1, Alg14). In contrast, Annexin A1 is a well-characterized member of the annexin superfamily of phospholipid binding proteins with documented roles in inflammation, cell proliferation and apoptosis (Lim and Pervaiz 2007; Tabe, Jin et al. 2007; Blume, Soeroes et al. 2009; Perretti and
D’Acquisto 2009; Zhang, Huang et al. 2010). Immunoblotting for Annexin A1 in SW620 THAP11 knockdown cells revealed increased Annexin A1 protein (Figure 3.20A) consistent with elevated Annexin A1 mRNA levels detected in those same cells (Figure 3.20B). Interestingly, we observe that knockdown of THAP11 in SW620 cells restores both Annexin A1 mRNA and protein to levels expressed in SW480 cells (Figure 3.20). Consistent with transcriptional de-repression of Annexin A1 upon THAP11 knockdown, we observed increased RNAPII and acetyl-histone H3 occupancy at the Annexin A1 proximal promoter by ChIP assay (data not shown). This data suggests that a gain in THAP11 function in SW620 cells may be at least partially responsible for the loss of Annexin A1 in these cells. Congruent with this hypothesis, we find that Annexin A1 protein levels inversely correlate with THAP11 protein levels in a panel of colon cancer cell lines as shown in Figure 3.21. However, we failed to detect direct binding of THAP11/HCF-1/OGT/HDAC1 on the Annexin A1 proximal promoter in SW620 cells suggesting THAP11 may indirectly regulate Annexin A1 gene expression (data not shown).

Cell growth suppression in THAP11 knockdown SW620 cells

During routine subculturing of THAP11 knockdown cells we noticed that cells expressing THAP11 shRNAs grew more slowly than those expressing control shRNA. To explore the possibility that THAP11 knockdown may affect cellular growth, we performed the Alamar Blue cell enumeration assay which detects the metabolic conversion of non-fluorescent resazurin to fluorescent resorufin in viable cells. SW620 cells were transduced with control (shNS) or
THAP11 shRNA (shT11A or shT11C) retrovirus and two days post-transduction selected with puromycin for an additional two days and then seeded into 96-well plates for the Alamar Blue assay. As shown in Figure 3.22, THAP11 knockdown resulted in a decrease in the number of viable cells over time with shRNA T11C slightly more effective than shRNA T11A. Decreased cell proliferation, colony formation and enhanced apoptosis have been reported in prostate cancer cell lines following restoration of Annexin A1 expression (Hsiang, Tunoda et al. 2006). Therefore, we next addressed if elevated Annexin A1 expression resulting from THAP11 knockdown may similarly contribute to the cell growth defect observed in those cells. To test this possibility, double-knockdown SW620 cells expressing control, THAP11, or Annexin A1 shRNAs were established. Cells were first transduced with control or THAP11 knockdown virus (pSRP) and 24 hours later transduced with control or Annexin A1 virus (pSRNG). Double-knockdown cells were selected by puromycin and Geneticin resistance and subjected to the Alamar Blue assay as described above. As shown in Figure 3.23A, knockdown of Annexin A1 in cells also depleted of THAP11 restores Annexin A1 expression to levels observed in control knockdown cells. However, Annexin A1/THAP11 double knockdown cells did not exhibit growth kinetics different from THAP11 knockdown alone (Figure 3.23B), suggesting that increased Annexin A1 is dispensable for the growth defect conferred by THAP11 knockdown in SW620 cells.
Discussion

Evidence of the transcriptional regulatory properties of THAP proteins has been previously demonstrated using luciferase reporter constructs regulated by Gal4 DNA-binding domain THAP protein fusions. However, very little is known regarding the identity of THAP protein gene targets, the mechanisms by which THAP proteins regulate transcription, or the cellular consequence of THAP protein function. The work presented within this chapter provides substantial evidence to suggest that endogenous THAP11 functions as a transcriptional repressor in living cells. An RNA interference based loss-of-function cell culture model was used to investigate the transcriptional regulatory properties of THAP11 in SW620 metastatic colon cancer cells. Using this model, a THAP11 dependent gene expression program was identified and found to contain both protein coding and non-coding genes. De-repressed genes were shown to have chromatin bound THAP11 near their transcription start sites suggesting direct transcriptional repression. THAP11 was found to recruit Host cell factor 1 (HCF-1) to repressed genes and depletion of chromatin bound THAP11 and/or HCF-1 resulted in increased transcription concomitant with elevated RNAPII occupancy. In addition, knockdown of endogenous THAP11 resulted in decreased cell proliferation suggesting THAP11 mediated transcriptional repression may contribute to cell growth regulation during colon cancer progression.

Loss-of-function cell culture models using RNA interference provide accessible biological systems to evaluate the function of novel transcription factors. We employed this strategy to identify a THAP11 dependent gene
expression program in SW620 metastatic colon cancer cells which exhibit robust expression of THAP11. Multiple THAP11 targeting and control short-hairpin RNAs were used to identify a core set of THAP11 responsive genes, the majority of which were de-repressed upon THAP11 knockdown. Additional validation was provided by expression of a non-silenceable THAP11 construct which rescued the gene expression profile induced by both THAP11 knockdown constructs. The functional redundancy of this approach was necessary for several reasons. First, gene expression profiling using RNA interference frequently results in off-target effects. Multiple shRNAs and non-silenceable rescue expression increase the likelihood of finding authentic THAP11 gene targets. Second, at the time of this undertaking no endogenous THAP11 regulated genes had been reported. While this work was in progress, an independent report identified c-myc as a directly repressed gene target of THAP11 in human HepG2 liver cancer cells (Zhu, Li et al. 2009). Similarly, characterization of THAP11 in mouse embryonic stem cells identified GATA4 and GATA6 as direct gene targets of THAP11 repression (Dejosez, Krumenacker et al. 2008). We find that knockdown of THAP11 in SW620 did not de-repress c-myc, GATA4 or GATA6 and suggest cell type specific mechanisms may be responsible. Importantly, several of the THAP11 gene targets identified here (Lsmd1, ncRNA00095, AA862256) were found to be induced by THAP11 knockdown in multiple human colon cancer cell lines and these genes may provide more tractable models to investigate the mechanism of THAP11 mediated repression.
Knockdown of THAP11 resulted in a moderate but reproducible decrease in cell proliferation. We speculated that increased Annexin A1 expression may explain the cell proliferation defect observed in THAP11 knockdown cells. The role of Annexin A1 in cancer cell function is complex but several reports have suggested Annexin A1 can function in an anti-proliferative capacity. Restoration of Annexin A1 in prostate cancer cell lines diminishes cell proliferation and sensitizes cells to apoptosis inducing agents (Hsiang, Tunoda et al. 2006). Similarly, HDAC inhibitor mediated apoptosis in Kasumi-1 leukemia cells is dependent on upregulation of endogenous Annexin A1 (Tabe, Jin et al. 2007). Of particular interest is the finding that Annexin A1 induction correlates with the differentiation status of colon cancer cell lines following butyrate treatment (Guzman-Aranguez, Olmo et al. 2005; Lecona, Barrasa et al. 2008). However, simultaneous knockdown of Annexin A1 in THAP11 knockdown SW620 cells failed to rescue the THAP11 knockdown induced cell proliferation defect. Furthermore, constitutive overexpression of Annexin A1 in normal SW620 also failed to phenocopy the reduced cell proliferation associated with THAP11 knockdown (data not shown). Since increased Annexin A1 expression in SW620 cells is neither necessary nor sufficient to perturb cell proliferation, an additional Annexin A1 independent mechanism must be at work. Additional experiments will be required to determine if altered cell cycle kinetics or cell attrition due to apoptosis explain the THAP11 dependent cell proliferation defect. It should be noted that while this data likely precludes Annexin A1 from a role in THAP11 dependent colon cancer cell proliferation, additional biological outcomes resulting
from THAP11 directed repression of Annexin A1 may be identified. For instance, cell surface externalized Annexin A1 functions as an “eat me” signal to promote phagocytosis of apoptotic cells while in necrotic cells Annexin A1 has been shown to be an important mechanism to prevent proinflammatory cytokine production following phagocytosis by macrophages (Tabe, Jin et al. 2007; Blume, Soeroes et al. 2009). An interesting but currently unexplored corollary suggests that Annexin A1 deficient apoptotic/necrotic tumor cells may trigger proinflammatory cytokine production following phagocytosis by tumor associated macrophages thereby promoting a microenvironment favorable for tumor formation and metastasis dissemination. This may be especially relevant in colon cancer metastasis to the liver where Kupffer cells (liver macrophages) line the hepatic sinusoids and provide the first barrier to intravasation and colonization by circulating tumor cells (Bayon, Izquierdo et al. 1996; Timmers, Vekemans et al. 2004).

Amongst the protein coding genes identified as directly repressed by THAP11 is Lsmd1. Lsmd1 (also referred to as hMak31p and hNAA38p) is a recently identified subunit of the Nα-terminal acetyltransferase complex C (NatC) (Starheim, Gromyko et al. 2009). Three Nα-terminal acetyltransferase complexes, NatA, NatB and NatC have been identified and partially characterized in human cells. Each NAT complex displays specificity for particular N-terminal peptide sequences but collectively account for the vast majority of protein Nα-terminal acetylation in eukaryotic cells (Starheim, Gromyko et al. 2009). More than 80% of human proteins contain Nα-terminal acetylation
and several studies have linked this modification with cell proliferation, survival, differentiation, and cancer (Starheim, Gromyko et al. 2009).

Repression of \textit{Lsmd1} resulting from increased THAP11 expression during colon cancer progression may decrease NatC function and N\(\alpha\)-terminal acetylation of NatC substrates. Only two NatC substrates have been identified to date; in human cells the Arf-like GTPase hArl8b and in zebrafish the target of rapamycin (TOR) (Wenzlau, Garl et al. 2006; Starheim, Gromyko et al. 2009). Experimental models of loss of NatC function have demonstrated mislocalization of hArl8b in HeLa cells and loss of TOR expression and signaling in zebrafish, thereby underscoring the relevance of N\(\alpha\)-terminal acetylation in proper biological function of NatC substrates (Wenzlau, Garl et al. 2006; Starheim, Gromyko et al. 2009). In addition, RNA interference mediated depletion of Lsmd1 or additional NatC subunits resulted in cell growth arrest and apoptosis in a p53 dependent manner. Knockdown of NatC subunits in HeLa or HCT-116 cells decreased Mdm2 protein levels resulting in stabilized p53 and increased transcription of pro-apoptotic p53 target genes (Starheim, Gromyko et al. 2009). Importantly, HCT-116 p53\(^{-/-}\) cells were unaffected by NatC complex knockdown confirming the role of functional p53 in the apoptotic response to loss of NatC function (Starheim, Gromyko et al. 2009). This apparent p53 requirement may explain the paradoxical finding that gain of THAP11 expression can repress \textit{Lsmd1} expression without triggering an apoptotic response. Human colon cancer cell lines expressing the highest THAP11 protein amounts (SW620, Colo320HSR) also contain gain-of-function p53 mutations which effectively
divorce p53 from the apoptotic transcriptional program (Brosh and Rotter 2009). Rather, gain-of-function mutant p53^{R273H/P309S}, basally expressed at high levels in SW480 and SW620 cells, has been shown to confer increased tumorigenic properties to SW480 cells (Yan and Chen 2009). Thus, gain-of-function mutation in p53 may provide a favorable genetic environment for increased THAP11 expression and concomitant reduction in Lsmd1.

In addition to protein coding genes, several annotated or putative long non-coding RNAs were also identified as direct targets of THAP11 mediated transcriptional repression including ncRNA00095 and AA862256. The fortuitous discovery of these ncRNAs as THAP11 regulated transcripts by microarray based gene expression profiling likely reflects their previous but erroneous annotation as protein coding genes with subsequent inclusion in the microarray design. Many long ncRNAs have been discovered to be regulated by the same transcriptional control mechanisms also utilized by protein coding mRNAs including recruitment of RNAPII, regulation by positive and negative trans-acting factors, and deposition of histone modifications associated with active promoters and RNA splicing (Guttman, Amit et al. 2009). Accordingly, we speculate that additional long ncRNAs are likely regulated by THAP11. Recent advances in massively parallel sequencing technology should greatly facilitate the complete annotation of both coding and non-coding THAP11 regulated RNAs.

Long ncRNAs are now well established to contribute to a diverse array of biological functions including transcriptional regulation, cell growth and apoptosis (Taft, Pang et al. 2010). However the vast majority of long ncRNAs are
uncharacterized. It is tempting to speculate that THAP11 mediated regulation of non-coding RNAs, including ncRNA00095 and AA862256, plays a critical role in cancer cell growth. Future studies will be necessary to identify the biological function of THAP11 repressed non-coding RNAs ncRNA00095 and AA862256. Knockdown of these ncRNAs in conjunction with THAP11 knockdown should reveal if their increased expression is necessary for the cell proliferation defect observed in THAP11 knockdown cells.

THAP11 was found to physically associate with HCF-1 in various colon cancer cell lines and recruit HCF-1 to repressed gene promoters. Since HCF-1 lacks intrinsic DNA binding capacity, THAP11 likely functions as the sequence specific determinant responsible for HCF-1 recruitment to chromatin. In support of this assertion, we find that the chromatin occupancy profile of HCF-1 closely parallels that observed for THAP11 and knockdown of THAP11 releases HCF-1 from chromatin. Loss of HCF-1 from THAP11 repressed promoters either by diminished recruitment resulting from THAP11 knockdown or reduction in total cellular HCF-1 arising from HCF-1 knockdown resulted in increased RNAPII occupancy coincident with increased transcription. This indicates an essential role for HCF-1 in THAP11 mediated repression. In agreement with these results, a previous biochemical purification of THAP11 from mouse embryonic stem cells also identified THAP11 in complex with HCF-1 but the contribution of HCF-1 towards THAP11 dependent transcriptional repression was not established (Dejosez, Krumenacker et al. 2008).
While the association with HCF-1 may be a general feature of THAP11 mediated repression, additional cell type specific mechanisms may also occur. For instance, loss of THAP11 from GATA4 and GATA6 promoters during ES cell differentiation correlated with a decrease in dimethylation of H3K9 at those promoters coincident with transcriptional activation (Dejosez, Krumenacker et al. 2008). In contrast, we find no detectable H3K9 dimethylation at THAP11 targets in SW620 cells (data not shown).

This specificity may be determined by THAP11 or alternatively may be dictated by HCF-1. HCF-1 is synthesized as a 230-kDa precursor which undergoes autocatalytic proteolysis at a series of 20-amino acid repeats present in the central domain of the protein (Vogel and Kristie 2000). The resulting proteolytic fragments remain tightly associated forming mature HCF-1; a heterodimeric complex of non-covalently associated amino- and carboxy-terminal subunits (Vogel and Kristie 2000). These site-specific processing events have been previously shown to influence both the protein interaction and transcriptional regulatory properties of HCF-1 and may account for the mechanistic differences observed in THAP11 mediated repression between mouse ES and human colon cancer cell lines (Vogel and Kristie 2006).

In the context of transcriptional repression, HCF-1 is known to associate with the SIN3/HDAC histone deacetylase complex and results presented in chapter 2 of this thesis demonstrate THAP11 can also associate with HDACs. Consistent with that data, a small but detectable amount of HDAC1 was found present in the THAP11/HCF-1 complex. However, de-repression associated with
release of THAP11 results in increased histone acetylation only at \textit{Lsmd1} and not at \textit{ncRNA00095}. This suggests that either THAP11 or HCF-1 tethers an additional, HDAC independent transcriptional repressor to THAP11 bound chromatin. An intriguing candidate to mediate the HDAC independent component of THAP11/HCF-1 repression is O-linked glycosyltransferase (OGT) which catalyzes the reversible addition of a single N-acetylglucosamine in O-glycosidic linkage (O-GlcNAc) to serine or threonine residues. OGT has previously been demonstrated to account for the HDAC independent component of SIN3 mediated transcriptional repression and has recently been shown to be necessary for Polycomb repression of \textit{Hox} genes during development in \textit{Drosophila} (Yang, Zhang et al. 2002; Gambetta, Oktaba et al. 2009; Sinclair, Syrzycka et al. 2009). We find OGT stably associated with the THAP11/HCF-1 complex and bound to THAP11 repressed promoters. Interestingly, SIN3 was not found associated with the THAP11/HCF-1 complex but because OGT is known to co-purify with HCF-1 this may obviate the requirement for SIN3.

Numerous transcription factors including Sp1 and RNAPII have been found to be O-GlcNAcylated with corresponding alterations in their transcriptional regulatory properties (Comer and Hart 2001; Ozcan, Andrali et al. 2010). Modification of Sp1 activation domains with O-GlcNAc inhibits Sp1 mediated transcriptional activation by disrupting the hydrophobic interactions between Sp1 and members of the transcriptional activation complex including the TFIID subunit TAF4 (Yang, Su et al. 2001). Similarly, O-GlcNAcylation of RNAPII at serine residues in its C-terminal domain (CTD) is thought to negatively regulate
transcriptional elongation by preventing CTD phosphorylation which is essential for the transition from transcription initiation to elongation (Comer and Hart 2001; Yang, Zhang et al. 2002). Each of the direct THAP11 gene targets examined here (Lsmd1, ncRNA00095, AA862256, Praf2, C1orf83, ZSCAN20) resides within a CpG island and accordingly contain several putative Sp1 binding GC-boxes. Chromatin immunoprecipitation experiments find RNAPII and Sp1 (data not shown) enriched at THAP11 bound regions. Thus, an intriguing possibility is that chromatin bound THAP11 may recruit an HCF-1/OGT complex resulting in O-GlcNAcylation of Sp1 and/or RNAPII CTD. The resulting transcriptional repression may arise from either decreased transcription initiation, elongation or a combination of the two. Consistent with this hypothesis we find that knockdown of THAP11 increases the occupancy of RNAPII at THAP11 repressed genes. Ongoing studies are actively engaged in determining the role of OGT and O-GlcNAc modification in THAP11 mediated transcriptional repression.

HCF-1 was initially identified as a transcriptional co-activator necessary for the expression of immediate early genes following α-herpesvirus infection and has also been shown to associate with several DNA binding proteins including both activating and repressing E2F family members (Tyagi, Chabes et al. 2007). With respect to E2F proteins, HCF-1 plays a directive role, tethering the appropriate transcriptional regulatory complex with the corresponding activating (E2F1) or repressing (E2F4) E2F protein (Tyagi, Chabes et al. 2007). When bound to E2F1, HCF-1 functions as a transcriptional co-activator by recruiting
histone H3 lysine 4 methyltransferase activities (Tyagi, Chabes et al. 2007). In association with E2F4 however, HCF-1 mediates transcriptional repression by recruiting the SIN3/HDAC complex (Tyagi, Chabes et al. 2007).

The role of HCF-1 in THAP protein regulated transcription appears to parallel the transcriptional dichotomy observed with E2F proteins. As described above, HCF-1 is an integral component of THAP11 mediated repression. In contrast, HCF-1 has recently been shown to be necessary for THAP1 and THAP3 mediated transcriptional activation (Mazars, Gonzalez-de-Peredo et al. 2010). Loss of HCF-1 diminished THAP1 activation of the RRM1 gene in human primary vascular endothelial cells (Mazars, Gonzalez-de-Peredo et al. 2010). Similarly, THAP3 tethered to the minimal promoter of an integrated luciferase reporter construct as a Gal4-DNA binding domain fusion protein was shown to positively regulate transcription in an HCF-1 dependent manner (Mazars, Gonzalez-de-Peredo et al. 2010). Thus, like E2F proteins, THAP domain proteins appear to utilize HCF-1 to direct opposing transcriptional outcomes.

The majority of human THAP proteins (THAP1-7, THAP9, and THAP11) have been shown to possess HCF-1 binding motifs and interact with HCF-1 by yeast two-hybrid assay (Mazars, Gonzalez-de-Peredo et al. 2010). Previously published results from our laboratory and presented within chapter 2 of this thesis has demonstrated that several of these putative HCF-1 binding THAP proteins including THAP7 and THAP4 repress basal transcription as Gal4-DNA binding domain fusion proteins. Future studies should determine if HCF-1 plays a similar role in regulating the transcriptional repressor functions of those THAP
proteins. In addition, subsequent work should also be performed to determine the mechanistic basis for THAP protein/HCF-1 complex mediated transcriptional activation versus repression.
Figure 3.1: THAP11 knockdown by retrovirally delivered shRNA in SW620 cells. (A) SW620 cells were transduced with retrovirus expressing the indicated shRNA and selected with puromycin as described in Experimental Procedures. Whole cell extracts were prepared four days post-transduction and immunoblotted for THAP11 and β-actin. (B) Quantitative RT-PCR analysis of THAP11 mRNA in cells expressing the indicated shRNA as in (A). RNA levels are normalized to β-actin and expressed as percentage of THAP11 mRNA relative to non-specific shRNA (shNS). Values represent the mean ± standard deviation of triplicate quantitative RT-PCR reactions from a representative experiment. Experiments in (A) and (B) were performed three times with similar results. (C) Schematic depicting the position of the indicated shRNA targeting sequences (black rectangles) on the intronless THAP11 gene. THAP11 coding sequence is represented by the thick blue rectangle while untranslated sequence is represented by thin blue rectangles.
Figure 3.2: Gene expression analysis in THAP11 knockdown SW620 cells. (A) Heat map depicting microarray-based gene expression profiling in SW620 cells expressing either control (shNS and shNS2) or THAP11 (shT11A and shT11C) shRNAs. (B) Validation of microarray determined gene expression changes by qRT-PCR. Values represent mean ± standard deviation of triplicate quantitative RT-PCR reactions of a representative experiment performed at least three times.
Figure 3.3: Nuclear run-on assay in THAP11 knockdown SW620 cells. (A) Schematic of modified nuclear run-on assay. Adapted from (Core, Waterfall et al. 2008). (B) Steady-state and nascent mRNA levels of putative THAP11 gene targets in THAP11 knockdown SW620 cells. Steady-state mRNA determinations of the indicated genes was performed by qRT-PCR from total RNA prepared from cells expressing either control (shNS) or THAP11 (shT11C) shRNA. Nascent mRNA levels were analyzed from identical cell populations as steady-state determinations using a modified nuclear run-on assay as described in Experimental Procedures. Steady-state and nascent mRNA levels were normalized to β-actin and expressed as fold change relative to SW620 expressing control (shNS) shRNA. Values represent mean ± standard deviation of triplicate quantitative RT-PCR reactions of a representative experiment performed at least twice.
Figure 3.4: Expression of non-silenceable THAP11 reverses gene expression induced by THAP11 knockdown.

(A) Schematic of the site-directed mutagenesis strategy used to create a non-silenceable THAP11 rescue construct.

Silent mutations (in red) were introduced into three consecutive codons of a THAP11-3xFLAG expression construct at targeting sequences corresponding to shRNAs shT11A and shT11C. The encoded amino acids are indicated below their respective codons. (B) Immunoblot analysis of SW620 cells expressing either control (Empty) or THAP11-Rescue-3xFLAG (Rescue) and the indicated shRNA. Whole cell extracts were first immunoblotted with anti-THAP11 antibody to detect endogenous (red arrow) or rescue (black arrow) THAP11. Blots were stripped and reprobed with anti-FLAG and anti-β-actin. (C) Quantitative RT-PCR analysis of THAP11 and Lsmd1 expression in SW620 cells expressing either control (Empty) or THAP11-Rescue-3xFLAG (Rescue) and the indicated shRNA. RNA levels are normalized to β-actin and expressed as fold change relative to SW620 expressing control shRNA (shNS) and rescue (Empty) constructs. Values represent mean ± standard deviation of triplicate quantitative RT-PCR reactions. (B) and (C) are representative experiments of at least three separate experiments with similar results.
Table 3.1: Summary of genes re-repressed by THAP11 rescue.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Empty</th>
<th></th>
<th></th>
<th>THAP11-3xFLAG-Rescue</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>shNS</td>
<td>shT11A</td>
<td>shT11C</td>
<td>shNS</td>
<td>shT11A</td>
<td>shT11C</td>
</tr>
<tr>
<td>LSMD1</td>
<td>1.00 ± 0.16</td>
<td>3.36 ± 0.04</td>
<td>4.30 ± 0.35</td>
<td>0.57 ± 0.05</td>
<td>0.69 ± 0.01</td>
<td>0.92 ± 0.10</td>
</tr>
<tr>
<td>AA862256</td>
<td>1.00 ± 0.11</td>
<td>6.04 ± 0.21</td>
<td>8.11 ± 0.98</td>
<td>0.75 ± 0.04</td>
<td>0.77 ± 0.04</td>
<td>0.99 ± 0.20</td>
</tr>
<tr>
<td>Praf2</td>
<td>1.00 ± 0.00</td>
<td>1.85 ± 0.05</td>
<td>2.42 ± 0.28</td>
<td>0.54 ± 0.06</td>
<td>0.44 ± 0.02</td>
<td>0.69 ± 0.06</td>
</tr>
<tr>
<td>Alg14</td>
<td>1.00 ± 0.64</td>
<td>4.10 ± 0.12</td>
<td>4.34 ± 0.44</td>
<td>0.82 ± 0.07</td>
<td>0.82 ± 0.06</td>
<td>0.99 ± 0.15</td>
</tr>
<tr>
<td>ncRNA00095</td>
<td>1.00 ± 0.05</td>
<td>3.09 ± 0.05</td>
<td>2.73 ± 0.43</td>
<td>0.67 ± 0.11</td>
<td>0.56 ± 0.01</td>
<td>0.67 ± 0.06</td>
</tr>
<tr>
<td>Atg4a</td>
<td>1.00 ± 0.13</td>
<td>1.61 ± 0.12</td>
<td>1.58 ± 0.06</td>
<td>0.71 ± 0.00</td>
<td>0.78 ± 0.09</td>
<td>0.88 ± 0.11</td>
</tr>
<tr>
<td>Smarca1</td>
<td>1.00 ± 0.06</td>
<td>1.25 ± 0.04</td>
<td>1.32 ± 0.11</td>
<td>0.54 ± 0.06</td>
<td>0.64 ± 0.01</td>
<td>0.75 ± 0.09</td>
</tr>
</tbody>
</table>

Quantitative RT-PCR analysis of THAP11 target genes in SW620 cells expressing either control (Empty) or THAP11-Rescue-3xFLAG (Rescue) and the indicated shRNA as in Figure 3.4. RNA levels are normalized to β-actin and expressed as fold change relative to SW620 expressing control shRNA (shNS) and rescue (Empty) constructs. Values represent mean ± standard deviation of triplicate quantitative RT-PCR reactions from a representative experiment.
Figure 3.5: THAP11 knockdown de-represses putative THAP11 gene targets in Colo320HSR cells. SW620 and Colo320HSR were transduced with retrovirus encoding control (shNS) or THAP11 (shT11A or shT11C) shRNA. Cells were harvested four days post-transduction and gene expression determined by quantitative RT-PCR. Values were normalized to β-actin and expressed as fold change relative to control (shNS) shRNA. Values represent mean ± standard deviation of triplicate quantitative RT-PCR reactions from a representative experiment performed at least three times with similar results.
Figure 3.6: Chromatin immunoprecipitation (ChIP) analysis of THAP11-Rescue-3xFLAG binding at Lsmd1 and ncRNA00095.

(A) Schematic depicting the location of the regions (ChIP Amplicons) analyzed for THAP11-Rescue-3xFLAG binding at Lsmd1 and ncRNA00095. Numbers correspond to amplicon midpoint relative to transcription start site (TSS). Amplicons exhibiting peak THAP11 binding are colored red. (B) SW620 cells depleted of endogenous THAP11 by shT11A and expressing THAP11-Rescue-3xFLAG as in Figure 3.4 were analyzed by ChIP using anti-FLAG M2 monoclonal antibody or control IgG as described in Experimental Procedures. ChIP enriched DNA was analyzed by quantitative PCR at the regions depicted in (A) and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed twice.
Figure 3.7: Chromatin immunoprecipitation (ChIP) analysis of THAP11-Rescue-3xFLAG binding at AA862256 and Praf2. (A) Schematic depicting the location of the regions (ChIP Amplicons) analyzed for THAP11-Rescue-3xFLAG binding at AA862256 and Praf2. Numbers correspond to amplicon midpoint relative to transcription start site (TSS). Amplicons exhibiting peak THAP11 binding are colored red. (B) SW620 cells depleted of endogenous THAP11 by shT11A and expressing THAP11-Rescue-3xFLAG and were analyzed by ChIP using anti-FLAG M2 monoclonal antibody or control IgG as described in Figure 3.6. ChIP enriched DNA was analyzed by quantitative PCR at the regions depicted in (A) and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed twice.
Figure 3.8: Chromatin immunoprecipitation (ChIP) analysis of THAP11-Rescue-3xFLAG binding at *C1orf83* and *ZSCAN20*. (A) Schematic depicting the location of the regions (ChIP Amplicons) analyzed for THAP11-Rescue-3xFLAG binding at *C1orf83* and *ZSCAN20*. Numbers correspond to amplicon midpoint relative to transcription start site (TSS). Amplicons exhibiting peak THAP11 binding are colored red. (B) SW620 cells depleted of endogenous THAP11 by shT11A and expressing THAP11-Rescue-3xFLAG and were analyzed by ChIP using anti-FLAG M2 monoclonal antibody or control IgG as described in Figure 3.6. ChIP enriched DNA was analyzed by quantitative PCR at the regions depicted in (A) and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed twice.
Figure 3.9: Chromatin immunoprecipitation (ChIP) analysis of endogenous THAP11. (A) THAP11 occupancy at the indicated regions in SW620 cells as determined by ChIP assay using anti-THAP11 or control IgG. (B) Control (shNS) or THAP11 (shT11C) knockdown SW620 cells were analyzed by ChIP as in (A). ChIP enriched DNA was analyzed by quantitative PCR at the indicated regions and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed three times with similar results.
Figure 3.10: Chromatin immunoprecipitation (ChIP) analysis of RNAPII at Lsmd1 and ncRNA00095 in THAP11 knockdown SW620 cells. RNAPII occupancy at (A) Lsmd1 and (B) ncRNA00095 in control (shNS) or THAP11 (shT11C) knockdown SW620 cells were determined by ChIP assay. ChIP enriched DNA was analyzed by quantitative PCR at the indicated regions and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed three times.
Figure 3.11: Chromatin immunoprecipitation (ChIP) analysis of RNAPII at Lsmd1 and β-actin in THAP11 knockdown and THAP11 rescue SW620 cells. (A) ChIP analysis of RNAPII occupancy at Lsmd1 in SW620 cells expressing either control (shNS) or THAP11 (shT11A) shRNA and either control (Empty) or THAP11-Rescue-3xFLAG (Rescue). (B) ChIP analysis of RNAPII occupancy at the β-actin promoter from the indicated SW620 cells as in (A) ChIP enriched DNA was analyzed by quantitative PCR at the indicated regions and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed three times.
Figure 3.12: Chromatin immunoprecipitation (ChIP) analysis of Acetyl-K9/K18 histone H3 at Lsmd1 and ncRNA00095 in THAP11 knockdown SW620 cells. Acetyl-K9/K18 histone H3 (AcH3) occupancy at (A) Lsmd1 and (B) ncRNA00095 in control (shNS) or THAP11 (shT11C) knockdown SW620 cells were analyzed by ChIP assay. ChIP enriched DNA was analyzed by quantitative PCR at the indicated regions and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed three times.
Figure 3.13: THAP11 interacts with HCF-1 in colon cancer cells. (A) Endogenous THAP11 co-precipitates with HCF-1 in SW620 cells. SW620 nuclear extract was immunoprecipitated with either control IgG or HCF-1 antibody and the immunoprecipitates were immunoblotted with either HCF-1 or THAP11 antibodies. Input corresponds to 2% (40 μg) of starting material. (B) THAP11 associates with HCF-1 in various colon cancer cell lines. Nuclear extracts from Colo320HSR, SW480, or HCT-116 were immunoprecipitated and immunoblotted as in (A). Input corresponds to 10% (40 μg) of starting material. (C) HCF-1 and THAP11 can stably associate in the absence of DNA binding. SW620 nuclear extracts were immunoprecipitated as in (A) (-EtBR panel) or preincubated with 100 μg/ml ethidium bromide for 30 minutes prior to antibody addition (+EtBr panel) as described in Experimental Procedures. Blots are representative of experiments performed at least twice. HCF-1 precursor and proteolytic polypeptides are denoted by the arrow and bracket, respectively. THAP11 is denoted by the arrow migrating beneath the IgG heavy chain which is indicated with an asterisk.
Figure 3.14: THAP11 and HCF-1 co-occupancy on chromatin. Chromatin immunoprecipitation of THAP11 (A) or HCF-1 (B) at Lsmd1 in SW620 cells. ChIP enriched DNA was analyzed by quantitative PCR and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed twice. (C) The chromatin occupancy profiles in (A) and (B) re-plotted as percent of maximum ChIP signal. (D) Chromatin immunoprecipitation of HCF-1 at THAP11 bound chromatin.
Figure 3.15: Knockdown of THAP11 decreases HCF-1 occupancy on THAP11 target genes Lsmd1 and ncRNA00095. SW620 cells expressing either control (shNS) or THAP11 (shT11C) shRNA were analyzed for HCF-1 occupancy at Lsmd1 (A) or ncRNA00095 (B) by ChIP assay using either HCF-1 antibody or control IgG. ChIP enriched DNA was detected by quantitative PCR and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed twice. (C) Immunoblot of HCF-1 in whole cell extracts of SW620 cells expressing either control (shNS) or THAP11 knockdown (shT11C). HCF-1 precursor and proteolytic polypeptides are indicated with the arrow and bracket, respectively.
Figure 3.16: Knockdown of HCF-1 de-represses THAP11 target genes. SW620 cells were retrovirally transduced with the indicated pSRNG shRNA and transduction was verified by GFP fluorescence (data not shown). Four days post-transduction cells were harvested for whole cell extract and total RNA isolation. (A) Whole cell extracts were immunoblotted with HCF-1, THAP11, and β-actin antibodies. HCF-1 precursor and proteolytic polypeptides are denoted by the arrow and bracket, respectively. THAP11 is indicated with an arrow and a non-specific band detected by anti-THAP11 antibody is denoted with an asterisk. (B) RNA levels for Lsmd1 and ncRNA00095 were determined by quantitative RT-PCR, normalized to β-actin and expressed as fold change relative to control (shNS) shRNA. Values represent the mean ± standard deviation of duplicate RT-PCR reactions. (A) and (B) represent a single experiment repeated twice with similar results.
Figure 3.17: Knockdown of HCF-1 increases RNA Polymerase II occupancy at Lsmd1. HCF-1 and RNAPII occupancy at Lsmd1 in SW620 cells expressing either control (shNS) or HCF-1 (shHCF-1B) shRNA was determined by ChIP assay using antibodies to HCF-1 (top left), RNAPII (top right), or control IgG (bottom left). ChIP enriched DNA was detected by quantitative PCR and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed twice.
Figure 3.18: Affinity purification of THAP11/HCF-1/OGT/HDAC-1 complex from SW620 cells. (A) Schematic of two-step affinity purification procedure using wheat germ agglutinin (WGA) and anti-FLAG agarose resins as described in Experimental Procedures. (B) Two-step purification of THAP11/HCF-1 complexes from SW620 cells expressing empty vector (E) or THAP11-3xFLAG (T11). Recovery of HCF-1 and THAP11 was monitored by immunoblotting 5% of each fraction: Nuclear extract (Input), WGA-FT (WGA-agarose flow through), WGA-Eluate (WGA-agarose bound and GlcNac eluted), FLAG-Eluate (FLAG-agarose bound and eluted by boiling). Red arrow indicates endogenous THAP11; black arrow indicates THAP11-3xFLAG. (C) THAP11/HCF-1 complex contains OGT and HDAC1. Two-step purified (FLAG Eluate) and WGA-purified (WGA Eluate) fractions from control (E) and THAP11-3xFLAG (T11) expressing SW620 cells were immunoblotted using antibodies for SIN3A, OGT, and HDAC1. Arrows indicate the protein of interest. IgG heavy chain is indicated with an asterisk.
Figure 3.19: OGT occupancy at THAP11 repressed gene promoters. ChIP assay was performed in SW620 using anti-OGT or normal rabbit IgG. ChIP enriched DNA was detected by quantitative PCR and expressed as percent recovery relative to input. Values represent the mean ± standard deviation of duplicate quantitative PCR reactions from a representative experiment performed twice.
Figure 3.20: Knockdown of THAP11 restores Annexin A1 expression in SW620 cells. (A) Annexin A1 (Anxa1) mRNA levels were determined in SW480, SW620 and SW620 cells expressing the indicated shRNA by quantitative RT-PCR. Values represent the mean ± standard deviation of triplicate quantitative PCR reactions from a representative experiment performed twice with similar results. (B) Immunoblot analysis of whole cell extracts from cells in (A).
Figure 3.21: Annexin A1 and THAP11 expression inversely correlate in colon cancer cell lines. Whole cell extracts prepared from the indicated colon cancer cell lines were immunoblotted for THAP11 and Annexin A1 (Anxa1).
Figure 3.22: Knockdown of THAP11 decreases cell proliferation. SW620 cells were transduced with retrovirus expressing the indicated shRNA. Four days post-transduction cells were plated in 96-well plates (5000 cells/well) and viable cells detected using Alamar Blue every 24 hours for 5 days as described in Experimental Procedures. Values represent the mean ± standard deviation (n=8 wells) from a representative experiment performed four times with similar results.
Figure 3.23: Annexin A1 de-repression is dispensable for decreased cell proliferation in THAP11 knockdown cells. SW620 cells were transduced with pSRP retrovirus expressing control (shNS) or THAP11 (shT11) shRNA. Twenty-four hours later cells were transduced with pSRNG retrovirus expressing either control (shNS) or Annexin A1 (shAnxa1) retrovirus. Double-knockdown cells were selected with 2μg/ml puromycin and 1mg/ml Geneticin for 7 days. (A) Immunoblot analysis of Annexin A1 (Anxa1) expression in double-knockdown SW620 cells. (B) Alamar Blue cell proliferation assay in double-knockdown SW620 cells. Values represent the mean ± standard deviation (n=8 wells) from a representative experiment performed twice with similar results.
Chapter 4: General Conclusions and Future Directions

THAP domain containing proteins are a recently identified and relatively uncharacterized family of proteins with putative function in DNA and chromatin dependent processes, including transcription. The data presented within this thesis demonstrates that THAP4, THAP8, THAP10 and THAP11 possess transcriptional repressor activity when tethered to promoters as heterologous Gal4-DNA binding domain fusion proteins. This finding is consistent with previous reports implicating THAP1 and THAP7 in transcriptional regulation and suggests that the majority of THAP proteins may function in this capacity. Further characterization of endogenous THAP11 revealed a THAP11 dependent gene expression program characterized by direct, THAP11 mediated transcriptional repression. Directly repressed gene targets were found to contain chromatin bound THAP11 near their transcription start sites. In addition, THAP11 was found to physically associate with and recruit the multi-functional transcriptional regulator HCF-1 to repressed promoters. HCF-1 in turn was shown to be a critical effector of THAP11 mediated repression; knockdown of HCF-1 mirrored the transcriptional de-repression observed by THAP11 knockdown. THAP11 was also shown to be differentially expressed during human colon cancer progression. Significantly higher THAP11 protein levels were observed in primary and metastatic tumor specimens as determined by immunohistochemical analysis of tissue microarrays. Similar differential THAP11 expression was also observed in human colon cancer cell lines including the SW480/SW620 isogenic cell culture model of colon cancer progression. The
increase in THAP11 expression in colon cancer tumors and cell lines suggests that THAP11 dependent transcriptional repression may contribute to disease progression. Consistent with this hypothesis we find that knockdown of THAP11 in metastatic SW620 colon cancer cells results in a modest but significant decrease in cell proliferation. Collectively, this data provides the first characterization of a directly regulated, THAP11 dependent gene expression program in human cells and suggests THAP11 may be an important transcriptional regulator in human colon cancer.

Because the cognate DNA binding sequences of most THAP proteins are currently unknown, the data presented herein relied initially on the use of Gal4-DNA binding domain fusion proteins to investigate the transcriptional regulatory properties of THAP proteins. Endogenous THAP11 was subsequently shown to directly repress transcription, suggesting Gal4-THAP fusion proteins faithfully recapitulate the biological function of endogenous THAP proteins. Nevertheless, definitive evidence of THAP protein mediated repression requires identification of directly regulated target genes. The strategy employed here, knockdown of endogenous THAP protein with short hairpin RNAs in conjunction with epitope tagged non-silenceable rescue expression, should prove useful in identifying additional THAP protein gene targets using genome wide expression and chromatin occupancy profiling. This approach should be immediately applicable to determining THAP10 gene targets under DNA damage conditions despite the current lack of THAP10 specific antibodies. Furthermore, DNA binding site determination using genome wide chromatin immunoprecipitation techniques
avoids the inherent limitations associated with determining binding sites using in vitro iterative selection procedures like SELEX, including binding site over-selection.

The mechanism governing THAP11 chromatin occupancy near transcription start sites of directly repressed gene targets has not yet been determined. The THAP domain of THAP11 may function as a sequence specific DNA binding zinc finger as reported for THAP1 and the *Drosophila* P element transposase. Consistent with this possibility, Dejosez et al. reported a SELEX derived THAP11 binding sequence during their characterization of THAP11 as a mouse embryonic stem cell pluripotency factor. However, a bioinformatic analysis has not identified this sequence in the THAP11 bound promoters described here. This may reflect flexibility in the nucleotide sequences recognized by THAP11 or may indicate over-selection in the SELEX derived sequence. Because the chromatin immunoprecipitation experiments described here have localized THAP11 to an approximately 500bp region, future experiments using DNase I footprinting and electrophoretic mobility shift assays should allow us to determine if THAP11 directly binds DNA. If THAP11 does indeed bind directly to DNA, then we anticipate several new lines of experimentation which should further our understanding of THAP11. First, since we have currently identified six directly repressed target gene promoters, DNase I footprinting and EMSA with each bound promoter region should allow us to derive a biologically relevant THAP11 binding DNA motif. Next, using this newly derived motif we should be able to identify the nucleotide determinants of
THAP11 binding using systematic nucleotide substitution and quantitative in vitro binding assays such as fluorescence polarization or EMSA. Finally, alanine scanning mutagenesis of THAP domain amino acids should be useful in identifying the residues responsible for THAP11 sequence specific binding. It remains possible that THAP11 does not bind to DNA directly. In this scenario, THAP11 may bind specifically to histones or non-histone chromatin associated proteins. THAP7 was previously shown to bind histones through its C-terminal histone interaction domain. However, the C-termini of THAP proteins are poorly conserved suggesting THAP11 likely does not contain this domain.

THAP11 mediated repression was shown to be dependent on HCF-1 but the precise molecular mechanism underpinning this repression remains to be determined (Figure 4.1). The THAP11/HCF-1 complex was shown to contain HDAC1 and OGT. Chromatin bound OGT was found at THAP11 repressed promoters but several attempts to ChIP HDAC1 were inconclusive. Furthermore, histone acetylation was increased only at Lsmd1 and not at other de-repressed gene promoters following THAP11 knockdown. This suggests that HCF-1 tethers an HDAC independent transcriptional repressor activity to chromatin bound THAP11. We speculate that this factor may be OGT but other factors may also be involved. For instance, the HCF-1 carboxy-terminal subunit has previously been shown to associate with protein phosphatase 1. Protein phosphatase 1 (PP1) is known to be capable of de-phosphorylating RNAPII CTD and, similarly to OGT, may repress transcription by inhibiting the conversion of RNAPII to the hyperphosphorylated elongating form. An intriguing possibility
involves the combined effect of THAP11 recruited PP1 and OGT to dephosphorylate and O-GlcNAcylate RNAPII CTD, locking RNAPII in the hypophosphorylated state rendering it stalled at transcriptional start sites. A similar mechanism involving dephosphorylation and O-GlcNAcylation of Sp1 may also account for THAP11 mediated repression. These possibilities may explain the prevalence of chromatin bound THAP11 near transcriptional start sites of CpG island promoters. Future experiments including co-immunoprecipitation, ChIP, and RNA interference should determine if either OGT or PP1 function in THAP11 mediated repression. An unbiased purification of the THAP11/HCF-1 protein complex is also warranted. We anticipate that using an affinity purification scheme similar to the one outlined in chapter 3 in conjunction with mass spectrometry based proteomics should identify relevant candidate co-effectors of THAP11 mediated transcriptional repression.

THAP10 and THAP11 were found to be differentially expressed during DNA damage and colon cancer progression, respectively. The contribution of these particular THAP proteins to the physiological processes associated with these distinct events remains to be determined. With respect to THAP10, DNA damage frequently results in either cell cycle arrest or apoptosis depending on the cell type and extent of damage incurred. Thus, future experiments with THAP10 will likely focus on its potential role in modulating these cellular responses to genotoxic insults. These experiments would encompass both overexpression and knockdown of THAP10 in a variety of human cancer cell lines followed by DNA damage induction and subsequent determination of cell
cycle arrest and apoptotic response. Future experiments aimed at determining the role of THAP11 in colon cancer progression will initially utilize the SW480/SW620 cell culture model. As shown in chapter 3, knockdown of THAP11 in metastatic SW620 cells resulted in decreased cell proliferation. This diminished proliferation may result from a cell cycle defect or cell attrition resulting from increased apoptosis. Cell cycle analysis and apoptosis assays in THAP11 knockdown SW620 cells should be helpful in discriminating between these possibilities. SW620 are more tumorigenic and metastatic than SW480 cells in nude mouse xenograft assays and the role of THAP11 in promoting this phenotype will also be explored using SW620 knockdown cells. To that end, we have recently created SW620 cells harboring both GFP and luciferase reporter constructs that should facilitate the detection of metastatic SW620 cells in live mice and dissected organs (Figure 4.2). Similar experiments using overexpression of THAP11 in low-expressing colon cancer cell lines such as HCT-116 and HT-29 may also be informative.

The experiments as described above will be critical in determining if THAP10 and THAP11 are “driver” or “bystander” genes in human DNA damage response and colon cancer progression, respectively. These results notwithstanding, elucidation of the transcriptional regulatory properties of these and other THAP proteins remains a worthwhile endeavor. Future lines of research may ascribe previously unforeseen roles for THAP proteins in normal or disease processes. Having previously established molecular mechanisms for these THAP proteins may expedite the functional characterization of these
proteins in novel biological contexts. Such a scenario appears to be playing out with respect to THAP1 where prior knowledge regarding its DNA binding and transcriptional regulatory properties has allowed dystonia researchers to hypothesize that mutations in THAP1 may account for DYT6 dystonia through disrupted THAP1 dependent gene expression.
Figure 4.1: Proposed model of THAP11 mediated transcriptional repression. THAP11 bound at promoters recruits HCF-1 to mediate transcriptional repression. Repression by THAP11/HCF-1 may require additional effectors including HDAC1, OGT, and unknown (?) proteins.
Figure 4.2: Creation of GFP and luciferase expressing SW620 THAP11 knockdown cells. A) Procedure for creating SW620 cells stably expressing luciferase, GFP and shRNA. SW620 cells were transduced with retrovirus expressing luciferase-IRES-GFP. GFP positive cells were selected by fluorescence activated cell sorting (FACS). SW620\textsuperscript{Luc-GFP} cells were then transduced with retrovirus expressing THAP11 (shT11A, shT11C, shT11D) or control (shNS) and selected with puromycin. (B, Left Panel) Phase contrast and fluorescence microscopy images of SW620\textsuperscript{Luc-GFP} cells. (B, Right Panel) Whole cell extracts from parental SW620\textsuperscript{Luc-GFP} cells and cells expressing the indicated shRNA were immunoblotted for THAP11, β-actin, luciferase and GFP.
## Appendix

### Table A.1: shRNA sequences

<table>
<thead>
<tr>
<th>Target</th>
<th>shRNA</th>
<th>Targeted mRNA Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>THAP11</td>
<td>T11A</td>
<td>GCCGAAGAACUGCUGAGAA</td>
</tr>
<tr>
<td>THAP11</td>
<td>T11C</td>
<td>UGAUGGAAGUGAAGAGGA</td>
</tr>
<tr>
<td>THAP11</td>
<td>T11D</td>
<td>GAGAAGACGUGAGAGCAGG</td>
</tr>
<tr>
<td>THAP11</td>
<td>T11E</td>
<td>GAAGUGAAGUGAGAGAGGA</td>
</tr>
<tr>
<td>THAP11</td>
<td>T11F</td>
<td>CUGAUAAGCGAGCGGGGACA</td>
</tr>
<tr>
<td>THAP11</td>
<td>T11I</td>
<td>CCGAAGAGCCACACCGGUA</td>
</tr>
<tr>
<td>Annexin A1</td>
<td>Anxa1A</td>
<td>CAAAGGAGAUAUGAGAAA</td>
</tr>
<tr>
<td>Annexin A1</td>
<td>Anxa1B</td>
<td>GGAUGAAGACUAAUAGAGA</td>
</tr>
<tr>
<td>Annexin A1</td>
<td>Anxa1C</td>
<td>GAGAGGAACUGAAGAGAGA</td>
</tr>
<tr>
<td>Annexin A1</td>
<td>Anxa1D</td>
<td>GAAAGGUGACAUUGAGAAA</td>
</tr>
<tr>
<td>HCF-1</td>
<td>HCF-1A</td>
<td>GGAGCUCAUGCGGGAGAGUA</td>
</tr>
<tr>
<td>HCF-1</td>
<td>HCF-1B</td>
<td>GGAAGAGACUAAAGGCAA</td>
</tr>
<tr>
<td>HCF-1</td>
<td>HCF-1C</td>
<td>AGUCCAGUGUGUGAGCUA</td>
</tr>
<tr>
<td>HCF-1</td>
<td>HCF-1D</td>
<td>GAGAAGAGUGAGAGUGAGUA</td>
</tr>
</tbody>
</table>

### Table A.2: shRNA cloning oligonucleotides

<table>
<thead>
<tr>
<th>Oligo Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T11A Top</td>
<td>GATCCCCGCGAAGAAGAACCTGGTGAAGATTCAAGAGATTCTCAACGCATTCTGCGTTTTTA</td>
</tr>
<tr>
<td>T11A Bottom</td>
<td>AGCTTAAAAACGGCAGAAGACTGCAGTGGAACATTTCTGCTACTGCTCTTTGCGGG</td>
</tr>
<tr>
<td>T11C Top</td>
<td>GATCCCCGCTGATGGAAGTGAAGATGAATTCAAGAGATTCTCAACGCATTCTACTTTCTTT</td>
</tr>
<tr>
<td>T11C Bottom</td>
<td>AGCTTAATATGTAGAATGAGATGATTCAAGAGATTCTCAACGCATTCTTTGAGGAGG</td>
</tr>
<tr>
<td>T11D Top</td>
<td>GATCCCCGCGAAGACGTGAAGCCCATTTCAAGAGATGGGCTTCACGTCTTCTCTTTTA</td>
</tr>
<tr>
<td>T11D Bottom</td>
<td>AGCTTAAAAATGATGGAAGTGAAGATGAAATCTCTTGAATTCATCTTCACTTCGAGG</td>
</tr>
<tr>
<td>T11E Top</td>
<td>GATCCCCGCGAAGAAGACTGGAAGAGAGATTCAAGAGATTCTCAACGCATTCTTCTTT</td>
</tr>
<tr>
<td>T11E Bottom</td>
<td>AGCTTAAAAGAAGATGGAAGAAGAGATGCTCACTGCTACTTACTCAGG</td>
</tr>
<tr>
<td>T11F Top</td>
<td>GATCCCCGCTGATGGAAGTGAAGATGAAATCTCTTGAATTCATCTTCACTTACGAGG</td>
</tr>
<tr>
<td>T11F Bottom</td>
<td>AGCTTAAAATGATGGAAGTGAAGATGAAATCTCTTGAATTCATCTTCACTTACGAGG</td>
</tr>
<tr>
<td>T11I Top</td>
<td>GATCCCCGCGAAGACCTACACCGTATTCAAGAGATACCGTGTAGGTCTTGAGTTTTTA</td>
</tr>
<tr>
<td>T11I Bottom</td>
<td>AGCTTAAAAACCGCAAGAAGCTACAGGTATCATCAGCTATCCGAGATTTCATCTCACAG</td>
</tr>
<tr>
<td>Anxa1A Top</td>
<td>GATCCCCGCGAAGAAGACTGGAAGAGAGATTCAAGAGATTCTCAACGCATTCTTTT</td>
</tr>
<tr>
<td>Anxa1A Bottom</td>
<td>AGCTTAAAAACCGCAGAAGATGGAAGAAGAGATGCTCACTGCTACTTACTT</td>
</tr>
<tr>
<td>Anxa1B Top</td>
<td>GATCCCCGCGAAGAAGACTGGAAGAGAGATTCAAGAGATTCTCAACGCATTCTTTT</td>
</tr>
<tr>
<td>Anxa1B Bottom</td>
<td>AGCTTAAAATGATGGAAGTGAAGATGAAATCTCTTGAATTCATCTTCACTTACGAGG</td>
</tr>
<tr>
<td>Anxa1C Top</td>
<td>GATCCCCGCGAAGAAGACTGGAAGAAGAGATGCTCACTGCTACTTACTTACTT</td>
</tr>
<tr>
<td>Anxa1C Bottom</td>
<td>AGCTTAAAATGATGGAAGTGAAGATGAAATCTCTTGAATTCATCTTCACTTACGAGG</td>
</tr>
<tr>
<td>Anxa1D Top</td>
<td>GATCCCCGCGAAGAAGACTGGAAGAAGAGATGCTCACTGCTACTTACTTACTT</td>
</tr>
<tr>
<td>Anxa1D Bottom</td>
<td>AGCTTAAAATGATGGAAGTGAAGATGAAATCTCTTGAATTCATCTTCACTTACGAGG</td>
</tr>
</tbody>
</table>
Table A.3: Non-silenceable THAP11 site-directed mutagenesis primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T11A Rescue F</td>
<td>GCTGCACCAGAGCTCCGAGAGGATGCGC</td>
</tr>
<tr>
<td>T11A Rescue R</td>
<td>GCCGATCCTTCTCGGAGCTCTTCGCGCAGC</td>
</tr>
<tr>
<td>T11CE Rescue F</td>
<td>GACATCCTGGCTCTGAGATGGAGGTCAAAATGAAAGAGATGAAAGGC</td>
</tr>
<tr>
<td>T11CE Rescue R</td>
<td>GCCTTTCATCTCTTTCATTTTGACCTCCATCAGAGCCAGGATGTC</td>
</tr>
</tbody>
</table>

Table A.4: ChIP primer sequences

<table>
<thead>
<tr>
<th>Amplicon (Gene, Center Position Relative to TSS)</th>
<th>Forward Primer Sequence</th>
<th>Reverse Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA862256, -1778</td>
<td>CAATTACGGCCAGAGCTTTGAAGT</td>
<td>TTATGTCCCACCCTTTGGAG</td>
</tr>
<tr>
<td>AA862256, -1340</td>
<td>CTAAGCCTCCCTCCAGATTCC</td>
<td>CACAGTCCCTTTTCTCAGC</td>
</tr>
<tr>
<td>AA862256, -909</td>
<td>CATCTAGCGCCTCATACAAAAA</td>
<td>TGGTTATGCGCAAGAAAACA</td>
</tr>
<tr>
<td>AA862256, -566</td>
<td>CTCCTGCTAGAGCCTTTCT</td>
<td>GACCCAGTGTTGGAAAACTT</td>
</tr>
<tr>
<td>AA862256, 101</td>
<td>GCGGACTCTCTAAGGCTCCAC</td>
<td>GAGGGAAGGTAAGGTGTCC</td>
</tr>
<tr>
<td>AA862256, 619</td>
<td>TTGATGAGCCGACGACACTT</td>
<td>AAACCTACAGGGCGGCTACT</td>
</tr>
<tr>
<td>AA862256, 1254</td>
<td>GCACGAGGTAACCCACCACCT</td>
<td>GTGACCCCAACAGGAAGGTC</td>
</tr>
<tr>
<td>AA862256, 1827</td>
<td>TGAATCCCCAGGAGACCTTCC</td>
<td>CAGGTAAGGGATGACCTGTTG</td>
</tr>
<tr>
<td>AA862256, 2454</td>
<td>CTTTTCAATTGCTGATGGTG</td>
<td>CCTCCCAAACAGTCTTGAT</td>
</tr>
<tr>
<td>AA862256, 2993</td>
<td>TCTGTTGAGATCCAGATGGT</td>
<td>CATCCTAGCATGCTTTGAAA</td>
</tr>
<tr>
<td>AA862256, 3656</td>
<td>AGGGGGCATGTAACGATGAG</td>
<td>TCCGTCTCTCACAAGAAAAGA</td>
</tr>
<tr>
<td>AA862256, 4430</td>
<td>GCCCTGACATGGTATGGTGT</td>
<td>ATGGGGTGAAAGCCTGATTA</td>
</tr>
<tr>
<td>AA862256, 5141</td>
<td>TTGAAGGGGCTACCTTATTG</td>
<td>TAAAGCCACATCACAAGC</td>
</tr>
<tr>
<td>AA862256, 5757</td>
<td>GAAACTGAGCAGGCCCCCTTG</td>
<td>TAAAGGGTACGGCTGAACT</td>
</tr>
<tr>
<td>AA862256, 5981</td>
<td>ACCTGAGGACACCGACACAT</td>
<td>TTTGCGTTGACCTGGATGAT</td>
</tr>
<tr>
<td>AA862256, 6426</td>
<td>CTGAAAAGGGCTAGGCGATTG</td>
<td>CATGAAAGGAAGGGAGCA</td>
</tr>
<tr>
<td>Alg14, -1869</td>
<td>ACGCATTTTGTAAAGGCCCCAG</td>
<td>AAGGTTGAGAAAGTTCATAGTTG</td>
</tr>
<tr>
<td>Alg14, -1247</td>
<td>GAATGGAGACACGGCTTTACA</td>
<td>CAGACTCCAGGGCTCATTA</td>
</tr>
<tr>
<td>Gene</td>
<td>Position</td>
<td>Sequence</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Alg14,  -767</td>
<td>CATCATTTTGGCTGAGGACA</td>
<td>TGCAAGCTCAGTGCTCTA</td>
</tr>
<tr>
<td>Alg14,  -80</td>
<td>ATTGTGACTGGGACCTCAGC</td>
<td>GTCTTGTTGGCTGGCTCTT</td>
</tr>
<tr>
<td>Alg14,  -659</td>
<td>AAAACAGTGCTAAGGCTGCTT</td>
<td>TGCAAGGTACTGTTGGCTCT</td>
</tr>
<tr>
<td>Alg14,  -1195</td>
<td>GTGCTATGGCAGTCTTACAGGA</td>
<td>GCCGTGCTCCTCAGTGGACT</td>
</tr>
<tr>
<td>Alg14,  1800</td>
<td>CCACAGCTCAGTGGCTCTA</td>
<td>AGCAAGTGTTACAGGGCTAT</td>
</tr>
<tr>
<td>Lsmd1,  1105</td>
<td>GGACCAGCACATTGCTCTT</td>
<td>TCTGAAAGTGTAAGCAGGCA</td>
</tr>
<tr>
<td>Lsmd1,  1195</td>
<td>ATTGTGACTGGGACCTCAGC</td>
<td>GTCTTGTTGGCTGGCTCTT</td>
</tr>
<tr>
<td>Lsmd1,  659</td>
<td>AAAACAGTGCTAAGGCTGCTT</td>
<td>TGCAAGGTACTGTTGGCTCT</td>
</tr>
<tr>
<td>Lsmd1,  1195</td>
<td>GTGCTATGGCAGTCTTACAGGA</td>
<td>GCCGTGCTCCTCAGTGGACT</td>
</tr>
<tr>
<td>Lsmd1,  1800</td>
<td>CCACAGCTCAGTGGCTCTA</td>
<td>AGCAAGTGTTACAGGGCTAT</td>
</tr>
<tr>
<td>Lsmd1,  129</td>
<td>AAAACAGTGCTAAGGCTGCTT</td>
<td>TGCAAGGTACTGTTGGCTCT</td>
</tr>
<tr>
<td>Lsmd1,  -506</td>
<td>ACTTCCACGATCGTTTCCAT</td>
<td>TCTGAAAGGTAAGCGCCATC</td>
</tr>
<tr>
<td>Lsmd1,  -944</td>
<td>GGGATTCCCCATTCCTTCTA</td>
<td>GTGGCAGGGGTCGTATTATG</td>
</tr>
<tr>
<td>Lsmd1,  -1502</td>
<td>AGGGGAGATAGCGGTGTAGG</td>
<td>GAAGCTGCTATCTGCCAAGG</td>
</tr>
<tr>
<td>Lsmd1,  -2139</td>
<td>AAGCATCCAGGAGGGTTGTT</td>
<td>TGGCCAGGAGGGTTTATG</td>
</tr>
<tr>
<td>Lsmd1,  -2473</td>
<td>ACTTCCACGATCGTTTCCAT</td>
<td>TCTGAAAGGTAAGCGCCATC</td>
</tr>
<tr>
<td>Lsmd1,  -2846</td>
<td>CTGCAAACCTCCCTCCTCCGT</td>
<td>GTTGGTTGGCTGGCTCTC</td>
</tr>
<tr>
<td>Lsmd1,  -3176</td>
<td>TTGCTAAGGAGGCTGCTGCT</td>
<td>AGCAAGTGTTACAGGGCTAT</td>
</tr>
<tr>
<td>Lsmd1,  -3434</td>
<td>AAGCCAAAGCAGAATGTCCTT</td>
<td>GTGGGCCCCCTGCAGATCTTC</td>
</tr>
<tr>
<td>Lsmd1,  -4080</td>
<td>GTGAGGTTGCTAAGGCAAGAC</td>
<td>AACTAGCTCAGTCTCGCAC</td>
</tr>
<tr>
<td>Lsmd1,  -4538</td>
<td>TTGCTAAGGAGGCTGCTGCT</td>
<td>AGCAAGTGTTACAGGGCTAT</td>
</tr>
<tr>
<td>Lsmd1,  -4912</td>
<td>TGCCATATGAGGAGAGGGTT</td>
<td>GGCAACAAGGAACAGAAACTT</td>
</tr>
<tr>
<td>Lsmd1,  -1739</td>
<td>CAAGATCTCAACCCCTTCTTG</td>
<td>CCCAAGGAATTTGGGAAGATT</td>
</tr>
<tr>
<td>Lsmd1,  -1218</td>
<td>CTTGAGCTGATCGTTTCCAT</td>
<td>GTTGATGAGGCTGCTGCT</td>
</tr>
<tr>
<td>Lsmd1,  -830</td>
<td>AGGGTGCTCAGCTTCTGC</td>
<td>GTTGATGAGGCTGCTGCT</td>
</tr>
<tr>
<td>Lsmd1,  -2139</td>
<td>AAGCATCCAGGAGGGTTGTT</td>
<td>TGGCCAGGAGGGTTTATG</td>
</tr>
<tr>
<td>Lsmd1,  -2473</td>
<td>ACTTCCACGATCGTTTCCAT</td>
<td>TCTGAAAGGTAAGCGCCATC</td>
</tr>
<tr>
<td>Lsmd1,  -2846</td>
<td>CTGCAAACCTCCCTCCTCCGT</td>
<td>GTTGGTTGGCTGGCTCTC</td>
</tr>
<tr>
<td>Lsmd1,  -3176</td>
<td>TTGCTAAGGAGGCTGCTGCT</td>
<td>AGCAAGTGTTACAGGGCTAT</td>
</tr>
<tr>
<td>Lsmd1,  -3434</td>
<td>AAGCCAAAGCAGAATGTCCTT</td>
<td>GTGGGCCCCCTGCAGATCTTC</td>
</tr>
<tr>
<td>Lsmd1,  -4080</td>
<td>GTGAGGTTGCTAAGGCAAGAC</td>
<td>AACTAGCTCAGTCTCGCAC</td>
</tr>
<tr>
<td>Lsmd1,  -4538</td>
<td>TTGCTAAGGAGGCTGCTGCT</td>
<td>AGCAAGTGTTACAGGGCTAT</td>
</tr>
<tr>
<td>Lsmd1,  -4912</td>
<td>TGCCATATGAGGAGAGGGTT</td>
<td>GGCAACAAGGAACAGAAACTT</td>
</tr>
<tr>
<td>Lsmd1,  -1739</td>
<td>CAAGATCTCAACCCCTTCTTG</td>
<td>CCCAAGGAATTTGGGAAGATT</td>
</tr>
<tr>
<td>Lsmd1,  -1218</td>
<td>CTTGAGCTGATCGTTTCCAT</td>
<td>GTTGATGAGGCTGCTGCT</td>
</tr>
<tr>
<td>Lsmd1,  -830</td>
<td>AGGGTGCTCAGCTTCTGC</td>
<td>GTTGATGAGGCTGCTGCT</td>
</tr>
<tr>
<td>Lsmd1,  -2139</td>
<td>AAGCATCCAGGAGGGTTGTT</td>
<td>TGGCCAGGAGGGTTTATG</td>
</tr>
<tr>
<td>Lsmd1,  -2473</td>
<td>ACTTCCACGATCGTTTCCAT</td>
<td>TCTGAAAGGTAAGCGCCATC</td>
</tr>
<tr>
<td>Lsmd1,  -2846</td>
<td>CTGCAAACCTCCCTCCTCCGT</td>
<td>GTTGGTTGGCTGGCTCTC</td>
</tr>
<tr>
<td>Lsmd1,  -3176</td>
<td>TTGCTAAGGAGGCTGCTGCT</td>
<td>AGCAAGTGTTACAGGGCTAT</td>
</tr>
<tr>
<td>Lsmd1,  -3434</td>
<td>AAGCCAAAGCAGAATGTCCTT</td>
<td>GTGGGCCCCCTGCAGATCTTC</td>
</tr>
<tr>
<td>Lsmd1,  -4080</td>
<td>GTGAGGTTGCTAAGGCAAGAC</td>
<td>AACTAGCTCAGTCTCGCAC</td>
</tr>
<tr>
<td>Lsmd1,  -4538</td>
<td>TTGCTAAGGAGGCTGCTGCT</td>
<td>AGCAAGTGTTACAGGGCTAT</td>
</tr>
<tr>
<td>Lsmd1,  -4912</td>
<td>TGCCATATGAGGAGAGGGTT</td>
<td>GGCAACAAGGAACAGAAACTT</td>
</tr>
<tr>
<td>ncRNA00095,-1739</td>
<td>CAAGATCTCAACCCCTTCTTG</td>
<td>CCCAAGGAATTTGGGAAGATT</td>
</tr>
<tr>
<td>ncRNA00095,-1218</td>
<td>CTTGAGCTGATCGTTTCCAT</td>
<td>GTTGATGAGGCTGCTGCT</td>
</tr>
<tr>
<td>ncRNA00095,-830</td>
<td>AGGGTGCTCAGCTTCTGC</td>
<td>GTTGATGAGGCTGCTGCT</td>
</tr>
<tr>
<td>ncRNA00095,-212</td>
<td>CCCCTCTTCTACTCCCTCCTTCTT</td>
<td>TGGCTTCCCTGAAATGTTGAG</td>
</tr>
<tr>
<td>ncRNA00095,335</td>
<td>CTTAACGATGGCGCAACAG</td>
<td>GAGCAGTCTGCTGCTGCT</td>
</tr>
<tr>
<td>ncRNA00095,955</td>
<td>TAAAGATGTCCGCTCTCC</td>
<td>GACTGATTCCTGCTGCTC</td>
</tr>
<tr>
<td>ncRNA00095,1368</td>
<td>GCCGTTGCTAAGGCCTCTTCT</td>
<td>TGGGGCTGAGGCTGCTGCT</td>
</tr>
<tr>
<td>ncRNA00095,1771</td>
<td>TTGCTTCCGCTTACTGGGTGTG</td>
<td>TGGCCAGGATGCCCCAGAT</td>
</tr>
<tr>
<td>ncRNA00095,2268</td>
<td>CCGCTCGGCTTCTTATCATG</td>
<td>GAACGCAGGGAGGGCTTAT</td>
</tr>
<tr>
<td>ncRNA00095,2902</td>
<td>GTGAGGCCCCCTTCTCTAAACA</td>
<td>TCTAGGCTTAACTTTCCAT</td>
</tr>
<tr>
<td>ncRNA00095,3400</td>
<td>GCGCAAGCTGCCAGCCTCTAC</td>
<td>GGGGCTTACTTCGTTGCTC</td>
</tr>
<tr>
<td>ncRNA00095,3883</td>
<td>AGGATCTGAGGGGACTGAGA</td>
<td>GACTGATACGCTGCTGCTC</td>
</tr>
<tr>
<td>Praf2,  -1769</td>
<td>TAGGTGTTGGAAGTGGACTTCTC</td>
<td>TGGGCTGAGTGGCAGAGA</td>
</tr>
<tr>
<td>Praf2,  -1184</td>
<td>TGCCTGCTGCTGCTGCTGCTT</td>
<td>CTGGCGACGAGCATGAGA</td>
</tr>
<tr>
<td>Praf2,  -829</td>
<td>TCCCTGATGGGAAACTGCAACA</td>
<td>CCGGCTGCTTAAAGTCT</td>
</tr>
<tr>
<td>Praf2,  -287</td>
<td>TCCCTTCGCTGAGTGGACTTCTC</td>
<td>TGGGCTGAGTGGCAGAGA</td>
</tr>
<tr>
<td>Praf2,  -112</td>
<td>CCGCTGAGCTGCTGCTGCTGCTT</td>
<td>CTGGCGACGAGCATGAGA</td>
</tr>
<tr>
<td>Praf2,  805</td>
<td>GCTGGGAAGGCCTGAGGCAAGC</td>
<td>CCGGCGACGAGCATGAGA</td>
</tr>
<tr>
<td>Praf2,  1158</td>
<td>GGACCTGAGGAGGACAGGAGAC</td>
<td>CCGGCGACGAGCATGAGA</td>
</tr>
<tr>
<td>Praf2,  1906</td>
<td>AGCTGGCAAGCAAGCCTTTCGCTG</td>
<td>CTGGCGACGAGCATGAGA</td>
</tr>
<tr>
<td>Praf2,  2298</td>
<td>GGGATGTTGAGGAGGATGAGA</td>
<td>CCGGCGACGAGCATGAGA</td>
</tr>
<tr>
<td>Praf2,  2839</td>
<td>AGGGATGTTGAGGAGGATGAGA</td>
<td>CCGGCGACGAGCATGAGA</td>
</tr>
<tr>
<td>ZSCAN20,  1602</td>
<td>TCTAGGAGGCACTCCAGT</td>
<td>CTGGCGACGAGCATGAGA</td>
</tr>
<tr>
<td>ZSCAN20,  1051</td>
<td>CTGAGGCTACGCTGCTGCTT</td>
<td>CTGGCGACGAGCATGAGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>ZSCAN20, 457</td>
<td>CCCTGGCAAGAAGTGAAGAG</td>
<td>CCACTGGGTTCCAGAGGTAA</td>
</tr>
<tr>
<td>ZSCAN20, -87</td>
<td>GGCTGAAGACGCTCTGAAAGG</td>
<td>CATCCTCCAGAAAGGAACC</td>
</tr>
<tr>
<td>ZSCAN20, -530</td>
<td>GGCCTGAGCTAATCGTTTT</td>
<td>TGCCATCCATGTCACAGAGT</td>
</tr>
<tr>
<td>ZSCAN20, -1307</td>
<td>TTCCCTAAGCCGCAATACC</td>
<td>GAGACGGAAGGATCGCTTTA</td>
</tr>
<tr>
<td>ZSCAN20, -1850</td>
<td>GTCTTACCTGACCCCTGCAA</td>
<td>AGCCGAAACATTTGAGAGA</td>
</tr>
<tr>
<td>C1orf83, 1313</td>
<td>GTGCGCTTTGATGCTGTCTT</td>
<td>GCCCTTTCCCAGTCTCAGC</td>
</tr>
<tr>
<td>C1orf83, 810</td>
<td>CTGCCCTCTTTTCTTGACCA</td>
<td>TTCTGGATTTCTAGCGTTCG</td>
</tr>
<tr>
<td>C1orf83, 360</td>
<td>GACCCCTCTGCTGTCTCTTG</td>
<td>ACCACAGCCATAATCCCAAG</td>
</tr>
<tr>
<td>C1orf83, -56</td>
<td>CTGTCGGAAGTTGTCCTC</td>
<td>CAGAATGCACTGCGAAGAAA</td>
</tr>
<tr>
<td>C1orf83, -303</td>
<td>TTTTCTCGGAAAGGTTTTCCT</td>
<td>TCGTAGTCCAGAGGTTAGAGG</td>
</tr>
<tr>
<td>C1orf83, -1012</td>
<td>TGCTGACTCCCGAAGACT</td>
<td>TAGAAGGGGAGCCAGATA</td>
</tr>
<tr>
<td>C1orf83, -1557</td>
<td>TCTCCGTAGCCACTCCGTTT</td>
<td>TTGGATTGGCTGCTGGT</td>
</tr>
</tbody>
</table>

**143**


