IDENTIFICATION AND CHARACTERIZATION OF THE MULTIFUNCTIONAL
EPIGENETIC REGULATOR CFP1 AS AN ERK1/2 SUBSTRATE

APPROVED BY SUPERVISORY COMMITTEE

Melanie H. Cobb, Ph.D.

Paul Sternweis, Ph.D.

Joel Goodman, Ph.D.

Nicholas Conrad, Ph.D.

DEDICATION

Thank you to all of the nerds, geeks and dorks who inspired and supported me through this

endeavor, especially TJ, my parents, and Melanie.

IDENTIFICATION AND CHARACTERIZATION OF THE MULTIFUNCTIONAL
EPIGENETIC REGULATOR CFP1 AS AN ERK1/2 SUBSTRATE

by

AILEEN MELANIE KLEIN

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

December, 2014

IDENTIFICATION AND CHARACTERIZATION OF THE MULTIFUNCTIONAL
EPIGENETIC REGULATOR CFP1 AS AN ERK1/2 SUBSTRATE

Aileen Melanie Klein

The University of Texas Southwestern Medical Center at Dallas, 2014

Supervising Professor: Melanie H. Cobb, Ph.D.

Epigenetic regulation of gene transcription occurs as an integration of multiple layers of signals at a genetic locus. These signals can include local chromatin structure, covalent modifications to both histone proteins and DNA, the presence of transcription factors, and modification directly to the transcriptional machinery. Our lab is interested in the control of cellular processes by the mitogen activated protein kinases ERK1/2. In a yeast two-hybrid screen with activated ERK2 (extracellular signal-regulated kinase 2) to find novel interacting partners, our lab identified CFP1 (CxxC finger protein 1), a DNA-binding protein that is a vital component of the H3K4 trimethylating Set1A/B complexes to promote gene transcription. CFP1 has also been shown to interact physically and functionally with the major maintenance DNA methyltransferase DNMT1. We are interested in defining how substrate targeting of CFP1 by ERK1/2 regulates downstream transcriptional outcomes. Interaction between ERK2 and CFP1 in cells was validated by co-immunoprecipitation from isolated mononucleosomes. Active ERK2 can phosphorylate CFP1 on multiple sites *in vitro*, an observation supported by studies in cells. Some of the most likely *in vivo* ERK1/2 phosphorylation sites include serine 224 and threonine 227. CFP1 is essential for focusing trimethylation of H3K4 at promoters, a histone modification that supports transcription from these loci. We hypothesized that phosphorylation of CFP1 by ERK1/2 during mitogenic signaling may support trimethylation of H3K4 and transcription of

ERK1/2-regulated target genes. Introduction of CFP1 containing the mutation T227V into HeLa cells blocked global H3K4 trimethylation to a similar extent as CFP1 depletion. On the other hand, CFP1 S224A shows diminished transactivation capacity against a model transcriptional substrate. Neither of these mutants fail to interact with Set1B in a pulldown, suggesting that these sites may be important for Set1 complex targeting or activity towards chromatin. Consistently, CFP1 knockdown hinders induction of several ERK1/2-regulated immediate early gene targets in response to serum treatment. It will be of interest to test whether this is dependent on stable or inducible H3K4 trimethylation and what impact overexpression of point mutants will play in their transcription. Regulation of H3K4 trimethylation through CFP1 phosphorylation might represent a novel regulatory input to support transcription of ERK1/2-regulated genes.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Klein A. M., and Cobb M. H. (2014) ERK5 signaling gets XIAPed: a role for ubiquitin in the disassembly of a MAPK cascade. *The EMBO journal* **33**, 1735-1736

Klein, A. M., Zaganjor, E., and Cobb, M. H. (2013) Chromatin-tethered MAPKs. *Current opinion in cell biology* **25**, 272-277

Klein, A. M., Dioum, E. M., and Cobb, M. H. (2010) Exposing contingency plans for kinase networks. *Cell* **143**, 867-869

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF DEFINITIONS

CFP1   CXXC finger protein 1
CGI   CpG island
ChIP   Chromatin immunoprecipitation
CMV   Cytomegalovirus
COMPASS   Complex of proteins associated with Set1
CTD   C-terminal domain
DMEM   Dulbecco's Modified Eagle Medium
DMSO   Dimethyl sulfoxide
DNMT1   DNA methyltransferase 1
EGF   Epidermal growth factor
ERK1/2   Extracellular signal regulated kinases 1 and 2
ESC   Embryonic stem cell
FBS   Fetal bovine serum
GFP   Green fluorescent protein
GST   Glutathione S-transferase
H3K4me3   Histone 3 Lysine 4 trimethylation
IEG   Immediate early gene
MAPK   Mitogen-activated protein kinase
MLL   Mixed-lineage leukemia
NGF   Nerve growth factor
ORF   Open reading frame
P-TEFb   Positive transcription elongation factor
PAA   Phosphoamino acid analysis
PHD finger   Plant homeodomain finger
PMA   Phorbol myristate acetate
PRC2   Polycomb repressive complex 2
PTM   Post-translational modification
qRT-PCR   Quantitative real-time PCR
RNAPII   RNA polymerase II
SET domain   Su(var)3-9, Enhancer of zeste and Trithorax domain
SILAC   Stable isotope labeling by amino acids in cell culture
SRF   Serum-response factor
TAF   TBP-associated factor
TSA   Trichostatin A
TSS   Transcription start site
UTR   Untranslated region

CHAPTER ONE
Introduction

Since initially cloning the mitogen-activated protein kinases (MAPKs) ERK1/2 over 20 years ago, our lab has been interested in identifying upstream mechanisms governing their regulation and downstream targets that carry out intracellular responses to their activation. ERK1/2 (extracellular signal regulated kinases) participate as key relays to couple extracellular signals, typically those involved in growth and proliferation, with appropriate intracellular responses (1). They are protein-serine/threonine kinases, and are widely regarded to be functionally redundant. Our search for novel substrates led us to perform a yeast two-hybrid screen for novel ERK2 interacting proteins that relied on kinase activation for enhanced binding, in the hope that identified proteins would prove to be new substrates.

This experiment identified CXXC Finger Protein 1 (CFP1) as an ERK2-interacting protein. CFP1 is an important regulator of at least two functionally antagonistic epigenetic processes – histone methylation and DNA methylation (2). As a constitutive component of the Set1A/B Histone 3 Lysine 4 (H3K4) trimethylase complexes, CFP1 binds to DNA directly at CpG islands, specialized DNA regions near or within gene promoters, to direct H3K4 trimethylation to these loci (3,4). Set1A/B catalyze at least 80% of H3K4 trimethylation in somatic cells (5). H3K4 trimethylation is tightly clustered around actively transcribed promoters, so much so that it can be used to predict the location of transcriptional start sites (6,7). Although many H3K4 trimethyl binding proteins have been identified, the particular mechanisms responsible for recognizing this mark and propagating it to induce transcription remain unclear (8). CFP1 loss results in broad mislocalization of H3K4 trimethylation, and the extent of promoter H3K4 trimethylation is greatly impacted at a majority of genes, although gene expression changes are more subtle (4).

DNA methylation occurs on cytidine residues, typically within CpG islands. CpG methylation is tightly associated with transcriptional repression of underlying genes, an event that is essential in dictating tissue-specific identity during development by silencing specific subsets of genes (9). In most cell types, the majority of CpG island-associated genes are devoid

of methylation, but may or may not be actively transcribed. Aberrant CpG methylation is a typical event in cancer progression, silencing expression of tumor suppressor genes (10). Cytosine methylation is catalyzed by DNA methyltransferase proteins (DNMTs). DNMT1 is the critical maintenance methyltransferase that recognizes hemimethylated sequences in newly-replicated DNA and catalyzes methyl group addition to the new strand. CFP1 interacts with DNMT1, and CFP1-null mice and embryonic stem cells exhibit profound defects in DNA methylation (11,12). The mechanistic role that CFP1 plays in DNMT1 targeting and activity is unclear.

Our goals were to validate interactions between ERK1/2 and CFP1, then address whether or not CFP1 can serve as a substrate for ERK1/2. Pursuing the diverse physiological outcomes that might be attributed to CFP1 phosphorylation was attractive, because potential roles for ERK1/2 in both histone and DNA methylation are easy to envision.

Acute ERK1/2 activation drives transcription of a small cross-section of genes, known as immediate early genes (IEGs; (13)). IEGs are predominantly transcription factors and signaling molecules that, when translated, go on to regulate successive rounds of transcription of secondary response genes. ERK1/2 are recruited to the promoters of IEGs in response to mitogenic stimuli, where they phosphorylate substrates to support transcriptional initiation. These substrates include transcription factors (14,15), histone modifying enzymes (16), and components of the transcriptional machinery (17,18), and their phosphorylation by ERK1/2 is a prerequisite for full transcriptional activation of target genes. Given the complexity of transcriptional activation and the fact that new nuclear phosphorylation targets are being continually identified, we hypothesized that CFP1 phosphorylation by ERK1/2 might support transcriptional activation of ERK1/2 target genes by directing H3K4 methylating activities to appropriate promoters (see schematic in Figure 1-1).

We could also foresee potential regulatory mechanisms for ERK1/2 phosphorylation of CFP1 in DNA methylation. ERK1/2 are regulated by the small GTPase family of Ras proteins. KRas is mutated in a wide variety of cancers, leading to constitutive activation of ERK1/2 and a pleiotropic host effects on transcriptional regulation by this pathway (19,20). One of the documented effects of Ras activation is aberrant gene silencing, mediated by DNA methylation and enforced by accumulation of histone modifications that ensure DNA compaction and
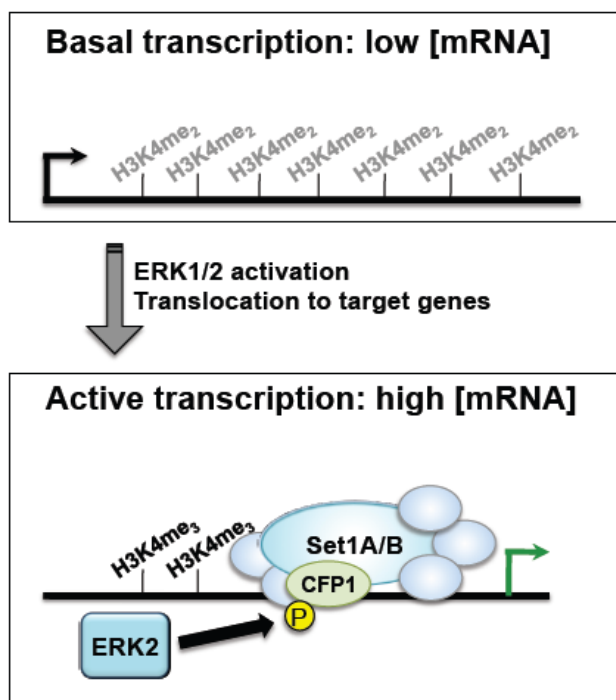
Figure 1-1: Hypothetical model of ERK1/2 contribution to H3K4 trimethylation through the Set1 methyltransferase complexes.

silencing (21,22). One outcome of this event can be the silencing of tumor suppressor genes, enhancing cancer development. A knockdown screen to identify signaling events that support Ras-driven gene silencing identified ERK2 and DNMT1 (21). Functional interactions between DNMT1 and CFP1 have not been well-studied outside of a developmental context (23-25). We envisioned that aberrant DNMT1 activity in Ras-activated cells could rely on ERK1/2 phosphorylation of CFP1, leading to silencing of target genes (see schematic in Figure 1-2).

We will discuss the topics raised here in greater detail in Chapter 2, giving the reader perspective on the current state of research on MAPK signaling, transcription, and their intersection, as well as in-depth accounts of studies on CFP1 by other research groups. In Chapter 3, we will detail experiments designed to validate our yeast two-hybrid findings, as well as catalogue post-translational modifications to CFP1 *in vitro* and in cells. The role that CFP1 plays in supporting histone methylation and transcription, as well as how ERK1/2 phosphorylation might impact these functions, is discussed in Chapter 4.
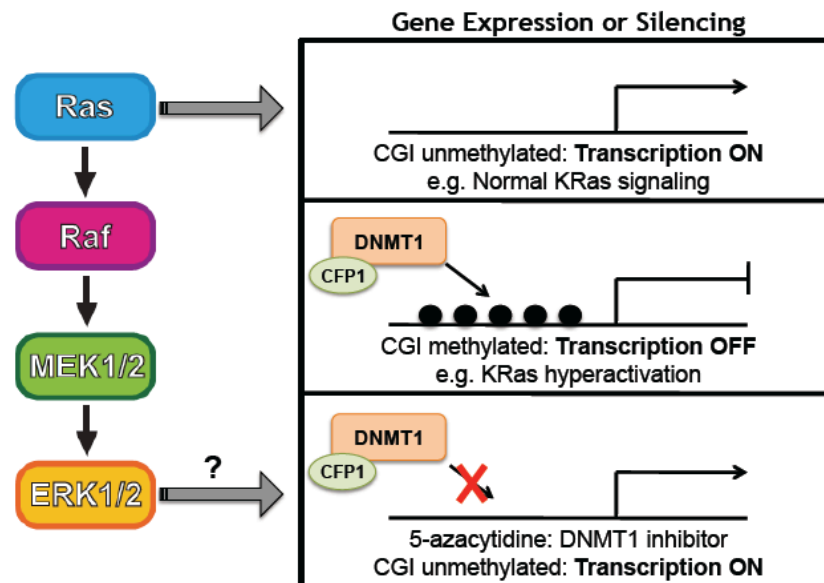
Figure 1-2: Working model of potential ERK1/2 contribution to CpG island methylation

CHAPTER TWO
Review of the Literature

ERK1/2 ARE PROTOTYPICAL MITOGEN ACTIVATED PROTEIN KINASES

Mitogen-activated protein kinases (MAPKs) are serine/threonine protein kinases that are essential mediators of cellular signaling. They are activated in response to a wide variety of extracellular signals, ranging from stress responses to growth and mitogenic signaling. Four major families exist in mammals: extracellular signal-regulated kinases 1 and 2 (ERK1/2); p38 $\alpha/\beta/\gamma/\delta$; c-Jun N-terminal kinases 1-3 (JNKs); and ERK5 (1). These kinases share extensive amino acid identity within their catalytic core regions and among the overall architectural features of their activation networks.

ERK1/2 were identified over 20 years ago as insulin-responsive serine/threonine kinases that share a high degree of homology with the yeast Fus3 and Kss1 kinases that mediate the *Saccharomyces cerevisiae* pheromone response (26-31). Today, they are recognized as key signal transduction molecules that regulate diverse processes such a proliferation (32), differentiation (33,34), motility (35) and stress (36), and their misregulation can have broad impact on normal physiology (37).

ERK1 and ERK2 share 84% amino acid identity, and are widely regarded to be functionally redundant (1). Knockout studies in mice may argue otherwise. Loss of ERK2 results in post-implantation embryonic lethality before embryonic day 8.5 as a consequence of improper trophoblast development (38). However, loss of ERK1 has a far milder impact on development, resulting in a largely normal adult animal with moderate defects in thymocyte development and activation (39). ERK2 is expressed to a greater extent than ERK1 in most tissues, so it is possible that total ERK1/2 dosage may account for these disparate phenotypic outcomes, and recent work has demonstrated that overexpression of ERK1 in mice can phenotypically rescue simultaneous loss of ERK2 (Christophe Frémin, personal communication). Whether ERK1/2 are fully functionally redundant remains a contentious issue, with little indisputable evidence to support or refute the claim (40-42).
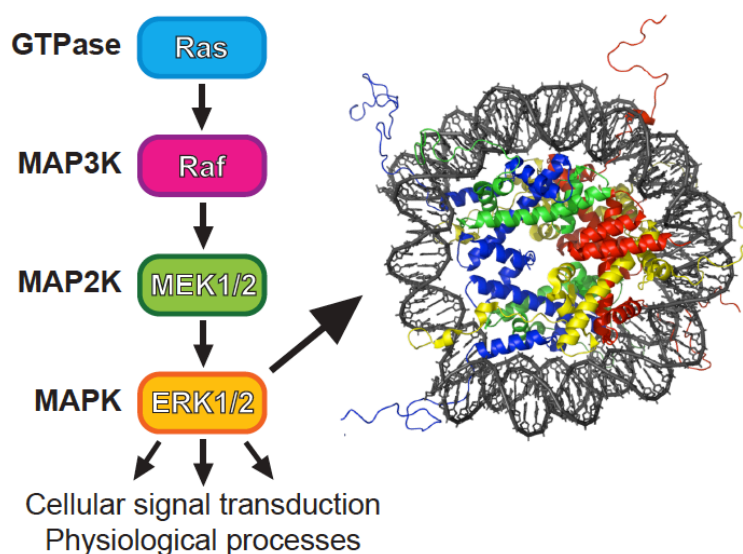
Figure 2-1: The ERK1/2 MAPK pathway. ERK1/2 are activated as the terminal MAP kinase in a three-kinase module, with MEK1/2 acting as the immediate upstream kinases (MAP2Ks), and in turn the Raf family proteins as MEK1/2 activators (MAP3Ks). Rafs are activated downstream of Ras GTPases. ERK1/2 phosphorylate a myriad of substrate proteins in many cellular compartments, including the nucleus, to effect appropriate cellular responses to proliferative, growth and stress signals.

ERK1/2 are directly phosphorylated and activated by the dual specificity mitogen-activated protein kinase kinases MEK1/2 (MAP2Ks, see Figure 2-1). Activating phosphorylation occurs on tyrosine 187 and threonine 185 within the activation loop of ERK1/2, leading to restructuring of the loop and activation site as well as reorganization of distal residues involved in interaction with upstream components and substrates (residue numbering from the human ERK2 protein, UniProt ID: P28482, (43-45)). MEK1/2 are the only known kinases capable of ERK1/2 activation, and pharmacological inhibitors exist that potently and specifically impair their ability to do so (46). In turn, MEK1/2 are phosphorylated and activated by the Raf family of MAP3Ks (MAP2K kinases) that form homo- and heterodimers when active (including A-Raf, B-Raf and C-Raf/Raf-1). Even more distally, Rafs are activated by Ras family GTPases (HRas, KRas and NRas), which are induced downstream of activated cell surface receptors or parallel signaling phenomena, or in the case of many cancers, mutated to a conformation that confers high levels of activity (47,48). Idiosyncratic and combinatorial utilization of Raf and Ras

proteins occurs as a consequence of cell type-specific expression and unique system wiring, the full determinants of which are an area of continuing research (49,50). Scaffolds that promote rapid, sequential pathway activation by binding multiple tiers of these signaling cascades also aid in appropriate subcellular localization for localized targeting of substrates (51).

Most studies to date have examined the role of ERK1/2 in a single cell type in response to a single stimulus, and frequently on a single downstream target or process. Different ligands can dictate differential signaling responses through ERK1/2. In one striking example, stimulation of ERK1/2 in the PC12 neuronal cell line by epidermal growth factor (EGF) induces a transient burst of activity and promotes proliferation (52-55). However, treatment with nerve growth factor (NGF) also stimulates ERK1/2, but activation is sustained over hours and results in differentiation. Both of these processes are mediated by induction of ERK1/2 activity (56). The basis for these differences lies in distinct signaling cues downstream of activated receptors, resulting in ERK1/2 activation, pathway feedback and cessation of signaling following EGF but not NGF treatment (57). This highlights how differential use of pathway components can impact signaling kinetics of the ERK1/2 pathway, yielding drastically different phenotypic outcomes.

Other studies have been data-driven, accruing large volumes of information about phosphorylation events but reaching little consensus on their meaning (58,59). Recapitulation and characterization of the consequences of combinatorial stimuli, as might be found in living tissues, has been relatively limited. Approaches to holistic interrogation are still restricted, but recent advancements in phosphoproteomics and substrate screening have made some headway into our understanding of cell-wide responses to stimulus and provided a framework upon which to deepen our understanding of global cellular responses to pathway activation (60-62).

Prediction of ERK1/2 phosphorylation sites from primary amino acid sequence was initially based on *in vitro* substrate specificity against oriented peptide libraries (63-65) and has been augmented in recent years by comparison of observed *in vivo* substrate sites numbering into the thousands (66,67). Consensus substrate sites are generally a serine or threonine immediately succeeded by a proline, or sometimes take the form Pro-X-Ser/Thr-Pro, where X may represent a neutral or basic amino acid.

Further support of *in silico* predictions can be derived from ERK1/2 binding sites within putative substrates. At least two clear amino acid motifs exist to support interaction between

ERK1/2 and their substrates, activating kinases and scaffolds. One is the Phe-X-Phe-Pro motif, usually present C-terminal to the site of phosphorylation (68). Alternatively, the D motif typically appears as several positively charged and hydrophobic residues, usually located N-terminally to the phosphoacceptor site (69,70).

In addition to being universally expressed, ERK1/2 are distributed broadly throughout many intracellular compartments. Nearly half of inactive ERK1/2 is bound on microtubules under basal signaling conditions (71) and in most unstimulated cells, some fraction of ERK1/2 resides in the nucleus (72). Following activation, ERK1/2 can move rapidly and transiently to the nucleus, presumably to target nuclear-localized substrates such as transcription factors. Nuclear import of activated ERK2 can occur via active transport through the nuclear pore complex, potentially aided by import factors (73,74). Inactive ERK2 can diffuse passively into the nucleus in the absence of energy and other factors, and binds to several components of the nuclear pore complex (60,75,76). Export is regulated by the export factor CRM1, and may depend on interactions between ERK1/2 and MEK1, the latter of which bears a strong nuclear export sequence (73,77).

AN OVERVIEW OF EUKARYOTIC MESSENGER RNA TRANSCRIPTION

IN THE CONTEXT OF CHROMATIN

Eukaryotic DNA is organized in the nucleus through interactions with histones into the basic repeating unit of chromatin, the nucleosome (78). Wrapping of DNA allows for the orderly compaction of over 2 meters of DNA into the volume of the nucleus, and provides an underlying protein structure that can be controlled by reorganization and post-translational modification. The nucleosome consists of ~147 bases of DNA, wound twice around a core of 8 canonical histone proteins (two copies each of H2A, H2B, H3 and H4), often with a more loosely tethered H1 associated outside of the central core (see Figure 2-1 for a cartoon of the crystal structure of the core histone octamer bound to DNA). The histone proteins are small, typically under 20 kDa in size, and highly basic in charge. Electrostatic interactions between histones and the negatively charged phosphate backbone of DNA are primary mediators of DNA-protein interactions, but hydrogen bonding, as well as nonpolar interactions with deoxyribose groups and solvent-

mediated contacts contribute, resulting in strong binding between DNA and proteins within the nucleosome (78,79).

The N-termini of the core histones, known as tails, are disordered and protrude outward from the center of the nucleosome (78). They are enriched for sites of post-translational modification (PTM), including relatively common PTMs such as acetylation, methylation, phosphorylation and ubiquitination (80). Chromatin is repressive to transcription in its native state (81,82). Modification and movement of nucleosomes to facilitate or further prevent interaction between transcription machinery and target DNA is a highly regulated process contingent on modifications to histone tails as well as the globular core of histone proteins (83,84). Modifications to chromatin can alter the binding affinity between histones and DNA, as well as serve as docking sites for regulatory proteins. Further regulatory complexity is conferred by alternate usage of paralogous noncanonical histone proteins (85).

To provide an additional level of regulation, chromatin can adopt a higher order structure known as heterochromatin that results in silencing of the encompassed DNA. This higher order folding generates the basis for even more complex structures (86). Heterochromatinization is an essential step in initiation of the cell cycle to ensure regulated division of chromatin into daughter cells. On a smaller scale, it regulates events like silencing of X-chromosomes as well as individual genes that are actively repressed. Formation of heterochromatin is specific to cell type and signaling state and is tightly regulated. Histone modifications, as well as ATP-dependent chromatin remodeling complexes feed into heterochromatin formation and maintenance. Conversely, euchromatin refers very generally to DNA that is in an "open" conformation, and can potentially be transcribed in response to an initiating stimulus (86). There is emerging evidence that complex, long-range interactions occur between actively transcribed genes and distal regulatory elements, often found thousands of bases away from the promoter or on another chromosome entirely (87).

In eukaryotes, transcription of tailed and capped transcripts including messenger RNAs (mRNAs), microRNAs (miRNAs) and small nucleolar RNAs (snRNAs) is mediated by RNA polymerase II (RNAPII; (88)). RNAPII is an incredibly complex and dynamic structure. Assembly and activation of the RNAPII holoenzyme at promoters occurs as a stepwise process. General transcription factors (including TFIIA, B, D, E, F and H) recognize specific promoter

elements on DNA, and bind sequentially with RNAPII to form the preinitiation complex (88,89). In vitro, the assembled preinitiation complex is extremely stable (90,91). Each of the general transcription factors performs a role vital in DNA recognition, DNA preparation for transcription and/or subsequent enzymatic activation of the RNAPII holoenzyme (88). Without addition of further factors, this complex is sufficient for productive transcription *in vitro*. More efficient activation is achieved by association of coactivators and post-translational modification of RNAPII subunits.

At this stage, just downstream of promoter clearance of RNAPII, a phenomenon known as promoter-proximal pausing often occurs (92). Chromatin immunoprecipitation (ChIP) studies of RNAPII localization reveal high levels surrounding the promoter. Several studies have demonstrated that elongation near the promoter is inefficient, collectively suggesting that promoter escape and productive transcription are regulated following preinitiation complex assembly (93,94). Complexes that induce pausing are recruited at this stage (93,95). Conversion to fully-active RNAPII requires the positive transcriptional elongation factor b (P-TEFb), a complex containing kinase activity that targets RNAPII and complexes that maintain pausing, collectively releasing RNAPII from repression (96). Promoter-paused RNAPII is resident at a large number of genes, especially those encoding genes responsive to signaling cues, including key components of the ERK1/2 pathway (97-99). This has been hypothesized to establish a low level of cycling transcription of these genes, enabling them to be rapidly responsive to external cues.

Phosphorylation of the disordered C-terminal domain (CTD) of Rpb1, the largest subunit of the RNAPII complex, is a critical step in transcriptional activation. In humans, the CTD is comprised of 52 heptad repeats of the sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (100). In the preinitiation complex, the CTD of Rbp1 is unphosphorylated (101). Phosphorylation of serines 2 and 5 is associated with a transcriptionally active RNAPII, and serine 5 phosphorylation is depleted as RNAPII moves through downstream genes (102). Several cyclin dependent kinases have been linked to this process. CDK7 is a component of the general transcription factor TFIIH complex, mediating phosphorylation at the promoter. CDK8 and CDK9 are part of the Mediator and P-TEFb complexes respectively, and appear to maintain phosphorylation of the CTD

throughout transcription downstream of the transcription start site. Other kinases have been shown to be capable of CTD serine phosphorylation, including ERK2 (103,104).

Coactivators and elongation factors enhance the ability of RNAPII to efficiently transcribe genes. Located at promoters, general coactivators include the TATA binding protein-associated factors (TAFs) and the Mediator complex. TAFs are enriched for domains that recognize histone modifications or DNA elements and are used combinatorially, seemingly to aid in the interaction of the preinitiation complex with chromatin at the promoter (105). Mediator makes critical contacts between the CTD of RNAPII and promoter-associated transcription factors, transmitting additional regulatory cues to RNAPII from the immediate environment (106). Many elongation factors, such as P-TEFb, are critical in ensuring processivity of RNAPII through to the 3' untranslated region of a target gene (107). The final step of transcription, termination, is similarly highly regulated and depends on yet another set of recruited factors that enhances disassembly of RNAPII and removal from target DNA (108).

Each of the transcriptional processes mentioned relies upon a complex underlying epigenetic framework of histone modifications and encoded DNA elements, some of which will be discussed in the following sections as they relate to the rest of this work.

HISTONE MODIFICATIONS IN PHYSIOLOGY AND PATHOLOGY

Histone modification represents an area of research that exploded with the advent of next-generation sequencing technologies. In the early days following the completion of the human genome project and subsequent cataloguing of enormous data sets defining global localization of histone modifications, it was widely anticipated that a "histone code" could be derived, which would serve as a basis for comprehensively predicting how the underlying DNA was transcriptionally regulated. To date, global maps of modifications in many systems have still not significantly furthered our understanding of this issue, and the consequences of small subsets of modifications are often used as "tells" for whether genes can and will be transcribed (109,110). Complicating this issue further is the fact that the complex nature of transcriptional regulation means that the causality of histone modification on transcriptional events is still poorly defined (111).

Specific modifications can inform upon the activity of the associated gene. One modification that is widely utilized to predict transcriptional activity is lysine methylation. Lysines can be covalently and reversibly mono-, di- and tri-methylated at the ζ-amine, modifications typically mediated by proteins containing a SET methyltransferase domain, initially identified in Drosophila (Su(var) 3-9, Enhancer of zeste and Trithorax; (112,113)). Different subsets of SET-domain containing methyltransferases are responsible for mediating lysine methylation at distinct residues. Lysine methylation occurs predominantly at histone tails on H3 and H4, and different sites have different consequences for the underlying DNA. For instances, dimethylation at H3K9 and trimethylation of H3K27 are both associated with silencing of the underlying DNA (80). Monomethylation of H3K4 marks enhancer elements, loci distant from promoters that aid in transcriptional activation of often-cryptic target genes (114,115). Trimethylation of H3K4 (H3K4me3), which will be a topic of focus in this document, is strongly associated with the 5' end of genes that are being actively transcribed in species ranging from yeast to humans (116-118).

Budding yeast and mammals differ in the machinery at their disposal to trimethylate H3K4, likely reflecting complexity gained through tissue differentiation. *S. cerevisiae* encodes only one H3K4 methyltransferase, Set1, which as a part of the COMPASS (*com*plex of *p*roteins *a*ssociated with *S*et) complex is responsible for mono-, di- and trimethylation of H3K4 (119-121). Surprisingly, Set1 is not essential, but loss results in aberrant expression of many genes and delayed doubling time (119). COMPASS is associated with elongating RNAPII, with deposition of multiple methyl groups an apparent consequence of repeated passage of RNAPII over target genes during multiple rounds of transcription (122,123). This may lay groundwork for positive feedback regulation of the marking, as at least one subunit of the COMPASS complex binds to H3K4me2/3 (124).

Mammals bear at least six complexes that resemble COMPASS in subunit composition and catalytic function (125). These include the Set1A and Set1B complexes, and the Mixed Lineage Leukemia complexes (MLL1-4). Set1A/B regulate global levels of H3K4 trimethylation (5), while the four MLL complexes regulate small subsets of genes that are typically associated with pluripotency and development, or enhancer monomethylation (114,126,127). Bolstering their importance at a relatively limited subset of genes, MLL1-4 are less abundant in HeLa nuclei

than the Set1A/B complexes (128). Genomic targets of each complex appear to be distinct (129), and loss of single complexes cannot be offset by other family members (130), although this matter has not been comprehensively addressed by global ChIP experiments.

All six complexes share a structural core of 4 proteins: Wdr5, Rbbp5, Ash2L and Dpy30, commonly termed the WRAD module. Compositionally, the Set1A and Set1B complexes appear to differ from one another only in distinct usage of their catalytic subunit, and the presence of an additional subunit in the Set1B complex (Bod1; (128)). Figure 2-2 details the usage of different subunits by the Set1A/B and MLL complexes. Human Set1A and Set1B are large proteins, each spanning over 1700 amino acids in length, with a C-terminally located SET methyltransferase domain. Overall, Set1A and B share 39% sequence identity and 56% similarity, although the majority of
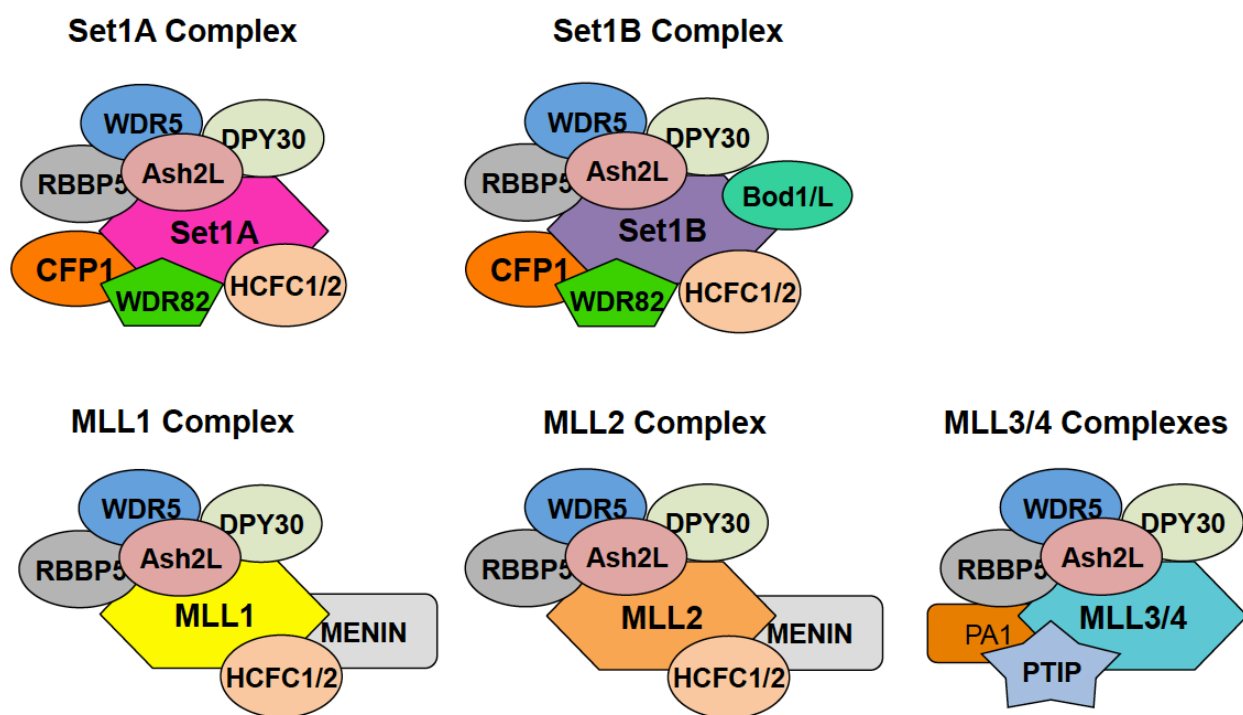
Figure 2-2: Comparative subunit composition of human COMPASS-like H3K4 trimethylating complexes. Analysis of subunit composition of Set1-like H3K4 methyltransferase complexes from HeLa cells. Set1A/B complexes share common structural elements with the MLL complexes (WDR5, RBBP5, Ash2L and DPY30). CFP1 and WDR82 are specifically associated with Set1A and Set1B. Modified from van Nuland, et al. Mol Cell Biol. 2013;33:2067-2077 and reprinted with permission from the American Society for Microbiology.

conservation is restricted to the SET domain through to the C-terminus, which are 85% identical and 95% similar. *S. cerevisiae* Set1 shares 35% identity with human Set1A and 37% identity with human Set1B (129).

The causative role that H3K4me3 plays in mammalian transcription is still an open question despite a clear global enrichment at actively transcribed promoters and some evidence that the marking can participate in cancer development. Knockout lines of many individual subunits of COMPASS family complexes are viable, arguing for a more passive role for H3K4me3, perhaps as a consequence of transcription (5,131). *In vitro*, H3K4me3 stimulates transcription from a chromatinized template using HeLa nuclear extracts (132). H3K4me3 is recognized by multiple transcriptional regulators (133), including TAF3, a core component of the TFIID general transcription factor complex whose binding to promoters is an essential step in RNAPII recruitment (88,134). Followup studies examined genes inducible through recruitment of the transcription factor p53 following doxorubicin treatment (135). Binding between the plant homeodomain finger (PHD finger, which binds methylated residues; (136)) of TAF3 and H3K4me3 is a prerequisite for TFIID recruitment to most H3K4 trimethyl-enriched promoters. Furthermore, this interaction is essential for the full transcriptional response to doxorubicin-induced genotoxic stress at many p53 target genes (135). H3K4me3 accumulates at p53 target genes following treatment with doxorubicin, and loss of H3K4me3 by knockdown of WDR5 prevents recruitment of TFIID and mRNA accumulation. To date, this represents the best evidence that H3K4me3 directly contributes to normal transcriptional activation in mammals.

PHD fingers are domains that can recognize H3K4me3, and in some cases other methylated histone motifs (137). They coordinate two Zn^{2+} ions via a characteristic Cys4-His-Cys3 motif, and form an aromatic cage that binds the methylated histone. They are present in various epigenetic regulators (133,138), including enzyme complexes that deposit and remove H3K4 methylation from nucleosomes (139,140). H3K4me3 levels are dysregulated in a variety of cancers (133). Driver mutations for a subset of hematopoeitic malignancies are translocations of the nuclear pore complex protein Nup98 with the PHD fingers of at least two distinct proteins, the H3K4 demethylase Jarid1A and PHF23, a protein of unknown function (139). In this case, mutation of the PHD finger prevents transformation, indicating that the interaction with

methylated histone is a prerequisite for cancer development. This further supports the idea that H3K4 trimethylation may play a role in promoting transcription.

DNA METHYLATION IN PHYSIOLOGY AND PATHOLOGY

The distribution of DNA bases – adenine, guanine, thymine and cytosine – throughout the eukaryotic genome is non-random, despite the fact that protein-coding sequences account for less than 2% of total genetic material (141). DNA sequence can directly inform the arrangement and spacing of nucleosomes (142-145). DNA elements surrounding the transcriptional start site provide binding information for both general and specific transcription factors (89,146). One widespread genetic phenomenon is the presence of CpG islands (CGIs) that occur at approximately 70% of all annotated gene promoters (147). CGIs are regions that average 1000 base pairs (bps) in length and are highly enriched for CpG dinucleotides (9). CpG dinucleotides are otherwise found at lower than expected ratios throughout the genome. Cytidine residues within these dinuclotides can be methylated at the 5 position. Global de-enrichment of this dinucleotide has been attributed to the propensity of methylated cytidine to be spontaneously deaminated, resulting in a single base change to thymine following semi-conservative gene replication during mitosis (9,148,149).

Cytidine methylation in mammals is mediated by the DNA methyltransferases (DNMTs). DNMT3a and DNMT3b are thought to be responsible for *de novo* methylation, a vital developmental mechanism that serves to define cell and tissue identity by silencing expression of nonessential genes within a tissue (150,151). It serves essential roles in the silencing of imprinted genes and transposable elements within the genome and is essential for genomic stability (152-154). The predominant maintenance methyltransferase, DNMT1, recognizes hemi-methylated sequences following replication and methylates the cognate strand (12,155). Regulated demethylation exists, and is an area of active research (156), further underscoring the dynamic nature of CpG methylation.

In adult tissues, most CGIs are unmethylated (9). Recent work has shed light on functions for unmethylated CGIs in promoting transcription by destabilizing nucleosome binding to DNA through nonfavorable DNA sequences (157,158). In the specific context examined, nucleosome

de-enrichment permitted binding of the transcription factor Sp1 to promote transcriptional activation (157) and P-TEFb recruitment initates paused RNAPII at these genes (99). This mechanism is likely to extend to other systems and probably many CpG-binding transcriptional regulators (159).

Dysregulated methylation of CGIs is widespread in cancer (160). DNA methylation is a contributing event in silencing programs that promote heterochromatin formation. Methylation recruits histone modifiers and chromatin remodelers, resulting in gene silencing through heterochromatin formation, often contributing to cancer progression through silencing of tumor suppressor genes.

ERK1/2 IN TRANSCRIPTIONAL REGULATION

The topic of MAPK-driven epigenetic regulation is not underrepresented in the literature, but reaching a consensus of steps downstream of MAPK activation between initiation of signal and consequent transcriptional effects is severely hampered by the sheer complexity of events, and perhaps by transcriptionally repressive or activity-independent functions. Observed transcriptional consequences are dependent on manner and duration of stimulus, cell type, environmental context and experimental approach (13). It is increasingly apparent that transcriptional regulation occurs at virtually every possible level of regulatory input. This includes phosphorylation of transcription factors, histone and chromatin modifying enzymes, transcriptional machinery and potentially some kinase activity-independent functions (Figure 2-3).

Early studies of the role of ERK1/2 in transcriptional regulation focused on phosphorylation-dependent transcription factor activities. Following translocation of ERK1/2 into the nucleus in response to serum activation, these kinases phosphorylate the transcription factor Elk-1 which can then bind to the serum response factor (SRF) to promote the expression of immediate early genes (161-163). To date, many transcription factor substrates and interdependent transcriptional coactivators have been uncovered (15,164-166). Additionally, activation of ERK1/2 substrate kinases can further support transcriptional outcomes (15).
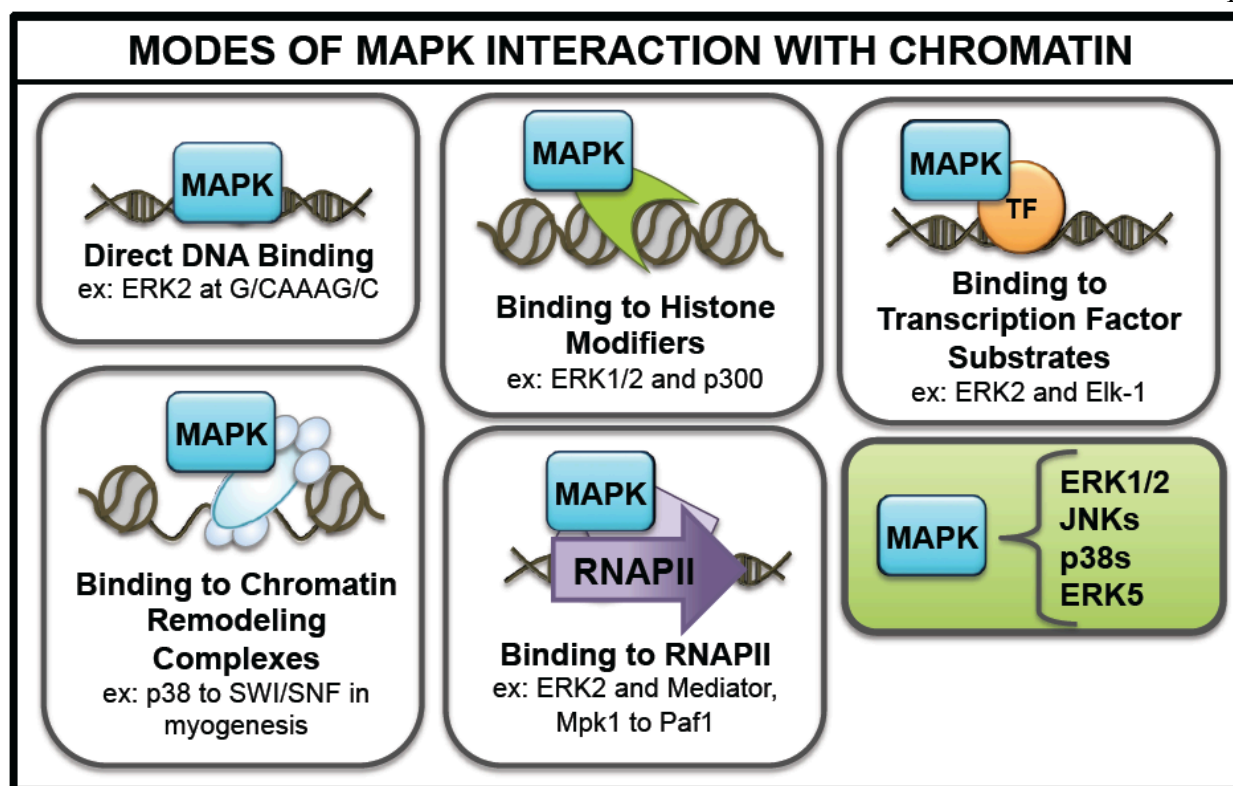
Figure 2-3: MAPKs are involved in many levels of transcriptional regulation. Modified from Current Opinion in Cell Biology, 25(2), A. M. Klein, E. Zaganjor & M. H. Cobb, Chromatin-tethered MAPKs, 272-277., Copyright 2013, with permission from Elsevier.

First reports of any MAPK physically occupying the promoter of target genes came in 2004, when a group reported recruitment of members of the p38 family to the promoters of muscle-specific genes expressed in differentiating myoblasts (167). This had been predicted by multiple groups (15,168). A more comprehensive study of MAPK interaction with chromatin was later completed in budding yeast, where multiple MAPKs were found to bind to many target genes following stimulation by their cognate activation signal (169). This included the p38 homolog Hog1 in response to osmotic stress, and the pheromone-responsive Fus3 and Kss1, as well as their scaffolding factor Ste5, following pheromone stimulation.

Work from our own lab subsequently demonstrated recruitment of ERK1/2 and their upstream activators MEK1/2 to the chromatin of the insulin promoter in pancreatic beta cell lines following stimulation by glucose (170). Chromatin recruitment was rapid (strongly detected within ten minutes) and transient (back to basal within 30 minutes), and pharmacological

inhibition of ERK1/2 activation reduced not only recruitment of kinases but also substrate transcription factors necessary for full induction of the insulin gene. Follow-up work in these same systems detected recruitment of ERK1/2, p38 and JNK to the insulin promoter in response to treatment of cells with the proinflammatory cytokine interleukin 1β (IL-1β) under hyperglycemic conditions (171). This stimulus resulted in promoter localization of these kinases that was sustained for at least four hours, refractory to ERK1/2 inactivation, and resulted in recruitment of a distinct complement of transcription factors and diminished transcription from the insulin promoter than glucose treatment alone. This clearly demonstrated that unique MAPK recruitment and transcriptional outputs may result from a single promoter, depending upon context.

Adaptation of chromatin immunoprecipitation (ChIP) to a 96-well plate format allowed investigators to multiplex interrogation of protein factor binding to many genomic loci over a range of time points (172). This was a particularly useful first step in forming a deeper understanding of inducible kinase-chromatin interactions, and to date has been used to probe recruitment of factors in the ERK1/2 MAPK pathway to promoters in response to insulin (173) and serum (174). Findings from these works expanded on established knowledge that upstream kinases (170,175) and scaffolds (169) are, with MAPKs, among factors recruited to target genes in response to stimuli. This includes members of the MAP3K class of enzymes, including B-Raf, and regulatory molecules even further upstream, including the scaffold GRB2, the guanine nucleotide exchange factor SOS and signal-initiating receptor tyrosine kinases themselves (173,174,176). Notably, pathway inhibition with MEK1/2 inhibitors reduced stimulus-driven recruitment of all of these factors to target genes, suggesting ERK1/2 activity is a necessary prerequisite to this process. It is possible this occurs through some sort of positive feedback signal, or active MAPKs may serve as tethering factors between chromatin and upstream regulatory molecules. One of the more interesting findings from this work was that signaling module recruitment was not limited to the promoter but persisted throughout the coding sequence of gene targets, implying sustained interaction of these molecules with transcriptionally active RNAPII (173,174).

These data are not alone in suggesting that MAPK interactions with RNAPII persist subsequent to promoter clearance, and may in fact be important for regulation of transcriptional

elongation and termination. ERK1/2 association with coding regions of immediate early genes such as Egr-1 is contingent on activity of P-TEFb and elongating RNAPII, suggesting interaction between ERK1/2 and the RNAPII holoenzyme (174). In yeast, the Hog1 MAPK maps to the open reading frames (ORFs) and 3' untranslated regions (UTRs) of osmotic stress-responsive target genes (177). Hog1 interacts specifically with the elongating form of RNAPII (bearing serine 5 phosphorylation in its C-terminal domain). Uncoupling of salt-responsive genes from their native promoters supports post-initiation roles for activated Hog1 in enhancing processivity of RNAPII through target genes and accumulation of associated mRNAs. Finally, the presence of an intact 3' UTR is essential for recruitment of Hog1 to ORFs of target genes, suggesting that some feature of 3' UTRs contributes to proper recruitment of transcription machinery and/or signaling molecules. Similar studies of the yeast Mpk1, a MAPK that regulates cell wall shape and integrity and most closely resembles mammalian ERK5, indicate that Mpk1 interacts with the Paf1 transcriptional elongation complex to support transcription of full-length target ORFs by blocking premature termination that occurs in its absence (178). Most surprisingly, this does not require the kinase activity of Mpk1, as a catalytically inactive mutant is sufficient to support elongation.

Advancements in the technologies used to map interactions between proteins and DNA as well as increased knowledge of genomic structure and regulation has enabled even further dissection of genome-wide transcriptional regulation by MAPKs, and recent focus has been trained on ERK1/2. Interrogation of global binding sites for ERK1/2 has been undertaken in embryonic stem cells (ESCs). ERK2 was ChIPed from human ESCs and global binding sites were compared to histone and protein markers for transcriptionally active and inactive genes (110). Knockdown of ERK2 results in a dramatic reduction in the expression of crucial pluripotency markers, although differentiation capacity was not assayed. ERK2 localized mainly to the promoters of transcriptionally active genes, overlapping substantially with histone trimethylation at H3K4 and acetylation H3K27 and H3K9 at gene targets, all marks of a transcriptionally competent promoter. ERK2 binding was particularly enriched at the promoters of noncoding genes, and genes that regulate metabolism, the cell cycle and transcriptional regulation. Motif identification within ERK2-bound promoters uncovered enrichment for binding motifs of many known ERK1/2 substrate transcription factors, including Klf4, Sp1, CREB1 and

Elk1. Interaction with these transcription factors may mediate recruitment of ERK2 to chromatin, or be dependent on locally active ERK1/2 for binding. Of the identified factors, Elk1 was employed to better define the distinct and combinatorial functions of transcription factor and ERK2 action at promoters. Genes that are bound and regulated by both proteins together segregated into distinct ontological classes than genes regulated by each component individually. These classes exhibit their own unique host of associated transcription factors, although all of these classes support self-renewal and pluripotency of hESCs.

Further mechanism underlying ERK1/2-regulated pluripotency of ESCs was uncovered in work from the Reinberg lab this year. Histone modifications and transcriptional capacity in ESCs are generally regulated by the same underlying mechanisms as differentiated cells, but ESCs exhibit a unique combination of histone markings at promoters of many developmental genes known as bivalent domains (179). This refers to the simultaneous modification of promoter nucleosomes with the activating histone modification H3K4me3 as well as the repressive transcriptional mark H3K27me3. Bivalent domains represent the capacity of these genes to rapidly adopt either transcriptionally active or repressed states depending on differentiation cues (180,181). In ESCs, H3K4me3 is deposited on these promoters by the mixed lineage leukemia (MLL) 2 complex (127) and H3K27me3 is mediated by the polycomb repressive complex 2 (PRC2; (182,183)). ERK2 and PRC2 components were found together at a subset of paused developmental genes in ESCs (104). ERK1/2 activity was required for H3K27me3, and these promoters appeared to be devoid of the general transcription factor TFIIH. This observation led the authors to hypothesize that perhaps ERK1/2 can function to phosphorylate the CTD of RNAPII in the absence of activating kinase activity from TFIIH, although mechanistic evidence for this was unconvincing. It is important to note that ERK1/2 activity can contribute to activating and repressive effects collaboratively in the same system to achieve a phenotypic outcome, and that it is capable of driving this effect through histone modification.

Where and when assembly of chromatin-associated ERK1/2 signaling modules occurs are also major outstanding questions in the field. ERK2 can bind directly to DNA at a defined consensus motif, as shown both *in vitro* and *in vivo* (104,184). Both reports indicate that ERK2 binding is repressive to transcription in different contexts. It will be interesting to see if this

holds true for other systems. Global analysis of ERK1/2-bound sites following ChIP indicates binding sites in the tens of thousands with a complex mixture of enriched binding motifs. This includes those of substrate transcription factors and frequently excludes the direct binding consensus, indicating that many interaction sites are indirect ((110); our unpublished data). Taken together with immunofluorescence studies, a pool of ERK1/2 is resident in the nucleus under basal simulation and inhibitory signaling conditions (75). This may represent a subset of molecules that is primed for rapid activation and perhaps relocation to target genes, but technological limitations have prevented resolution of this issue.

MAPKs can direct activity and targeting of histone- and chromatin-modifying activities, with an ever-expanding list of substrates and regulated target gene sets depending upon cell type and stimulus. Chromatin remodeling and histone modification mediated by the yeast Hog1 MAPK in response to osmotic stress, described above, is a particularly comprehensively-studied example, where the kinase contributes to multiple steps at salt-responsive genes including transcription factor recruitment (185,186), histone modification (187), chromatin remodeling (188) and support of RNAPII recruitment and elongation (177,189,190).

Similarly, ERK1/2 and other mammalian MAPKs are instrumental in promoting and repressing various histone modifications and chromatin remodeling complexes to support directed physiological outputs, and transcriptional misregulation by these pathways has been implicated in disease states. For example, the ERK-MAPK pathway has been implicated in regulation of histone modifications such as H3K4me3 in *Drosophila melanogaster* development (191) and H3K28 phosphorylation through downstream kinases that promote transcription by blocking PRC2 complex recruitment to a neighboring residue (192). p38 phosphorylation of a substrate transcription factor during myogenesis promotes recruitment of H3K4 trimethylation activities to differentiation-specific promoters, driving muscle development (193). p38 additionally promotes transcription of gene targets during myogenesis through phosphorylation and recruitment of a member of the SWI/SNF family of chromatin remodelers, which are ATP-dependent (167). Some evidence exists for substrate targeting of other ATP-dependent chromatin remodeling complexes by ERK1/2, leading to distinct transcriptional outcomes in response to external cues (194), though the physiological implications are less clear.

ERK1/2 activation downstream of serum and other mitogenic stimuli orchestrates a number of effects that likely work in concert to promote transcription of immediate early genes. In response to EGF, ERK1/2 phosphorylate and activate the lysine acetyltransferase p300 at multiple sites, which promotes p300 activity and target gene expression (16). p300 catalyzes histone acetylation at the immediate early gene *c-fos* following Elk-1 recruitment, promoting nucleosome eviction from the promoter and target gene expression (195). Phosphorylation of Med14, a subunit of the Mediator complex, is required for RNAPII recruitment at ERK/Elk-1 directed target genes (17). Loss of any one of these events impairs transcription of immediate early genes, suggesting they act collaboratively to support targeted transcription. Another study showed that this type of paradigm is cell-type specific, but other tissues and stimuli may share common regulatory features, such as adoption of transcription factors other than Elk-1, to achieve the same ends (196).

ERK1/2 activity has been implicated in global transcriptional remodeling events that depend on DNA methylation. Genomic reorganization due to sustained ERK1/2 pathway activation may be refractory to reestablishment of the native state following removal of the initiating stimulus (197). Sustained activation of Ras has been shown to lead to silencing of a specific subset of genes, including the pro-apoptotic Fas gene (21,198). Silencing occurs through DNA methylation and recruitment of a series of histone modifiers and chromatin remodelers, and sustained silencing is dependent on DNMT1 as well as ERK2 (21,22).

CFP1 DIRECTS MULTIPLE TRANSCRIPTIONAL OUTCOMES IN THE CELL VIA EPIGENETIC REGULATION

CXXC Finger Protein 1 (CFP1) is a ubiquitously-expressed epigenetic regulator in animals and fungi (140). It is essential for mammalian development; knockout in the mouse results in peri-implantation lethality between embryonic days 4.5-6.5 (Carlone & Skalnik 2001). It is required for at least two indispensable developmental programs carried out in animals: DNA cytosine methylation and Histone 3 Lysine 4 trimethylation, discussed at length in the following two sections.

CFP1 interacts with the *de novo* DNA methyltransferase DNMT1 to direct cytosine methylation and gene silencing

CFP1 was initially identified in a phage display screen for proteins that preferentially bind to unmethylated CpG dinucleotides (140). Primary sequence analysis revealed that CFP1 contains a zinc finger-CXXC (ZF-CXXC) domain, at the time already known to be a DNA-binding element due to its presence in a host of other CpG-directed proteins. The domain is a tandem repeat of the amino acid sequence Cys-Gly-X-Cys-X-X-Cys and binds to two zinc ions (199). To date, the list of ZF-CXXC domain-containing proteins comprises functionally diverse epigenetic regulators including the Tet family of DNA demethylases, the KDM2 histone lysine demethylases, and the Set1-related H3K4 methyltransferases mixed lineage leukemia 1 and 2 (MLL1/2; reviewed in (2)).

The ZF-CXXC domain of CFP1 (amino acids 160-209; See Figure 2-4) has been well characterized, and primary sequence of the domain is highly conserved (200). *In vitro*, experiments have shown CFP1 can bind to DNA, exhibiting a strong preference for sequences that contain unmethylated CpG dinucleotides (201). A crystal structure of the ZF-CXXC domain of CFP1 complexed to multiple distinct CpG-containing DNA probes was solved in an attempt to understand the basis of non-methylated CpG specificity (200). This structure revealed that this domain fits into the major groove of target DNA, making extensive electrostatic contacts with both the negatively-charged DNA backbone and the CpG motif itself. Specificity towards
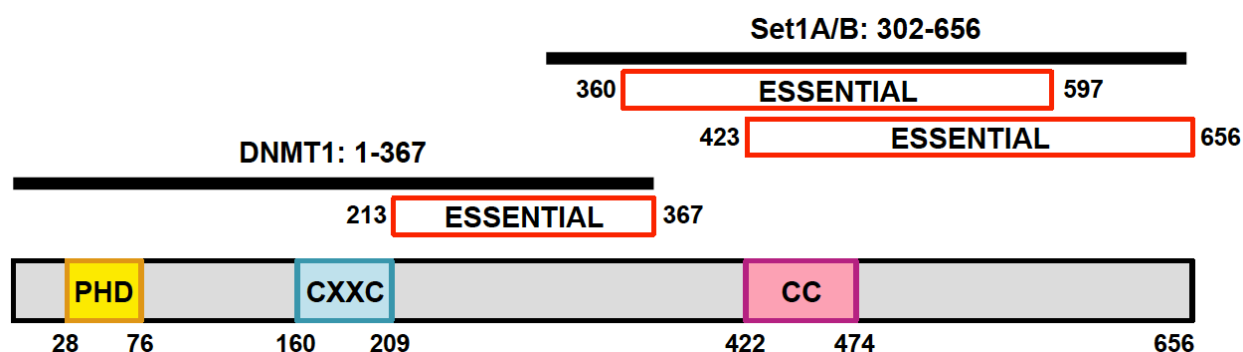
Figure 2-4: Domain structure of human CFP1. The canonical isoform of CFP1 (UniProt ID: Q9P0U4-1) is 656 amino acids in length and is comprised of at least three structured domains. A splice isoform exists, where K340 is replaced by the sequence "KVMER." PHD finger: Plant Homeodomain, binds methylated H3K4; CXXC: Unmethylated CpG binding domain; CC:

putative coiled-coil domain. Minimal interaction domain for DNMT1 and Set1A are shown, and regions within those domains that are essential for binding are marked.

nonmethylated DNA appears to be a consequence of the location of the key binding residues Ile-199/Arg-200/Gln-201 within an inflexible loop that is unable to sterically accomodate cytosine methylation.

Genomic deletion of CFP1 results in embryonic lethality in mice (202). CFP1 is expressed consistently throughout development, implantation and embryogenesis (203). Death in the null animals coincides with a global reprogramming event where all genomic methylation is ablated, setting the stage for tissue specific *de novo* DNA methylation patterning during embryonic implantation, mediated by DNMT1 (11,150,204). Defects associated with CFP1 loss are more severe than a DNMT1 knockout; mice lacking DNMT1 survive until E9.5-11 (11). This suggests that DNA methylation defects associated with CFP1 loss do not fully account for the observed phenotype, indicating that other CFP1-dependent activities, include histone methylation (discussed in the following section), contribute substantially to the early lethality.

Embryonic stem cells isolated from CFP1 knockout mice self-renew but cannot differentiate following removal of the cytokine leukemia inhibitory factor (LIF) from the culture medium (24,205). These cells exhibit a 70% global deficit in DNA methylation, with detectable hypomethylation at satellite repeats, stably methylated gene promoters and imprinted genes. *In vitro* assays with crude lysates exhibit normal DNA methyltransferase activity towards unmethylated probes but more than a two-fold reduction in activity towards hemimethylated targets, suggesting impairment of maintenance, but not *de novo* methyltransferase activity. This decreased activity can be accounted for by the steady state loss of nearly half of DNMT1 protein in the cells (24). Although CFP1 can promote transcriptional transactivation when associated with the Set1 complexes ((140); see next section), transcriptional changes do not account for the loss of DNMT1 protein - in fact, DNMT1 mRNA is 1.5-fold higher in CFP1 knockout cells (206). CFP1-null cells exhibit globally decreased protein synthesis and reduced polysome levels, including decreased synthesis rates for DNMT1 protein. The half-life of DNMT1 is also reduced under these conditions (206).

Efforts to understand post-embryonic developmental functions for CFP1 were undertaken in zebrafish, which have a CFP1 homologue that is 60% identical to its murine counterpart (25). Oligonucleotides were designed that knocked down CFP1 expression in a dose-dependent fashion, and functional rescues were successfully undertaken with murine CFP1. Morpholino dosage correlated with defect severity, and at high doses was lethal. Following 24 hours of knockdown, significant loss of genomic DNA methylation was observed, consistent with reports in other systems. Animals with incomplete CFP1 knockdown exhibited highly-penetrant defects in yolk sac morphology and primitive erythropoeisis, and the majority exhibited cardiac edema. Irregularity in hematopoeitic development was an anticipated outcome, as inhibition of DNMT1 with the drug 5-azacytidine or histone deacetylases with Trichostatin A displayed similar phenotypes (207). Increased apoptosis at the site of early erythropoeitic development prompted follow-up studies in human myeloid cell line PLB-985 (208). Clonal populations depleted of CFP1 were unable to undergo stimulus-dependent differentiation. An increased doubling time was observed, although the high rate of apoptosis seen in zebrafish, as well as in mouse ESCs (24), was not recapitulated. These results, as well as the inability of mouse ESCs to undergo differentiation upon LIF removal (24), suggest that CFP1 is largely dispensable for the function of pluripotent cells, but may be generally required for DNA methylation-driven lineage commitment.

CFP1 and DNMT1 interact in cells (23). Based on studies with a series of truncation mutants, multiple regions of each protein apparently contribute to their interaction. Residues on CFP1 that are required for detectable interaction with DNMT1 were restricted to the N-terminus of the protein (fragments encompassing residues 1-123 and 103-367 both bind, which include both the PHD histone binding domain and CXXC domains; see Figure 2-4). Further truncations excluding the CXXC domain (between residues 103-213) do not bind. Direct binding of recombinant CFP1 and DNMT1 synthesized in *E. coli* was not detected, suggesting interaction between these two proteins is either dependent on other binding partners or upon post-translational modification (23). DNA methylation defects in mESCs were rescued by expression of CFP1 truncation mutants from both the N- and C-termini of the protein, including portions that are insufficient for DNMT1 binding (209). This may indicate that there is a low level of binding in some truncations that was not observed and is sufficient to rescue activities, or

alternatively that this system is more complicated than initially supposed. Multiple studies have failed to detect interaction between DNMT1 and the Set1 complexes, suggesting that these interactions occur with distinct pools of CFP1 (128,129,210), however the mechanism leading to linked phenotypes despite independent interactions is unclear.

A mechanistic rationale for interaction between CFP1 and DNMT1 has yet to be uncovered. Both proteins contain redundant DNA targeting domains (201,211), and a dependence on CFP1 for the genomic targeting or activity of DNMT1 has not been directly demonstrated. Furthermore, information regarding stoichiometry of interactions between CFP1 and DNMT1 has not been reported. One possibility is that interactions are restricted to a subset of targeted genes or responsive to specific signaling conditions. Studies to determine whether CXXC domains from other proteins can functionally compensate for one another revealed that the DNMT1 CXXC domain functioned more like that of MLL1 than the same domain from CFP1, implying clear differences between these factors that are not yet understood (212).

CFP1 appears to have acquired the ZF-CXXC domain in organisms that utilize DNA methylation to fine-tune gene expression. This domain is absent in eukaryotes that lack DNA methylation including budding and fission yeasts as well as the model nematode *Caenorhabditis elegans*, but is present in species that developed DNA methylation to support tissue specific gene expression, including fruit flies, zebrafish, mice and humans (9,49,213). Although this is one of many possibilities underlying species-restricted use of this domain, these findings support the importance of ZF-CXXC domain of CFP1 in appropriate catalysis of DNA methylation.

CFP1 directs Histone 3 Lysine 4 Trimethylation via the Set1A/B Complexes

Initial characterization of human CFP1 demonstrated that it was competent to transactivate transcription from a CpG-rich promoter in mammalian cells (140). Soon thereafter, the CFP1 homolog Spp1 was identified in the multisubunit COMPASS (*com*plex of *p*roteins *as*sociated with *S*et) complex in *S. cerevisiae* (119), the sole mediator of H3K4 trimethylation in this organism (121,214). Spp1 is dispensable for viability (119,214), but loss results in anomalous gene expression and growth delay similar to that observed in deletion strains of 5 out of 6 of the other COMPASS subunits (deletion of the sixth subunit, Cps35, could not be assayed because it is not viable; (119)). The most apparent gene expression changes were in de-

repression of telomeric (119) and rDNA (214) genes. However, at the time of publication of these data work, evidence for a role for H3K4 methylation in facilitating transcription was accumulating, and prior work implicated Set1 in transcriptional regulation of genes involved in cell growth, division, cycling and transcription (215,216). Similarly, CFP1 is associated with H3K4 methylation activities in mammalian cells. Of the six known COMPASS-like histone methyltransferase complexes that regulate H3K4 methylation, only the Set1A and Set1B complexes contain CFP1 (Figure 2-2) (128).

Much of the biochemical work undertaken to define the basic requirements for activity of the H3K4 methyltransferases was performed on the yeast COMPASS complex, as it is the sole mediator of H3K4 methylation in this system and based on subunit composition seemed like the most analogous simplified system. Spp1/CFP1 appears to be a dedicated subunit in both complexes (128,212), although loss in cells does not impact formation of the remainder of the COMPASS complex (121) and is dispensable for activity *in vitro* (217). Deletion of Spp1 from *S. cerevisiae* has a severe and specific influence on global levels of H3K4 methylation, with an 80% reduction in levels of trimethylation and a 10% disruption in dimethylation (121). Knockout of the Spp1 homolog from *Schizosaccharomyces pombe* results in the complete loss of this marking (218), a phenomenon that can be attributed to differential domain structures of the COMPASS complexes in these two yeast species. As mentioned previously, CFP1 deletion in mice results in early embryonic lethality (202), but undifferentiated ESCs derived from these embryos have normal global levels of H3K4 methylation (210). However, upon induction of differentiation by LIF removal, H3K4 trimethylation accumulates to nearly four-fold basal levels over six days, suggesting that CFP1 somehow restricts Set1 complex activity.

Interaction mapping with the Set1A/B complexes identified a large fragment of the C-terminus of CFP1 that mediated binding (binding was detected with residues 302-656 and 361-656, but not shorter fragments, see Figure 2-4), and mutation of several conserved residues to alanine (C375 and C580 individually; YCS 390-392 together) ablated interaction with Set1A/B complexes but not with DNMT1 (23). In addition to the DNA binding CXXC domain, discussed in the previous section, CFP1 bears a PHD finger. *In vitro* binding studies with the Spp1 PHD finger demonstrated binding specificity towards di- and tri-methylated H3K4, but not the monomethylated species (124). Homology modeling of this domain based on the related protein

ING2 for which there is a crystal structure (219) indicates that interaction with the modified histone is mediated by hydrophobic base stacking interactions. Murine CFP1 binds to trimethylated H3K4 in a pulldown screen for histone-interacting proteins (8). The H3K4me2/3 binding capacity of CFP1 may act as a feed-forward loop to help drive accumulation of H3K4me3 at discrete points by tethering the histone methyltransferase complexes to genomic loci where di- and tri-methylated histones are already present.

These findings are consistent with genome-wide analysis of the role that CFP1 plays in localization of the Set1A/B complexes to chromatin and their role in supporting transcription. Loss of CFP1 in mESCs results in global disruption in the subnuclear distribution of H3K4me3 as well as Set1A. Typical localization of these signals occurs almost exclusively at euchromatin, but loss of CFP1 leads to redistribution of a pool of these signals to silenced chromatin. Rescue was not achieved with truncations of CFP1 encompassing large segments of the protein (including 1-481 and 302-656; the latter segment was previously shown to be sufficient for Set1 interaction (23,220)) or mutants that ablate binding of CFP1 to DNA or the Set1A/B complexes (220).

Large-scale ChIP-sequencing analyses were undertaken to define regions where CFP1 binds, and to gain insight on the global requirements for and consequences of association on histone and DNA methylation (3,4). Thomson, et al. immunoprecipitated CFP1 from mouse brain, and identified a 93% overlap of CFP1 binding and H3K4me3, and an 81% overlap with CpG islands (CGIs). To test whether CGIs were independently capable of driving H3K4me3 through recruitment of CFP1, an exogenous CGI consisting of an enhanced green fluorescent protein (eGFP) and a puromycin resistance gene (both of which are CpG-rich) were artificially inserted into the genome of an ESC line. These artificial sequences accumulated high levels of H3K4me3, and were free of RNAPII. This is important, as it suggests that the Set1A/B complexes do not require RNAPII for their recruitment to target sequences and that transcription is not a prerequisite for this modification.

This study was taken one step further by assaying for genes that rely on the presence of CFP1 for appropriate promoter H3K4me3 in mouse ESCs (4). CFP1-dependent histone methylation was detected at over 18,000 promoters. CFP1 knockout resulted in a detectable decrease in H3K4me3 in nearly half of these genes, 95% of which had CGIs. A composite view

of affected promoters indicates that the most obvious defects occur just downstream of the transcription start site (TSS), but the 20% most highly impacted genes exhibit pronounced loss immediately upstream of the TSS as well. DNA methylation at these promoters was largely unaffected. Transcriptional competency, as measured by GRO-seq, a technique that profiles transcription rates on a genomic scale (221), indicated that transcription was largely unaffected. However, highly expressed genes with the largest CFP1-dependent loss of H3K4me3 were impacted, although they were still highly transcribed relative to promoters whose H3K4me3 status was unaffected by CFP1 loss. DNA binding mutants of CFP1 were used to rescue expression of CFP1, and these restored promoter H3K4me3, although spurious accumulation H3K4me3 outside of CGIs occurred, consistent with previous reports that CFP1 limited activity of the Set1A/B complexes in mammalian and yeast systems (119,155,210,214).

Based on these data, it seems that CFP1 has developed a role within the mammalian Set1A/B complexes to target these activities to the appropriate promoter-proximal sites in the context of CGIs. Given that this histone modification is insufficient to drive transcription in isolation, it may be able to set the groundwork for a transcriptionally competent environment in combination with other factors (132,222).

Major outstanding questions lie in our understanding of the functional redundancy of Set1-related complexes in mammals. MLL methyltransferase activity has been well-documented at developmentally-regulated gene promoters (223). Antibodies distinct to Set1A and Set1B show a mutually exclusive nuclear distribution, suggesting that these complexes perform non-redundant functions (129). MLL1 regulates a defined subset of genes, and loss of this methyltransferase cannot be functionally compensated by other cellular methyltransferases (118). However, no studies have been undertaken to define the subset of promoters regulated by each, whether loss of one can be functionally compensated by another, and how their relative distribution is regulated. Additionally, MLL complexes have apparently obviated the requirement for CFP1 in their targeting, though why this occurred is somewhat unclear. It is reasonable to speculate that CFP1 may have been lost from MLL complexes following their acquisition of a CXXC domain. Both MLLs and CFP1 perhaps have subsequently developed distinct inputs, such as phosphoacceptor sites, to respond to different signaling cues.

CHAPTER THREE
CFP1 is a Chromatin-Associated ERK1/2 Interacting
Partner and Substrate

INTRODUCTION

Post-translational modifications, such as phosphorylation, acetylation and methylation, dynamically support physiological changes within cells in response to external stimuli. The functional consequences of phosphorylation can vary widely between substrates, and targeting of distinct phosphoacceptor sites in a single substrate can often direct unique outcomes. Defining how individual phosphorylation events contribute to a concerted phenotypic outcome can carry broad implications for understanding the complex underpinnings of physiological and pathological states.

Efforts to identify novel kinase substrates on a large scale have utilized a relatively limited repertoire of experimental approaches. Major impediments to substrate identification are the low abundance and transience of modifications in cells. One method takes advantage of a yeast two-hybrid system to detect interactions between kinases and interacting partners, potentially including novel substrates. Individual substrates are then validated by traditional means, including *in vitro* and *in vivo* radiolabeling assays and immunoblotting with antibodies raised against specific modification states. In the last decade, improvements in approaches to phosphorylation site identification by mass spectrometry, including enrichment of phosphorylated peptides (224), substrate tagging with specially-engineering kinases (66) and semi-quantitative techniques such as MudPIT and SILAC (225,226) have identified a plethora of novel sites (59,227-230).

We were interested in identifying novel substrates of the mitogen activated protein kinases, ERK1/2. ERK1/2 share a high degree of homology, and are generally considered functionally redundant kinases in terms of substrate specificity (1,231). Our lab devised a yeast-two hybrid approach that took advantage of our ability to selectively activate ERK2 in yeast through co-expression with a constitutively active upstream kinase, MEK1 R4F (231). In this way, we were able to compare relative affinity of ERK2 for binding partners in its activated and

quiescent states. Using this approach, we identified CXXC Finger Protein 1 (CFP1) as a novel ERK2 interacting protein.

CFP1 is a nuclear protein involved in at least two opposing transcriptional regulatory processes: DNA methylation, a gene silencing event (209), and Histone 3 Lysine 4 (H3K4) trimethylation, which supports transcription of associated genes (4). Validated post-translational modifications of CFP1 have not been reported in the literature, so determining whether ERK1/2 phosphorylates CFP1 and what this modification might mean for CFP1 function represented an opportunity to address unanswered questions regarding the role that ERK1/2 might play in dynamic regulation of DNA and H3K4 methylation.

Our characterization of post-translational modifications of CFP1 led us to examine not only phosphorylation, but lysine acetylation as well. Lysine acetylation was initially reported on, and long thought to be restricted to, histone proteins (232). Identification of non-histone acetylation targets have only become prevalent within the last 20 years, and the majority of identified lysine acetylated proteins are restricted to either the nucleus or mitochondria (233-236). Functional outcomes of acetylation are diverse, impacting protein stability, targeting, activity and interactions. Like phosphorylation, lysine acetylation is a reversible event. Lysine acetylation is catalyzed by three major families of acetyltransferases (KATs): GCN5, p300/CBP and MYST proteins (236,237).

EXPERIMENTAL PROCEDURES

Cloning and constructs

Constructs for bacterial expression of CFP1 were generated from human CFP1 cDNA ligated into the pGEX-6PI (GE Healthcare Life Sciences) vector that encodes an N-terminal glutathione S-transferase (GST) tag. Cloned fragments were generated by polymerase chain reaction (PCR) to encompass the following protein fragments: 1-481 (nucleotides 1-1443); 302-656 (nucleotides 906-1971) and 423-656 (nucleotides 1269-1971). Fragments and vector were digested with EcoRI and XhoI restriction enzymes and ligated. For mammalian cell expression, CFP1 was amplified by PCR to generate NotI and SalI restriction sites, then digested and ligated into p3xFLAG-CMV-7.1 (Sigma E4023). p3xFLAG-CMV-7.1 CFP1 K63R was generated using standard site-directed mutagenesis techniques with the following primer: 5'-

CTGAGAAGATGGCCAGGGCCATCCGGGAGTG-3'. All constructs were sequenced to ensure no spurious mutations were introduced.

Recombinant protein expression and purification

His$_6$-tagged ERK2 protein was expressed in Origami *E. coli* (EMD Millipore) using the NpT7-His$_6$-ERK2 wild type construct as described previously (238). Briefly, ERK2 was activated in vivo by coexpressing ERK2 and MEK1 R4F (231), a constitutively active form of MEK1. Activated ERK2 was purified by His60 Ni Superflow Resin (Clontech) and eluted over an imidazole gradient. pERK2-rich fractions were further purified by MonoQ 5/50 GL column (GE Life Sciences). Activity of purified pERK2 was verified by activity assay against myelin basic protein (MBP) and immunoblotting with pERK1/2 antibody.

To produce recombinant CFP1 and truncation constructs, pGEX-6PI-CFP1 constructs were transformed into Origami competent cells (EMD Millipore) and starter cultures were seeded from a single colony in Luria-Bertani (LB) medium containing 100 ug/mL ampicillin. The starter culture was incubated overnight at 30°C with shaking and used to inoculate 1L cultures of Terrific Broth (TB) containing 100 ug/mL ampicillin and 34 ug/mL chloramphenicol. Cells were grown in log phase at 37°C at 250 rpm until an OD$_{600}$ of 0.6 was reached. Cells were incubated at 20°C and 150 rpm for one hour prior to induction of protein expression with 0.4 mM isopropyl-1-thio-b-D-galactopyranoside (IPTG) for 16 hours. Bacteria was pelleted by centrifugation, then suspended in buffer containing 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, pH 7.5), 1 mM EDTA, 1 mM DTT with 1:500 protease inhibitor cocktail (in 62.5 mL of DMSO: 25 mg of pepstatin A, 25 mg of leupeptin, 250 mg of N$_a$-p-tosyl-L-arginine methyl ester, 250 mg of N$_a$-p-tosyl-L-lysine chloromethyl ketone hydrochloride, 250 mg of N$_a$-Benzoyl-L-arginine methyl ester, and 250 mg of soybean trypsin inhibitor). Resuspended pellets were frozen overnight at -20°C. Thawed pellets were treated with 1 mg/mL lysozyme, 1% 3-[(3-cholamidopropyl)dimethylammonio]-1- propanesulfonate (CHAPS) and 1:1000 benzonase (EMD Millipore) on ice, then extensively Dounce homogenized. Lysates were clarified by centrifugation for 1 hour at 35,000 rpm at 4°C on a Beckman Ti45 rotor. Supernatant was collected, and an equal volume of buffer containing 50 mM HEPES (pH 7.6), 300 mM NaCl, 1% Triton X-100, 2 mM EDTA and 2 mM DTT was

added. This mixture was allowed to bind overnight to buffer-equilibrated glutathione agarose (Pierce; 0.5 mL bed volume/ L culture). Beads were isolated and washed extensively with buffer containing 20 mM HEPES (pH 7.6), 150mM NaCl, 1% CHAPS, 1 mM EDTA and 1 mM DTT. Bound protein was eluted with 20 mM glutathione in this same buffer, and GST tags were cleaved using PreScission protease (GE Healthcare Life Sciences) according to the manufacturer's instructions. Isolated protein was dialyzed against 25% glycerol, 150 mM NaCl, 10 μM $ZnCl_2$, 1 mM EDTA and 1 mM DTT, frozen in liquid nitrogen and stored at -80ºC.

In vitro binding assay

Full-length uncut GST-CFP1 was bound to glutathione resin for 1.5 hours at 4°C. Recombinant pERK2 was added for 30 minutes, then resin was washed 4 times with buffer containing 20mM Tris, 150mM NaCl and 0.1% Triton X-100. Resin was boiled at 100°C in 5x Laemelli sample buffer prior to separation by separation by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and blotting with antibodies raised against ERK1/2 and pERK1/2 (pT185/pY187).

In vitro kinase assay and mass spectrometry

In vitro kinase reactions were carried out using recombinant activated pERK2 and various GST-CFP1 constructs with the GST tag removed by PreScission protease cleavage. Reactions proceeded at 30°C for either 30 minutes (for truncation mutant kinase reactions) or the time indicated. Reactions were carried out in the presence of 10 mM Tris (pH 7.5), 10 mM $MgCl_2$, 200 μM ATP and 10 μCi ATP^{32}. Reactions were quenched by addition of 5x Laemmli sample buffer and boiled for 5 minutes at 100°C. Samples for mass spectrometry were carried out for 60 minutes without radiolabeled ATP, separated by SDS-PAGE, Coomassie stained and submitted for analysis.

Phosphoamino acid analysis

CFP1 1-481 was phosphorylated *in vitro* with pERK2 for 30 minutes. The phosphorylated CFP1 fragment was immunoblotted and exposed to film to verify phosphate incorporation, then the band was excised and hydrolyzed with 6 N HCl. Hydrolysis products

were resolved by electrophoresis on a thin layer plate using a Hunter apparatus as described previously (239,240). The plate was exposed to film, and identities of phosphorylated residues were made by comparison to migration of phosphoamino acid standards visualized with ninhydrin.

Mammalian cell culture and labeling

HeLa cells were acquired from the ATCC and cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (Sigma) and 2 mM L-glutamine. Cells were transfected with p3xFLAG-CMV-7.1-CFP1 for 48 hours using Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. Medium was aspirated and plates were washed with phosphate buffered saline (PBS), then replaced with serum-free DMEM for 2 hours. Medium was again aspirated and cells were incubated with phosphate-depleted DMEM (Life Technologies) containing 1.0 mCi/mL $^{32}PO_4$ (MP Biomedical). Prior to stimulation, cells were pretreated with either 500 nM PD0325901 for 30 minutes or 10 μM wortmannin for 10 minutes (LC Laboratories) prior to stimulation with 100 nM phorbol myristate acetate (PMA; Sigma-Aldrich) for 15 minutes. Cells were lysed with RIPA Buffer (50 mM Tris (pH 7.4), 150 mM NaCl, 1mM EDTA, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 80 mM β-glycerophosphate, 100 mM NaF and 2 μM Na_3VO_4) with 1:1000 protease inhibitor cocktail, then extensively Dounce homogenized. Lysates were clarified by centrifugation, protein concentration was determined by μBCA assay (Pierce) and lysates were brought to the same concentration for immunoprecipitation with M2 Flag-agarose resin (Sigma) at 4°C for 4 hours. Resin was washed 3 times with RIPA buffer and denatured by boiling at 100°C in 5x Laemmli sample buffer prior to SDS-PAGE.

Subcellular fractionation, mononucleosome preparation and immunoprecipitation

HeLa cells were grown to confluency in 10 cm dishes and treated as described. For Flag-ERK2 immunoprecipitations, cells were transfected with p3xFlag-CMV7-ERK2 (241) for 48 hours with Lipofectamine 2000. Cells were rinsed with PBS and pellets were resuspended at $1.0x10^7$ cells per mL in Buffer A (10 mM HEPES (pH 7.9), 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, 1 mM DTT). Triton X-100 was added to a final concentration of 0.1%

and cells were incubated on ice for 5 minutes. Crude nuclear fractions were pelleted by microfuge at 3500 rpm for 5 minutes at 4°C. For further isolation of chromatin, nuclear pellets were washed once in detergent-free Buffer A, then resuspended in Buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT) for 15 minutes on ice with periodic vortexing. Chromatin was pelleted by microfuge at 4000 rpm for 5 minutes at 4°C.

Mononucleosomes were prepared by resuspending chromatin pellets with Buffer A containing 1 mM CaCl2, then digested at room temperature with micrococcal nuclease (Thermo Scientific). Digestions were stopped by adding EGTA to a final concentration of 1 mM. Undigested chromatin was pelleted by centrifugation at 4000 rpm for 5 minutes at 4°C. Supernatants were checked for mononucleosome isolation by phenol/chloroform extraction of genomic DNA and segregation on an ethidium bromide-stained agarose gel. For immunoprecipitation, salt concentrations of the mononucleosome fraction were adjusted to 150 mM with KCl and Triton X-100 was added to 0.2%. Flag-ERK2 was immunoprecipitated using Flag M2-conjugated agarose (Sigma) and washed extensively with salt-adjusted Buffer A prior to denaturation by boiling at 100°C in 5x Laemmli sample buffer.

Immunoprecipitation from crosslinked lysates

HeLa cells were crosslinked with 1% formaldehyde in PBS for 10 minutes at room temperature. Formaldehyde was quenched by addition of glycine to a final concentration of 0.125 M. Cells were washed with PBS and scraped into RIPA buffer with protease inhibitors (~1x10^7 cells per 300 μL of buffer). Lysates were bath sonicated using a Bioruptor system (Diagenode) for 28 minutes on highest setting with 30 seconds on/off cycles. Immunoprecipitations were performed using the following antibodies: normal mouse IgG (Santa Cruz), ERK1/2 (X837, generated in-house), and mouse anti-Histone 3 and H3K4 trimethyl (Abcam). Immunoprecipitated proteins were collected using either sheep anti-mouse or sheep anti-rabbit magnetic Dynabeads and denatured as previously described.

Immmunoblotting and antibodies

SDS-PAGE or tricine gels (242) were used to separate proteins that were transferred to nitrocellulose membrane (Millipore). Membranes were blocked with LiCor blocking buffer and

primary antibodies were incubated with membranes overnight. LiCor fluorescent secondary antibodies were used at 1:15,000 and immunoblots were imaged and bands quantified using a LiCor system. Western blot analysis was performed using the following primary antibodies: rabbit anti-CFP1 (Bethyl Laboratories); rabbit anti-ERK1/2 (produced in-house, (243)); mouse anti-phospho-ERK1/2 (T185/Y187) and mouse anti-Flag M2 (Sigma); α-tubulin (mouse monoclonal antiserum generated in-house from cells obtained through Developmental Studies Hybridoma Bank); mouse anti-Histone 3 and rabbit anti-H3K4 trimethyl (Active Motif); goat anti-Lamin A/C (Santa Cruz); Acetylated lysine (rabbit polyclonal from Cell Signaling; mouse monoclonal from Millipore).

RESULTS

Yeast two-hybrid screen identifies CFP1 as an ERK2-interacting protein

A yeast two-hybrid screen was performed by Svetlana Earnest, a research associate in our lab, in an effort to identify novel ERK2 substrates and interacting partners. The screen was designed to identify differential binding of interacting partners based on the kinase activation state of ERK2, with an eye towards identifying potential novel substrates. Two yeast strains were assayed in parallel -- one bearing a construct expressing ERK2 alone, and one with ERK2 co-expressed with a constitutively active form of the upstream activating kinase, MEK1 (231,244).

A mouse neonatal cDNA library was used as prey, and among interacting proteins identified was a fragment of CFP1 encompassing residues 373-660, a segment of the protein encompassing the C-terminus. A beta-galactosidase assay was employed to quantitatively assess the interaction (Figure 3-1).

Yeast two-hybrid for ERK2 interactors

Clone	pERK2	ERK2
CFP1	0.025	Not detected
Synapsin II	0.075	0.005
RSHL2	0.025	0.010
KIF2A	0.140	Not detected

Table 3-1: Summary of yeast two-hybrid results. Liquid β-galactosidase assays of two-hybrid interactions between prey and ERK2 or pERK2. Values are OD_{420}.
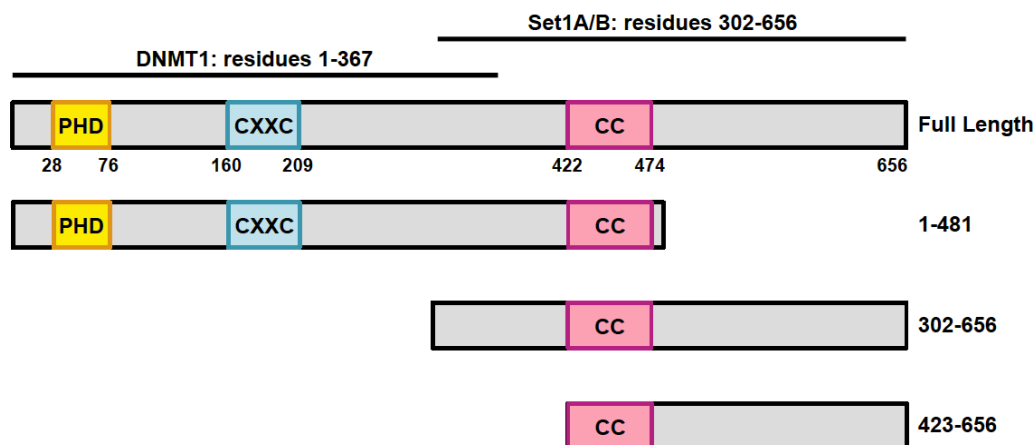
Figure 3-1: Domains and interacting regions of CFP1 truncation mutants used for *in vitro* interaction and kinase assays. PHD finger: Plant Homeodomain, binds methylated H3K4; CXXC: Unmethylated CpG binding domain; CC: putative coiled-coil domain. Known regions of interaction with DNMT1 and Set1A/B complexes are shown.

Recombinant CFP1 does not interact with ERK1/2

To determine whether the interaction between CFP1 and ERK2 was direct, Svetlana generated a glutathione S-transferase (GST)-tagged version of human CFP1, as well as a series of truncation mutants excluding various regions of the protein known from the literature to mediate interactions with binding partners, as well as domains that bind to DNA and H3K4me2/3 (See Figure 3-1; (23,124,140,245)). Amino acids 1-481 localize properly to the nucleus by immunofluorescence and are sufficient for interaction with DNMT1 but not with Set1. Residues 302-656 can interact with Set1 but not DNMT1, and likewise show proper gross localization, despite the absence of DNA and histone interacting domains. Finally, residues 423-656 interact with neither DNMT1 nor Set1, and although immunofluorescence data for this truncation has not been published, it lacks the nuclear localization signal at residues 109-121 (245) and therefore likely would not localize properly.

GST-CFP1 and fragments were generated in *E. coli*. ERK1 and ERK2 were purified and activated *in vitro* with purified active MEK1-R4F (231). GST-CFP1 and activated ERK1 or ERK2 were co-incubated to test *in vitro* binding. No evidence of binding was detected. Taken

	Cytosolic		Soluble Nuclear		Chromatin	
EGF	+	+	+	+	+	+
PD0325901		+ +		+ +		+ +

CFP1
pERK1/2
ERK1/2
α-tubulin
Histone 3
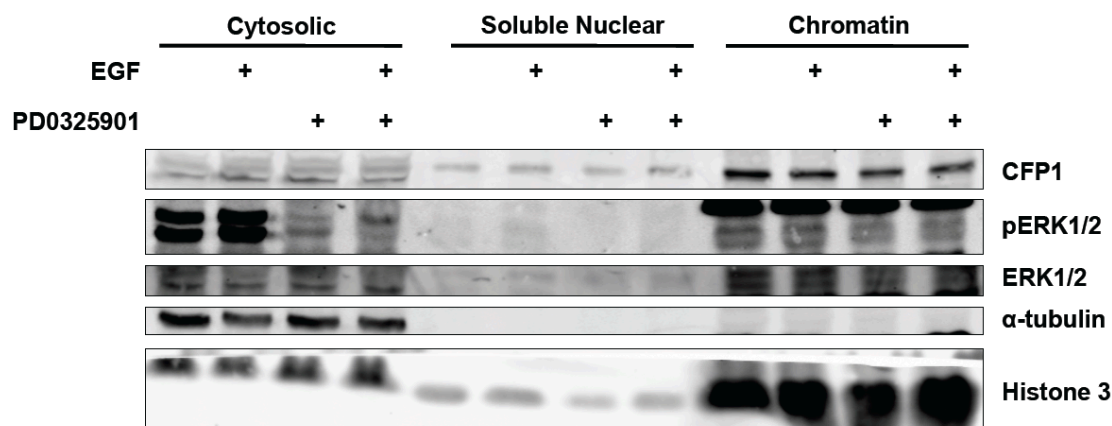
Figure 3-2: Subcellular fractionation of HeLa cells following stimulation and inhibition of ERK1/2. HeLa cells were starved to quiescence in serum-free DMEM, then pretreated with 500 nM PD0325901 or DMSO for one hour prior to stimulation. EGF was added for 15 minutes at 10 ng/mL. Cells were fractionated to separate crude cytosolic, soluble nuclear and insoluble chromatin pools, and fractions were volume-normalized prior to separation by SDS-PAGE and immunoblotting with the indicated antibodies. Antibodies against tubulin and Histone 3 were used to validate the quality and consistency of cytosolic and nuclear fractions, respectively. Representative of two separate experiments.

together with the yeast two-hybrid results, this suggests that interaction between ERK2 and CFP1 is weak, indirect, or relies on conformation of CFP1 stabilized by interaction with binding partners or in a complex.

CFP1 and ERK1/2 subcellular localization

CFP1 localizes predominantly to the nucleus, where it is mainly restricted to euchromatin (245). ERK1/2 localization is cell-type dependent, but is found in many compartments, including the nucleus (1). Transient nuclear redistribution of cytosolic ERK1/2 can be stimulated through ERK1/2 pathway activation by FBS (75).

We checked the relative distribution of these proteins in HeLa cells under different states of ERK1/2 pathway activation (Figure 3-2). Briefly, we starved cells in serum-free medium and either pretreated with the MEK1/2 inhibitor PD0325901 or dimethylsulfoxide (DMSO), then activated ERK1/2 by brief treatment with epidermal growth factor (EGF). We fractionated

cytosolic, soluble nuclear and insoluble chromatin fractions by mild detergent treatment and differential centrifugation. Tubulin and Histone 3 were blotted to roughly verify the quality and consistency of cytosolic and nuclear fractionation. Individual fractions were brought to the same volume prior to denaturation and SDS-PAGE, so that the relative distribution of these proteins could be roughly assessed.

As expected, ERK1/2 was stimulated by EGF treatment and all detectable phospho-ERK1/2 signal was blocked by the inhibitor, especially apparent in the cytosolic fraction. Bulk redistribution of ERK1/2 following activation was not readily detectable by this method, although a pool of ERK1/2 is apparent in the chromatin fraction, even under unstimulated and inhibitor-blocked conditions.

CFP1 almost exclusively localizes to the chromatin fraction, although some signal is detectable in both the soluble nuclear and cytosolic fractions. A doublet of bands appears in the cytosolic CFP1 blot, where a lower molecular weight species is detected in addition to the single band observed in nuclear fractions. It is unclear whether this represents a spliced or modified variant of CFP1, or is simply a cross-reacting protein.

CFP1 and ERK2 interact on chromatin

Efforts to reproduce the interaction between CFP1 and ERK2 in mammalian whole cell lysates were unsuccessful. Various approaches were taken in HeLa cells utilizing both overexpressed and endogenous protein pairs with various affinity tags, but co-immunoprecipitation from a total cellular lysate consistently failed.

Two alternative approaches were taken to validate the interaction between CFP1 and ERK2 in human cells. First, cells were briefly cross-linked with formaldehyde, lysed and sonicated. Lysates were cleared, and pulldowns were performed with antibodies targeting ERK1/2, Histone 3, trimethylated H3K4 and IgG as a negative control (Figure 3-3). Immunoprecipitated proteins were separated on the same SDS-PAGE gel and membranes were blotted for the pulldown targets, as well as for CFP1. CFP1 signal was enriched in the ERK1/2 and H3K4me3 pulldowns, but absent in the H3 and IgG pulldowns.
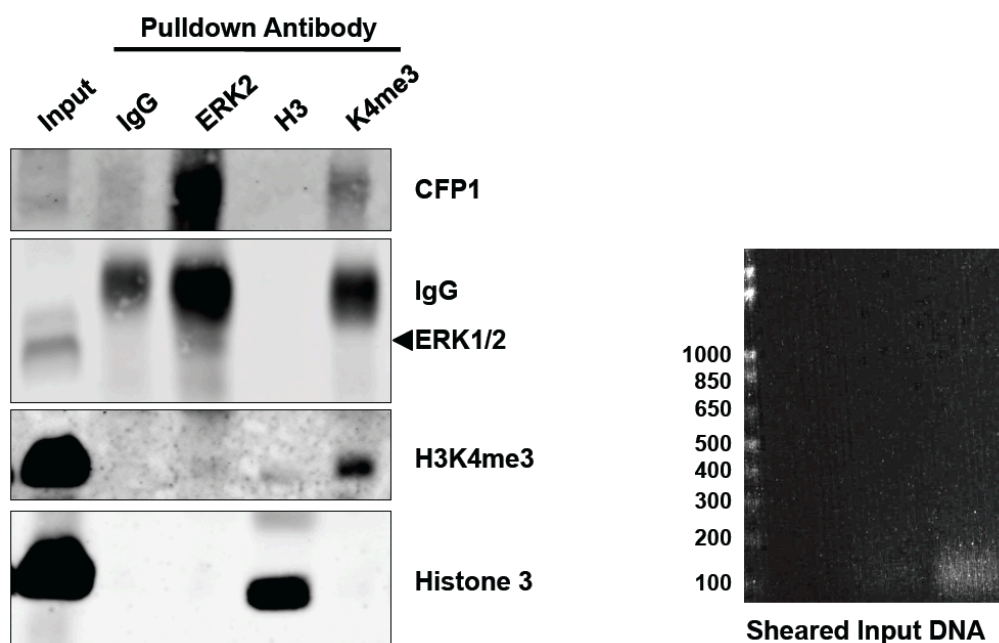
Figure 3-3: ERK2 and CFP1 co-immunoprecipitate from cells crosslinked with formaldehyde. Quiescent cells were crosslinked with 1% formaldehyde in PBS for 10 minutes at room temperature then lysed with RIPA buffer, sonicated, and clarified. Supernatants were subjected to immunoprecipitation with the indicated antibodies, resolved by SDS-PAGE and immunoblotted with the indicated antibodies. The IgG band spreads into the ERK1/2 band, which is marked. Sheared DNA size was assessed by heat-reversing crosslinks, purification by phenol/chloroform extraction, and resolution of sheared DNA on a 1.2% agarose gel. Representative of two experiments.

In the second approach, HeLa cells were transfected with Flag-ERK2 for 48 hours prior to isolation of chromatin and preparation of an enriched population of mononucleosomes by limited digestion of chromatin with micrococcal nuclease, an endonuclease that shows preference to DNA sequences not bound by protein (Figure 3-4). Enrichment of mononucleosomes was verified by isolating DNA by phenol/chloroform extraction and electrophoresing isolated DNA on an agarose gel. The only detectable band in this fraction fell between DNA standards for 100 and 200 bases, as expected for single mononucleosomes that contain roughly 150 bases of DNA. Presence of Flag-ERK2 was verified in all fractions assayed, as observed in prior fractionations.
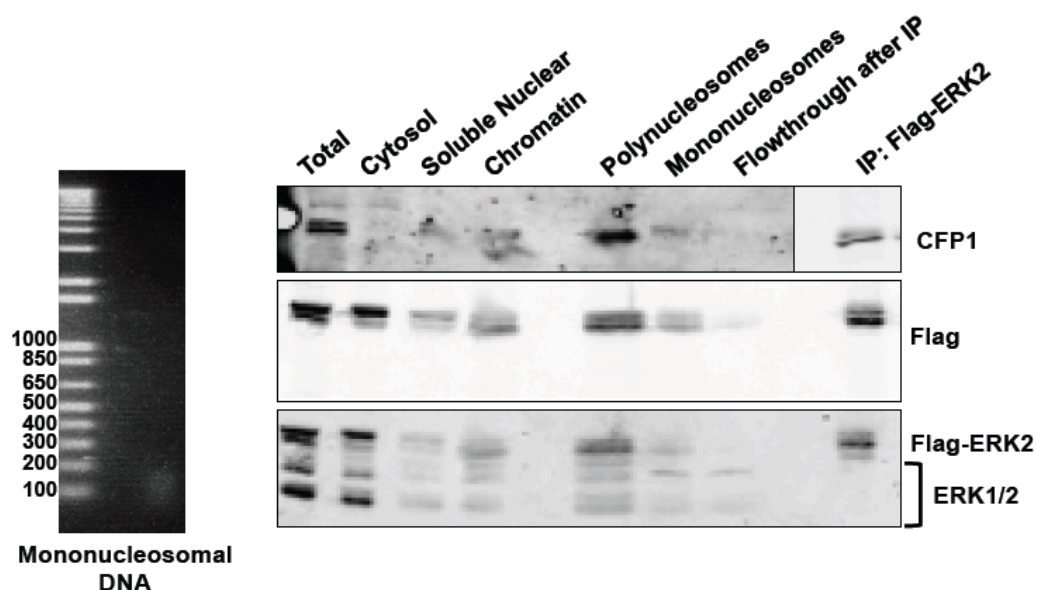
Figure 3-4: CFP1 and ERK2 co-immunoprecipitate from mononucleosomes. HeLa cells were transfected with Flag-tagged ERK2 for 48 hours, then fractionated. Chromatin was digested with micrococcal nuclease to isolate mononucleosomes, and this fraction underwent immunoprecipitation with resin conjugated to a monoclonal Flag antibody. Fractions and immunoprecipitated proteins were separated by SDS-PAGE and blotted with the indicated antibodies. DNA was isolated from the mononucleosomal fraction and separated by agarose gel electrophoresis to verify appropriate fragment size. Representative of three similar experiments. *Courtesy of Svetlana Earnest.*

Flag-ERK2 was immunoprecipitated from the mononucleosomal fraction using resin-conjugated monoclonal Flag antibody. ERK2-enriched mononucleosomes were then blotted for endogenous CFP1, which was apparent as a doublet of bands, similar to those seen in the total cellular lysate. Unlike previous fractionation experiments, fractions were not brought to the same volume, so comparisons of relative protein enrichment cannot be estimated from the blots shown, but the presence and absence of CFP1 is consistent with other fractionation experiments.

CFP1 is phosphorylated *in vitro* by ERK2

Because ERK2 and CFP1 interact in cells, we next wanted to determine if ERK1/2 could phosphorylate CFP1. The GST-CFP1 constructs described earlier were cleaved of their GST tags and used as substrates in kinase reactions with activated recombinant ERK2 (Figure 3-5).
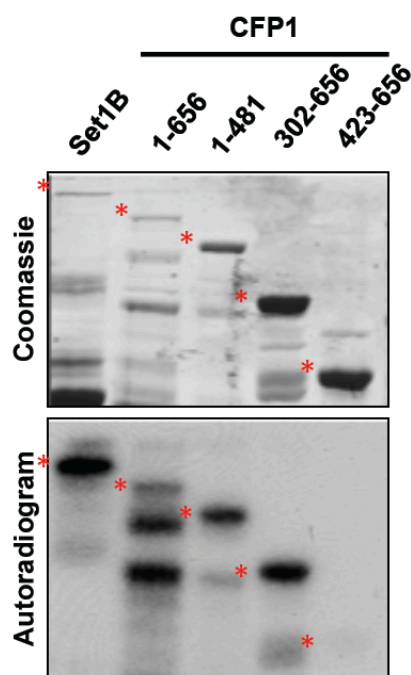
Figure 3-5: Truncation mutants of CFP1 are phosphorylated by pERK2 *in vitro*. Full-length and fragments of GST-tagged CFP1 were expressed in *E. coli* and purified on glutathione resin. GST tags were cleaved, and fragments were subjected to *in vitro* kinase assay for 30 minutes at 30°C. Reactions were separated by SDS-PAGE, Coomassie stained and exposed to film. Purified mouse Set1B (residues 1185-1985) was included as a positive control for phosphorylation. Red asterisks correspond to the indicated fragments. Representative of two experiments. *Courtesy of Svetlana Earnest.*

Reactions were run for 30 minutes at 30°C. Mouse Set1B 1185-1985 was included as a positive control for ERK2 activity. Relative incorporation of phosphate into each CFP1 truncation construct was not calculated, but an autoradiogram of the gel clearly indicates phosphate incorporation into all fragments, except for residues 423-656.

The full-length recombinant protein was used for phosphoamino acid analysis (PAA; Figure 3-6). *In vitro* kinase reactions were performed, and quenched reactions were separated by SDS-PAGE, Coomassie stained, and exposed to film to verify incorporation of ^{32}P into CFP1. The band corresponding to phosphorylated CFP1 was excised and acid hydrolyzed prior to separation of phosphorylated species by electrophoresis on a silica gel plate alongside phosphorylated amino acid standards. The plate was then exposed to film to determine which amino acids were phosphorylated. Phospho-serine and phospho-threonine were both evident.
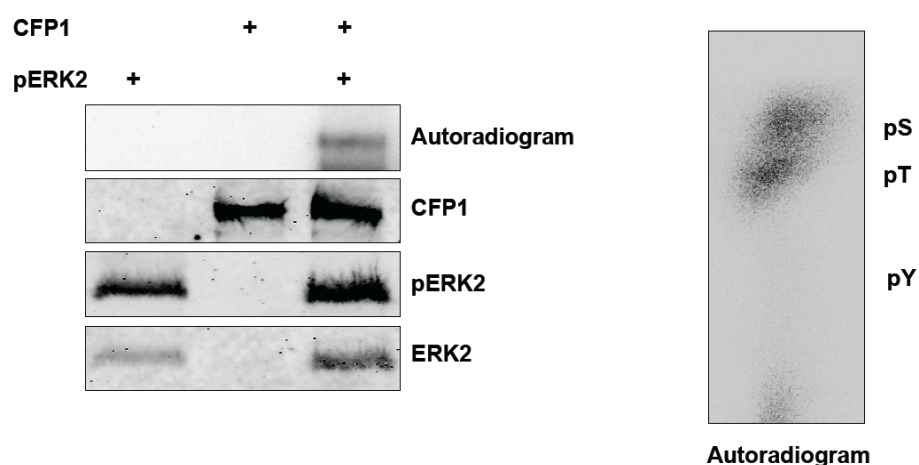
Figure 3-6: Phosphoamino acid analysis reveals that CFP1 is phosphorylated on serine and threonine by ERK2 *in vitro*. *In vitro* kinase reactions included full-length recombinant CFP1 and activated recombinant ERK2 in the presence of [γ-^{32}P]ATP at 30°C. Proteins were resolved by SDS-PAGE, and bands corresponding to phosphorylated CFP1 were hydrolyzed in acid prior to separation of amino acid species by electrophoresis on a silica gel plate. Assignment of phosphorylated amino acids was made based on relative migration of purified standards. Representative of two experiments.

There was no detectable phospho-tyrosine, as expected from ERK2 that exhibits serine/threonine substrate specificity.

The fragment encompassing residues 1-481 was selected for a more detailed analysis of CFP1 phosphorylation by ERK2, including calculation of stoichiometry and mass spectrometry for sites that incorporated detectable phosphate. CFP1 1-481 localizes to the nucleus in cells (245) and interacts with both DNMT1 and Set1 (23). From our purifications, we found that it was much more stable than the full-length protein. A time course assay was performed with this CFP1 fragment and recombinant active ERK2 (Figure 3-7). Reactions were quenched, analyzed by SDS-PAGE and Coomassie stained. CFP1 1-481 undergoes a pronounced band shift over the duration of incubation with ERK2. Incorporation of phosphate, as assessed by scintillation counting of labeled bands as well as visual inspection of band shifting, appears to be complete at 90 minutes. Stoichiometry of phosphorylation was calculated based on protein and ATP standard curves to be 1.5 mol phosphate per mol CFP1.
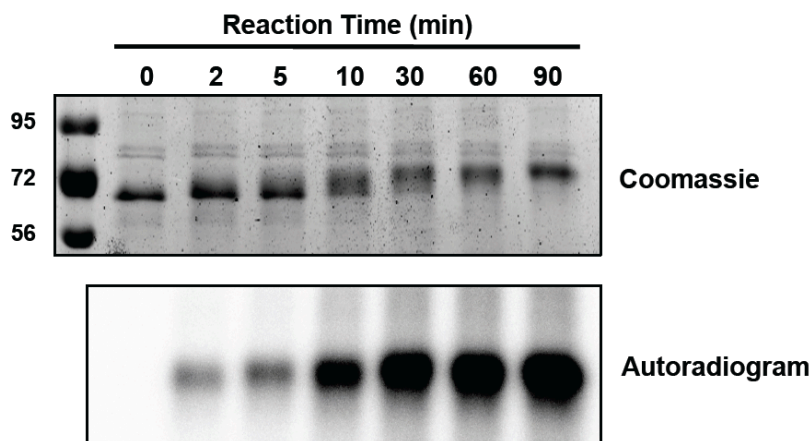
Figure 3-7: ERK2 phosphorylates CFP1 in *vitro* with a stoichiometry of ~1.5 mol P_i/mol CFP1. A recombinant human CFP1 fragment encompassing residues 1-481 was incubated with $[\gamma\text{-}^{32}P]$ATP and activated recombinant ERK2 for the indicated times at 30°C. Kinase reactions were quenched with SDS buffer, then subjected to SDS-PAGE and exposed to film. Bands were excised and counted alongside $[\gamma\text{-}^{32}P]$ATP standards to calculate a molar ratio of ^{32}P incorporation. Representative of three experiments.

Mass spectrometry of phosphorylated CFP1 suggests multiple sites of modification

ERK1/2 substrate sites are most typically serines and threonines, followed immediately by a proline. In an optimal scenario, another proline is located two residues upstream from the phosphoacceptor sites (Pro-X-Ser/Thr-Pro), and ERK1/2 docking motifs, such as D motifs and FXF motifs are proximal to the target residue (1). We sought to identify putative ERK1/2 phosphorylation sites on CFP1, both by manual assessment using these guidelines, as well as phosphorylation site prediction software. Scansite predicts phosphorylation and binding sites on input sequences based on information derived from oriented peptide library screens for kinase substrate specificity (246). Phosphosite compiles *in vivo* post-translational modification data from both low and high throughput sources in the literature, though often entries do not report the kinase responsible for modification (230). Predicted ERK1/2 phosphorylation sites from Scansite and observed phosphorylated residues on CFP1 from Phosphosite are summarized in Table 3-2.

Unlabeled *in vitro* phosphorylation reactions from a 60-minute time point were submitted to the UT Southwestern Proteomics Core facility for modification analysis. The samples were

Site	ModLS score (≥27; FDR <1%) Trypsin	ModLS score (≥27; FDR <1%) Elastase	Phosphosite (# of times observed)	Scansite Predicted	Other Predicted Kinases for Site
S124	128	ND	2		Akt
T126	79	25	1		
S143	1	19		Y	CDK1/5, CDC2
T150	80	22			CDK1/5, CDC2
S215	21	156			Aurora B, PKA, PKC
S221	8	ND	2		
S222	1	3	1		
S224	53	29	5	Y	CDK1/5, CDC2
T227	79	71	6	Y	CDK5, CDC2, GSK3
T241	165	312			ATM, DNAPK
S248	48	Unamb.			ATM
T273	ND	91		Y	GSK3
T275	75	48		Y	ATM, CDK1, DNAPK, GSK3
S371	5	Unamb.			CK1
S387	16	36			PKC
S418	Unamb.	Unamb.			CDK1

Table 3-2: Mass spectrometry results compared to predicted sites. Phosphorylated residues are indicated in the first column. ModLS scores are represented as reported by the UT Southwestern Proteomics Core for both trypsin and elastase digests. A ModLS score of 27 was used as an arbitrary cutoff, as this score is considered to have an associated false discovery rate < 1%. Reported sites are compared to publically-available phosphorylation information from Phosphosite.org, as well as predicted substrate sites for ERK1/2 by Scansite.org. If other kinases were considered potential enzymes for that site, they are also reported. ND, Not Detected; Unamb., Unambiguous assignment.

split between two different protease reactions to maximize coverage of the protein, in an attempt to definitively map as many sites as possible. Many phosphorylation sites were identified, although the relative incorporation of phosphate into each site could not be determined from the analysis performed.

Phosphorylation sites were called using the ModLS algorithm, developed by members of the core facility (247). This algorithm scores individual phosphoacceptor residues in phosphorylated protein fragments based on their relative probabilities of detection given the mixture of fragments identified and provides a score for each potential site of modification. Scores are calculated independently for each experiment, so values reported by trypsin and

elastase digest are often different. Based on recommendations by members of the core facility, we used a score of 27 as an arbitrary cutoff for positive identification, as this score is considered to represent a scenario where the false discovery rate is less than 1%. Table 3-2 summarizes all sites with at least one ModLS score greater than the false discovery rate cutoff in comparison with results from Phosphosite and Scansite. A full report of potential sites identified by trypsin and elastase digests are located in Appendix A and Appendix B, respectively.

Phosphorylation sites that we were particularly interested to follow up on are those that were positively identified in the mass spectrometry screen, were predicted to be ERK1/2 substrate sites and have been previously detected *in vivo*. Of the 16 sites outlined in Table 3-2, our focus fell most especially to S224 and T227. These sites were both individually mutated to aspartate in the context of recombinant GST-CFP1 1-481, and purified protein was compared to wild type protein in a time course of phosphorylation by activated ERK2 (Figure 3-8). Band shifting of both mutants was more pronounced at an earlier time point than wild type protein, despite all unmodified proteins migrating similarly. Bands were excised and counted for phosphate incorporation, then normalized to relative loaded protein. Over the range of time
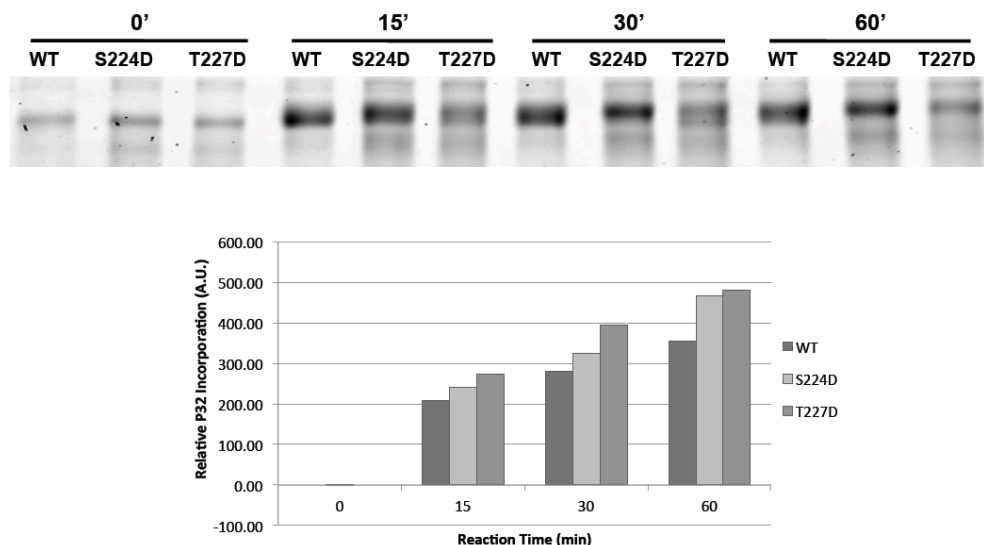
Figure 3-8: *In vitro* kinase assay with active ERK2 and select CFP1 point mutants. Recombinant human wild type and mutant CFP1 fragments encompassing residues 1-481 were incubated with [γ-^{32}P]ATP and activated recombinant ERK2 for the indicated times at 30°C. Kinase reactions were then subjected to SDS-PAGE and Coomassie stained prior to imaging using a LiCor system. Bands were excised and underwent scintillation counting, and relative incorporation of radioactive phosphate was calculated, and normalized to relative amounts of each CFP1 mutant detected at the zero time point. Representative of a single experiment.
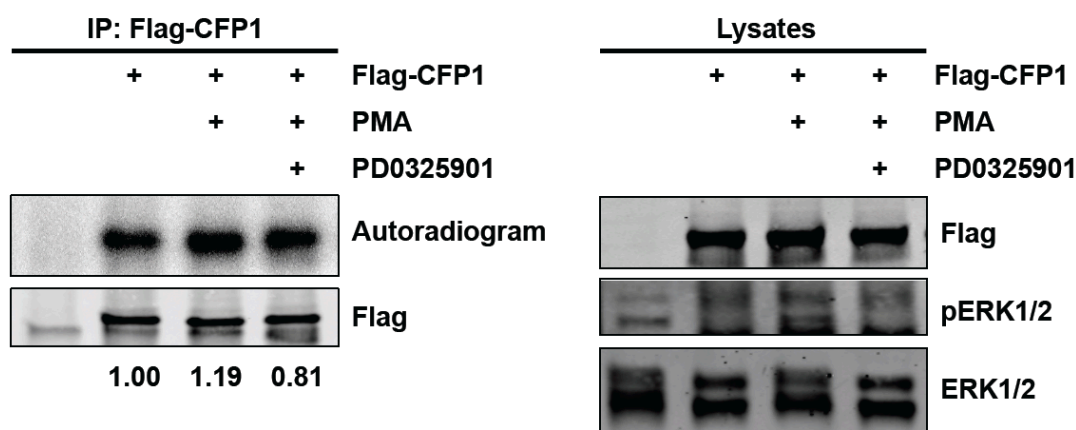
Figure 3-9: CFP1 is an ERK1/2 substrate in cells. HeLa cells were transfected with Flag-CFP1, then serum starved for 2 hours followed by 1 hour of incubation with 1.0 mCi/mL $^{32}PO_4$ in the presence or absence of the MEK1/2 inhibitor PD0325901 (500nM), and then stimulated with 100 nM PMA for 10 minutes. CFP1 was immunoprecipitated with Flag antibody-conjugated agarose and immunoprecipitated protein was resolved by SDS-PAGE. Fold change in ^{32}P incorporation normalized to Flag-CFP1 LiCor signal in each lane is displayed. Representative of two similar experiments.

points both mutants showed phosphate incorporation values that were similar or greater than the wild type protein. From these data, it is unclear if these are *bona fide* phosphorylation sites, but clearly neither constitutes a major site of phosphorylation *in vitro*. Alternatively, these sites may serve as priming phosphorylations, promoting modification at other sites.

CFP1 is phosphorylated in cells by ERK1/2

Mass spectrometry data suggested that multiple residues could serve as ERK1/2 substrate sites *in vitro*. Since recombinant CFP1 is not structurally confined by interaction with its native complexes, we expected that some of these sites might not be recapitulated *in vivo*. Unfortunately, we were unable to purify enough CFP1 from mammalian cells to perform modification analysis by mass spectrometry.

We performed cell labeling experiments in HeLa cells with overexpressed Flag-tagged CFP1 (Figure 3-9). Cells were transfected for two days and serum starved prior to the experiment. Labeling took place for one hour with 1.0mCi of $^{32}PO_4$ in phosphate-depleted DMEM, and cells were briefly treated with PD0325901 or DMSO prior to a 15 minute ERK1/2 pathway stimulation with phorbol myristate acetate (PMA). Immunoprecipitated Flag-CFP1 and

total cellular lysates were separated by SDS-PAGE gel, transferred and blotted with antibodies against the Flag epitope as well as for total and activated ERK1/2 to verify that drug treatments worked. Phosphate incorporation was visualized by exposing membranes to film, and relative incorporation of radioactive phosphate was quantified by scintillation counting and normalized to blotted Flag-CFP1.

PMA treatment induced a nearly 20% increase in phosphate labeling of Flag-CFP1. Pretreatment with PD0325901 resulted in a decrease in phosphate incorporation 20% below the control value. Substantial phosphate incorporation into Flag-CFP1 was observed in the absence of stimulus, which likely reflects phosphorylation events stimulated by kinases other than ERK1/2. Alternatively, it may be possible that nuclear ERK1/2 is refractory to starvation conditions, an effect that we have been unable to conclusively demonstrate.

As an alternative approach to radiolabeling, we attempted to optimize electrophoretic separation of phosphorylated CFP1 species using Phos-tag gels (248). Phos-tag is a commercially available compound that stably incorporates manganese ions into polyacrylamide gels to promote segregation of phosphorylated from unmodified protein targets by retarding the migration of the more negatively-charged phosphorylated protein. We assayed both endogenous and overexpressed proteins (Figure 3-10). In the first panel, starved HeLa cells were either treated with PD0325901 or DMSO, then stimulated with PMA over a time course. A laddering effect is evident in a blot of endogenous CFP1, suggesting a complex mixture of phosphorylation states. Co-treatment with PD0325901 appears to collapse some of these bands, but it is unclear how many phosphorylated species this might represent.

To verify that the banding pattern that we saw was indeed due to CFP1, we transfected HeLa cells with Flag-CFP1 prior to Phos-tag blotting. We starved cells and pretreated with either PD0325901 or 10 μM wortmannin, a phosphoinositide 3-kinase (PI3K) inhibitor (249). Wortmannin was included in an attempt to determine if other Ras-activated pathways were contributing to CFP1 phosphorylation. We then treated with either PMA or 20% fetal bovine serum (FBS) to stimulate ERK1/2 pathway activation. Again, banding patterns were complex for CFP1, as observed for the endogenous protein, lending further credence to the idea that multiple phosphorylated states for this protein simultaneously exist in the cell. Changes in banding

complexity are not obvious with either stimulation or with inhibitor treatments. We can conclude that separation worked well for stoichiometrically phosphorylated proteins, as we were able to
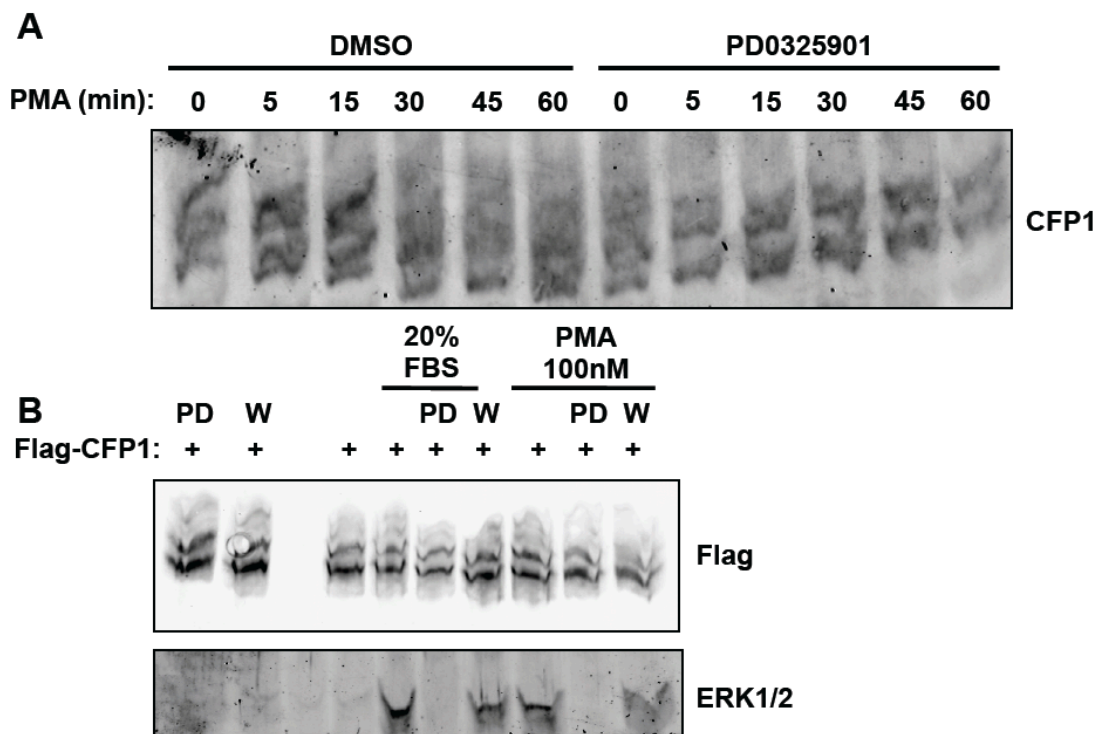
Figure 3-10: Phos-tag gel electrophoresis reveals multiple stable phosphorylation states of CFP1. A. HeLa cells were starved in serum-free DMEM, then pretreated for 30 minutes with either 500 nM PD0325901 or DMSO. Cells were then stimulated for the indicated durations, lysed, and resolved on a Phos-tag polyacrylamide gel. Immunoblotting was performed with an antibody against endogenous CFP1. Representative of two experiments. B. Cells were transfected with Flag-tagged CFP1 for 48 hours, then pretreated either with PD0325901 as described in A, or for 10 minutes with 10 uM wortmannin. Lysates were resolved on a Phos-tag gel. Membranes were immunoblotted with the indicated antibodies. Representative of a single replicate.

blot a shifted form of ERK1 with a total ERK1/2 antibody near the bottom of the gel. This suggested to us that phosphorylation of CFP1 was sub-stoichiometric and likely occurs at multiple sites.

CFP1 is acetylated in cells

We were interested in other unexplored post-translational modifications reported on CFP1 by Phosphosite. Acetylation caught our interest based on the widespread nuclear
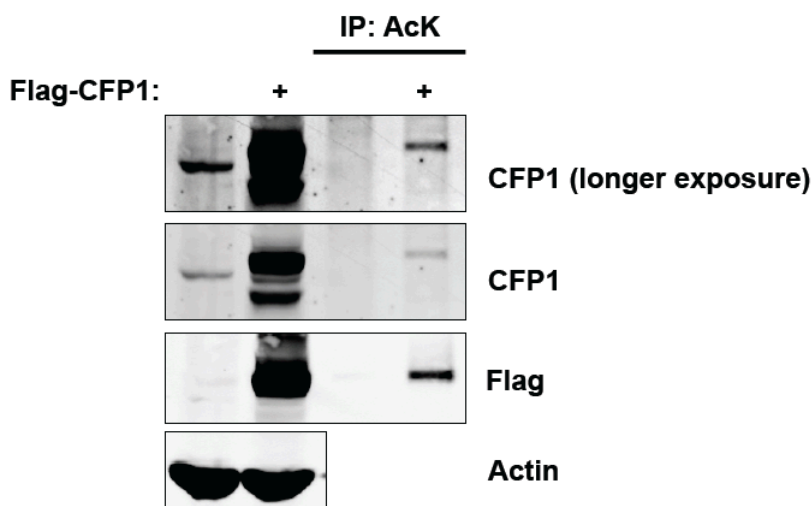
Figure 3-11: Flag-CFP1 is acetylated under basal conditions. HeLa cells were transfected with Flag-tagged CFP1 for 48 hours. Cell lysates were precleared with Protein A resin and subsequently immunoprecipitated with an antibody targeting acetylated lysine and collected on Protein A beads. Beads were washed extensively, then boiled in SDS sample buffer prior to resolution of bands by SDS-PAGE. Membranes were blotted with the indicated antibodies. Representative of a single replicate.

functionality of this modification as well as its ease of detection using broadly-specific acetylated lysine antibodies. Acetylation was reported for a single residue, lysine 63, that occurs within the Plant Homeodomain (PHD) of CFP1, a domain that promotes recognition of di- and trimethylated H3K4 (124,250). We speculated that acetylation at this site might play a role in modulating the interaction of CFP1 with chromatin.

We transfected HeLa cells with Flag-CFP1 or an empty vector control and performed immunoprecipitations with acetyl-lysine specific antibody (Figure 3-11). Enriched proteins were analyzed by SDS-PAGE and blotted with antibodies targeting both the Flag epitope as well as CFP1. Both antibodies specifically recognized a band in the Flag-CFP1 transfected pulldowns at a slightly higher apparent molecular weight than Flag-CFP1.

To further verify that CFP1 is acetylated, we split Flag-CFP1 transfected lysates into two separate immunoprecipitations (Figure 3-12). One was treated with trichostatin A (TSA), a broad

specificity lysine deacetylase inhibitor that we predicted would stabilize acetylated lysine through the course of an immunoprecipitation. Acetyl-lysine blots of the immunoprecipitated
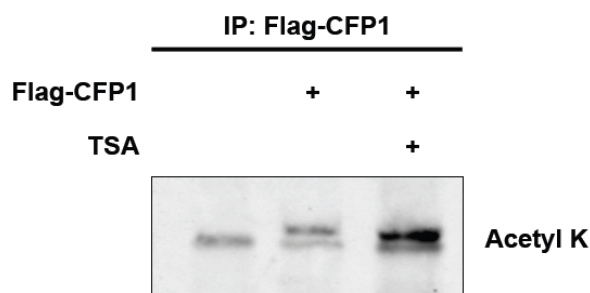
Figure 3-12: Addition of Trichostatin A to HeLa lysates stabilizes an acetyl-lysine signal at the same molecular weight as Flag-CFP1. Cells were transfected with Flag-CFP1 or an empty vector control for 48 hours. Cell lysates were split equally between immunoprecipitation conditions either containing 400 nM Trichostatin A or not. Following 2 hours of incubation with anti-Flag conjugated resin, immunoprecipitated protein was washed extensively with lysis buffer, then boiled in SDS sample buffer and resolved by SDS-PAGE. The membrane was blotted as indicated. Representative of two similar experiments.

proteins at the expected molecular weight of Flag-CFP1 shows a clear stabilization of signal, further validating acetylation of CFP1.

In an attempt to unambiguously assign the site of acetylation, we mutated lysine 63 to arginine (K63R) and immunprecipitated it alongside Flag-CFP1 (Figure 3-13). This site was identified in a global proteomics screen of acetylated proteins (250). Mutation at this residue did not reduce the level of acetylated lysine detected relative to the wild type protein. This may be the result of incorrect assignment in the initial screen, since another lysine is present within the tryptic peptide, K60. Alternatively, many sites of acetylation may exist on CFP1. Acetylation sites are nearly impossible to predict based on primary amino acid sequence, because lysine acetyltransferases do not appear to conform to any strict sequence specificity rules (251).

Further attempts were made to define acetyltransferases that mark CFP1 by using a targeted siRNA screen against the four major mammalian lysine acetyltransferases, then blotting for acetyl-lysine on immunoprecipitated Flag-CFP1. These experiments were not successful in revealing further regulatory information about this modification (data not shown).
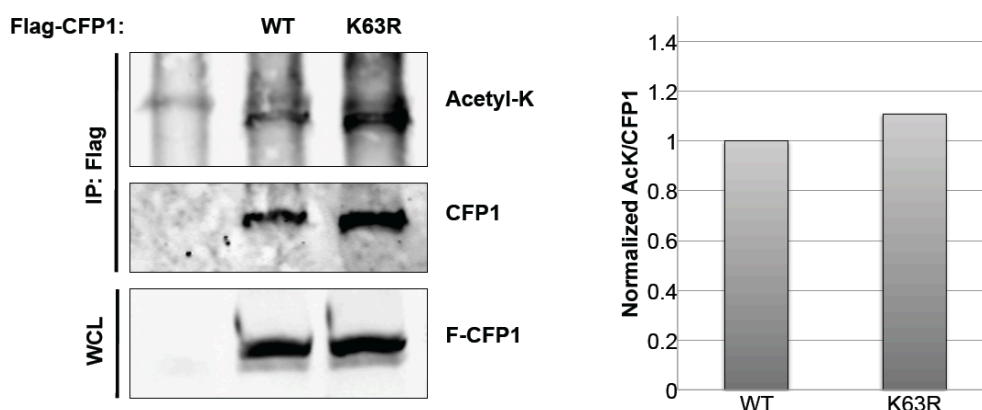
Figure 3-13: Mutagenesis of K63, identified in the literature as a site of acetylation by mass spectrometry, does not impact recognition of CFP1 by an antibody targeting acetylated lysine. Cells were transfected for 48 hours with either wild type (WT) or a K63R mutant of Flag-CFP1. Following 2 hours of incubation with anti-Flag conjugated resin, immunoprecipitated protein was washed extensively with lysis buffer, then boiled in SDS sample buffer and resolved by SDS-PAGE. The membrane was blotted as indicated. Acetyl lysine and CFP1 signals were quantified using the LiCor imaging system. Representative of two similar experiments.

DISCUSSION

Our yeast two-hybrid study successfully identified two other novel ERK1/2 substrates in addition to CFP1, both of which were relatively straightforward in their characterization and phosphorylation site assignment. We hoped that CFP1 would represent a similarly straightforward challenge, but that was not the case.

One early concern that we had was that CFP1 was a false positive identification from the yeast two-hybrid screen, and that it was in fact direct binding of ERK2 to *S. cerevisiae*'s endogenous Set1 that supported colony growth on selection media. While we are still unable to unequivocally rule this out without interaction testing between murine CFP1 373-660 and yeast Set1, alignment of the yeast CFP1 homolog Spp1 completely lacks the identified domain, supporting the idea that binding to Set1 did not contribute to the detected interaction. This concern was not put to rest by initial interaction studies. If binding between CFP1 and ERK1/2 is direct, it is too weak to detect in either *in vitro* binding experiments or immunoprecipitations from whole cell lysates, and may only occur during substrate targeting.

Successful co-immunoprecipitations from mononucleosomes and cross-linked lysates might suggest that interaction between CFP1 and ERK1/2 is stabilized by other nuclear proteins or by nucleosomes themselves. It might also suggest that the locations where CFP1 and ERK1/2

interact are very restricted and that interaction may be transient. This would be consistent with the fact that CFP1 binds to chromatin almost exclusively at promoters and ERK1/2 promoter recruitment may be short-lived and dependent on activity in many cases. Restricted CFP1 localization is supported by robust immunoprecipitation with trimethylated H3K4, a marking concentrated at promoters, but not with total H3 that is globally distributed.

Phosphorylation site determination was much more complicated than we anticipated. *In vitro* sites from mass spectrometry numbered too many to reasonably be tested in a cell-based setting by a cell labeling approach, and preliminary results from mutagenesis and *in vitro* phosphorylation were relatively uninformative in paring down the list of possible sites. Use of point mutants in cell labeling experiments was additionally hampered by large differences in expression of the mutants that we made prior to obtaining mass spectrometry results (S224A and T227V) relative to the wild type protein (discussed further in Chapter 4), making definitive quantitation of phosphate incorporation challenging. Furthermore, if many phosphorylation sites exist in a cell-based setting, as suggested by labeling studies with wild type Flag-CFP1, mutagenesis of single sites may not greatly impact overall phosphate incorporation. Some of the best sites by mass spectrometry were in agreement with existing data from Phosphosite as well as site prediction information. Which sites are actually modified *in vivo* by ERK1/2 targeting remains to be more fully explored, but might be addressed through cell labeling experiments with point mutants (S224A and T227V, for example) that cannot be phosphorylated. Such assays were not attempted due to stability differences of these overexpressed proteins relative to their wild type counterpart in HeLa cells (discussed in Chapter 4). Interestingly, none of the sites identified by mass spectrometry occur within identified structured domains of CFP1, perhaps indicating that CFP1 phosphorylation might impact interactions or conformational changes of these regions.

One approach that may aid in assignment of *bona fide* ERK1/2-regulated phosphorylation sites is SILAC labeling (stable isotope labeling of amino acids in cell culture). Briefly, this technique relies on two parallel sets of cells, one labeled with heavy amino acid isotopes and one grown in normal medium. Cells are treated differentially – for instance in this case, with or without the MEK1/2 inhibitor PD0325901 – then lysates are mixed, proteolyzed and analyzed by mass spectrometry. Since protein input is normalized prior to mass spectrometry, quantitation of

relative phosphate incorporation into a fragment, and thus determination of important regulated sites, is possible. Initial attempts to purify enough Flag-CFP1 from cells for this technique failed. Another technical impediment that must be faced with this technique was apparent from *in vitro* phosphorylation and mass spectrometry studies. Proteolysis of CFP1 with various proteases yielded fragments that were rich in serine and threonine residues. Assignment of phosphorylation sites was ambiguous for these fragments, a limitation that would extend to other mass spectrometry-based techniques for CFP1 analysis.

The outcome of *in vitro* kinase assays to compare incorporation of phosphate into point mutants versus the wild type protein were not a surprise in the face of the mass spectrometry results. Anticipating that any one phosphorylation site would substantially reduce the overall phosphate incorporation when so many sites were modified seemed improbable. However, in the single replicate of the time course that we performed comparing wild type, S224D and T227D CFP1, the observed increase in phosphorylation of both point mutants was surprising. There are a few possibilities that may account for this finding. Technical limitations related to calculating relative incorporation included difficulty definitively determining the amount of protein in each lane. Despite equivalent loading of protein into each reaction, the observed band shifts coupled with band detection using the LiCor system gave very uneven calculations of relative loading, and instead of normalizing each band individually, all samples were normalized to the zero time point which was clearest between mutants. A more interesting possibility may be that phosphomimetics at these sites could serve as priming phosphorylation events for other sites. Other potential phosphorylation sites are currently being mutated for future analysis.

Endogenous CFP1 often appears as a doublet by immunoblotting, although the underlying reasons for this are unclear. This may represent a splice isoform that has been previously identified, bearing four additional amino acids (UniProt ID: Q9P0U4-2). Alternatively, this may represent a post-translationally modified form of CFP1, potentially acetylated or phosphorylated, and CFP1 methylation has also been reported (252). It seems likely that these bands represent differentially modified forms of CFP1. Standard 30:1 acrylamide-bis acrylamide ratios usually do not clearly distinguish two bands for overexpressed Flag-tagged CFP1. However, blots of Flag-CFP1 for acetylation (Figure 3-11) do appear to be shifted relative to the bulk of the total pool of Flag-CFP1, suggesting that the second band represents acetylated

protein. Finally, two predominant bands appear in the Phos-tag blots (Figure 3-10), both for endogenous and overexpressed protein. These major bands appear to be insensitive to stimulatory treatments as well as ERK1/2 and PI3K inhibition, suggesting that they may not represent a phosphorylation event. In order to definitively show that a given band represents acetylated CFP1, we would need to identify and mutate sites of modification or the acetyltransferase(s) responsible, two goals we were unable to achieve in this body of work. An alternative strategy would be to test deletions encompassing groups of sites.

CHAPTER FOUR
Outcomes of ERK1/2 Activity on CFP1 Function as a Component of the Set1A/B Complexes

INTRODUCTION

We identified CXXC finger protein 1 (CFP1), a conserved component of the Set1A/B complexes (125), as a novel ERK1/2-interacting protein and substrate. CFP1 has previously been demonstrated to play an essential role in properly targeting the Set1A/B complexes to promoters where they trimethylate H3K4 (4), a histone modification that supports transcription of the underlying DNA. Inducible post-translational modification to proteins that mediate epigenetic events, such as the Set1A/B and related MLL 1-4 (mixed lineage leukemia) complexes, represents a relatively underexplored field.

One report indicates that a shared subunit of all of these complexes, Ash2L, and its associated H3K4 trimethylating activities are recruited to muscle-specific loci during myogenesis by the ERK1/2-related MAPK p38 (193), suggesting that MAPK activity may be generally important for dictating inducible H3K4 trimethylation. The best evidence to date that ERK1/2 is important for coordinating this activity comes from *Drosophila melanogaster*, where the transcriptional regulatory protein Corto, which shares best homology to the human protein MLL2, binds to the ERK1/2 homolog *rolled* as well as the ERK1/2 scaffold MP1 (191). Importantly, the Corto and *rolled* knockout phenotypes are similar. Additionally, ERK1/2 and H3K4 trimethylation colocalize extensively on chromatin by ChIP-sequencing analysis in mouse embryonic stem cells, further implicating a functional relationship (110).

Several recent studies have suggested that H3K4 trimethylation is induced by various stimuli to support transcriptional responses. Following DNA damage, a host of transcriptional regulators are recruited to target gene promoters, including the transcription factor p53, the lysine acetyltransferase p300 and Set1 complexes to stress-responsive target gene promoters, leading to accumulation of H3K4 trimethylation and transcription of associated genes (253). CFP1 makes an essential contribution to deposition of H3K4 trimethylation to these promoters, but is not universally essential for transcriptional output (254). The impact of post-translational modifications in targeting the Set1A/B complexes to specific subsets of genes remains unclear.

We were particularly intrigued by the role that CFP1 phosphorylation might play in inducibly directing H3K4 trimethylation towards immediate early gene (IEG) promoters, which integrate ERK1/2 signaling through multiple effector proteins to activate transcription.

IEGs represent the first wave of transcriptional response following mitogenic stimulation as they require no new protein synthesis for their expression (255). These genes frequently encode transcription factors that mediate a broad range of secondary responses. ERK1/2 are essential for inducible IEG mRNA changes through substrate targeting of multiple factors to these genes, including transcription factors (161), histone modifiers (16), and the RNA polymerase II-associated Mediator complex (17,18). Collaborative ERK1/2-driven recruitment and activity of these components is a prerequisite for transcription. ERK1/2, upstream signaling components and the ERK1/2 kinase substrate MSK (mitogen and stress-activated kinase) are recruited directly to IEG promoters (174,256), where they may target these and other promoter-restricted substrates, further enforcing transcriptional activation of target genes.

We chose a standard signaling paradigm (serum treatment of HeLa cells) in which to characterize the interplay between CFP1-driven transcription and ERK1/2 activation. There were several reasons for this choice. First, the Set1A/B complexes have been extensively characterized in this cell line by other groups (128,210). We knew that serum treatment induced ERK1/2 activation and robust nuclear translocation (75), and many other rapidly-inducible ERK1/2-dependent transcriptional effects have been characterized in response to serum treatment. ERK1/2 induction frequently increases IEG mRNA concentrations as a consequence of transcription. We hypothesized that part of the rapid ERK1/2-dependent transcriptional response hinged on direct phosphorylation of CFP1, potentially altering the activity or localization of Set1A/B complexes to support IEG transcription.

EXPERIMENTAL PROCEDURES

Cloning and Constructs

pCMV5-eGFP C1 was used as a readout for transactivation assays. p3xFLAG-CMV-7.1 CFP1 mutants were generated by standard site mutagenesis techniques with the following primers: S224A 5'- CCCTTCCTCGCTCGCCCCAGTGACGCCCTC-3'; S224D 5'-CCCTTCCTCGCTCGACCCAGTGACGCCCTC-3'; T227V 5'-CTCGCTCTCACCAGTGG

TGCCCTCAGAGTCCC-3';   T227D   5'-CTCGCTCTCACCAGTGGACCCCTCAGAGTCCC-3';. pCMV5-MycSet1B 1185-1985 was generated by PCR with murine cDNA from Origene (BC038367). The amplified fragment and vector were digested with EcoRI and SalI and ligated. All constructs were sequenced to ensure no spurious mutations were introduced.

siRNA

Duplex siRNA oligonucleotides targeting human CFP1 (s26937) or a nontargeting control (4390843) were acquired from Life Technologies. Oligos were reverse transfected into HeLa cells at 20 nM with RNAiMax (Life Technologies) according to the manufacturer's instructions.

Immunoprecipitation from nuclei

HeLa cell nuclei were isolated following 48 hours of transfection as described in Chapter 3. For co-immunoprecipitations between Flag-CFP1 and Myc-Set1, nuclei were resuspended in buffer containing (20 mM Tris HCl (pH 7.4), 400 mM NaCl, 25% glycerol, 5 mM EDTA, 0.1% NP-40, 1 mM DTT, 80 mM β-glycerophosphate, 100 mM NaF and 2 μM Na_3VO_4) supplemented with protease inhibitors and 400 nM Trichostatin A and extensively Dounce homogenized. Nuclei were clarified by centrifugation and supernatants were used for immunoprecipitation with anti-Myc antibody and anti-mouse magnetic Dynabeads for 2 hours at 4°C. Resin was washed 3 times with immunoprecipitation buffer, and denatured with Laemmli buffer.

Immunoblotting and antibodies

SDS-PAGE or tricine gels (242) were used to separate proteins that were transferred to nitrocellulose membrane (Millipore). Membranes were blocked with LiCor blocking buffer and primary antibodies were incubated with membranes overnight. LiCor fluorescent secondary antibodies were used at 1:15,000 and immunoblots were imaged and bands quantified using a LiCor system. Western blot analysis was performed using the following primary antibodies: rabbit anti-CFP1 (Bethyl Laboratories); rabbit anti-ERK1/2 (produced in-house, (243)); mouse anti-phospho-ERK1/2 (T185/Y187) and mouse anti-Flag M2 (Sigma); α-tubulin (mouse

monoclonal antiserum generated in-house from cells obtained through Developmental Studies Hybridoma Bank); mouse anti-Histone 3 and rabbit anti-H3K4 trimethyl (Active Motif); rabbit anti-green fluorescent protein (Santa Cruz); mouse anti-Myc (National Cell Culture Center).

Microarray

Total RNA was prepared from HeLa cells with by Pure-link RNA Mini Kit (Life Technologies) and submitted to the UT Southwestern Microarray Core for analysis. Briefly, RNA was checked for concentration and quality using an Agilent 2100 Biolanalyzer, then cRNA was synthesized and labeled prior to hybridization to an Affymetrix Human Transcriptome 2.0 Array chip and detection. Raw data was analyzed using Affymetrix Transcriptome Analysis Console 2.0 software.

qRT-PCR

Total RNA was isolated with TRI reagent (Applied Biosystems). Complementary DNA was synthesized using iScript cDNA Synthesis Kit (Bio-Rad) according to the manufacturer's instructions. Reactions were carried out with diluted cDNA, SYBR Green Supermix (Bio-Rad) and fluorescence was measured using a quantitative real-time thermocycler (Bio-Rad). Relative changes in gene expression were calculated by $2^{-\Delta Ct}$, and Actin was used as an internal expression control. Primers were as follows: Actin 5'- AGGTCATCACTATTGGCAACGA -3'(forward) and 5'-CACTTCATGATGGAATTGAATGTAGTT-3' (reverse); Set1A 5'- CAGTGGCGGAACTACAAGCTC-3' (forward) and 5'-CATAGCGGTACACCTTCTGAGA-3' (reverse); Set1B 5'- CCGGTGGAAATTGTCGAAGAT-3' (forward) and 5'-GCTCCTTG TTTTTGGTCCAGAT-3' (reverse); EGR1 5'- CAGCACCTTCAACCCTCAG-3' (forward) and 5'- AGCGGCCAGTATAGGTGATG-3' (reverse); EGR3 5'- GACATCGGTCTGACCA ACGAG-3' (forward) and 5'- GGCGAACTTTCCCAAGTAGGT-3' (reverse); FOSB 5'- GCTGCAAGATCCCCTACGAAG-3' (forward) and 5'- GCTGCAAGATCCCCTACGAAG-3' (reverse); DUSP1 5'- ACCACCACCGTGTTCAACTTC-3' (forward) and 5'- TGGGAGAGG TCGTAATGGGG-3' (reverse); DUSP5 5'- GCCAGCTTATGACCAGGGTG-3' (forward) and 5'-GTCCGTCGGGAGACATTCAG-3' (reverse); KLF10 5'- ACTGCCAAACCTCACATTGC-

3' (forward) and 5'- ACGAATCACACTTGTTGCCTG-3' (reverse); BHLHE40 5'-GACGGG GAATAAAGCGGAGC-3' (forward) and 5'- CCGGTCACGTCTCTTTTCTC-3' (reverse).

Statistical Analysis

Results were expressed as means ± Standard Deviation determined from three independent experiments. Statistical significance was calculated using the GraphPad Prism software package using Student's *t* test.

RESULTS

CFP1 knockdown results in a global deficit of trimethylated H3K4

We initially wanted to characterize our ability to knock down endogenous CFP1 with targeted siRNAs and evaluate the impact of CFP1 loss on ERK1/2 pathway activation (Figure 4-1). HeLa cells were treated with either control or CFP1-targeted siRNA oligonucleotides for 72
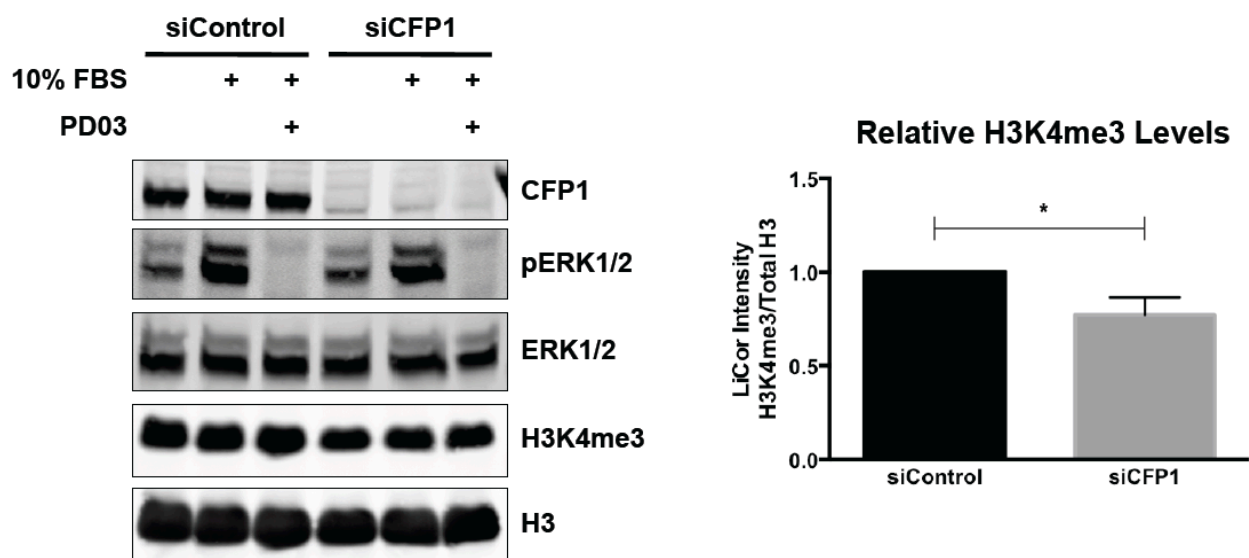
Figure 4-1: FBS-stimulated ERK1/2 signaling and H3K4 trimethylation in CFP1-depleted HeLa cells. Cells were transfected with either siRNA against CFP1 or a control target for 72 hours, then incubated in serum-free DMEM prior to a 15 minute treatment with 10% FBS. Five minutes before stimulation, cells were treated with either 500 nM PD0325901 or DMSO. Cells were then lysed and proteins resolved by SDS-PAGE prior to blotting with the indicated antibodies. Signal intensity of H3K4me3 relative to total H3 under the untreated conditions for each siRNA treatment was quantified using the LiCor imaging system, and relative modification

normalized to total H3 was plotted. Error bars indicate standard deviation of three separate experiments. Significance was quantified by Student's two-tailed t-test; * = p < 0.05.

hours prior to a brief serum starvation. We then stimulated ERK1/2 with 10% FBS, with or without pretreatment with the MEK1/2 inhibitor PD0325901. CFP1 knockdown was consistently efficient, typically approaching a depletion of 90% reduction of immunoreactive protein. This had no significant impact on our ability to stimulate ERK1/2 activity with serum. In the same experiment, we blotted for H3K4 trimethylation, as well as total H3. H3K4 trimethylation was insensitive to short treatments with PD0325901 and FBS. However, knockdown of CFP1 resulted in an average reduction in trimethylation of over 20% compared to controls, indicating that it is essential for some H3K4 trimethylation in this cell type.

CFP1 depletion results in the specific upregulation of Set1B transcript

In a prior report, loss of CFP1 in mouse embryonic stem cells resulted in reduced Set1A methyltransferase protein expression (220). However, global H3K4 trimethylation remained unaltered in these cells (210). CFP1 restricts methyltransferase activity to specific genomic loci, and knockout in mouse embryonic stem cells results in diminished promoter H3K4 trimethylation at a subset of targets, as well as spurious catalysis of this marking outside of normal target sites (4). We wanted to measure the relative expression levels of Set1A and Set1B in our system, since we saw a pronounced loss of H3K4 trimethylation with CFP1 knockdown. We hypothesized that Set1A/B levels might be similarly diminished under these conditions.

Attempts to immunoblot Set1A and Set1B methyltransferases with several commercial antibodies failed. Therefore, we chose to measure relative mRNA concentrations of these subunits by quantitative real-time PCR (qRT-PCR) following 72 hours of CFP1 knockdown, compared to a control condition (Figure 4-2). We expected that, as CFP1 is a stable component of both the Set1A and Set1B complexes, and because their subunit compositions are nearly identical (128), that CFP1 depletion would have a similar impact on both. To our surprise, CFP1 loss did not have a significant impact on the expression of Set1A, but specifically increased expression of Set1B mRNA by nearly 1.5 fold.

Although we were unable to confirm that this change in mRNA expression extended to Set1B protein, this does suggest that compensation for CFP1 loss may occur as an upregulation of remaining complex components. We have not yet explored expression changes in other shared components of these complexes. It is also notable that reports on events that differentially
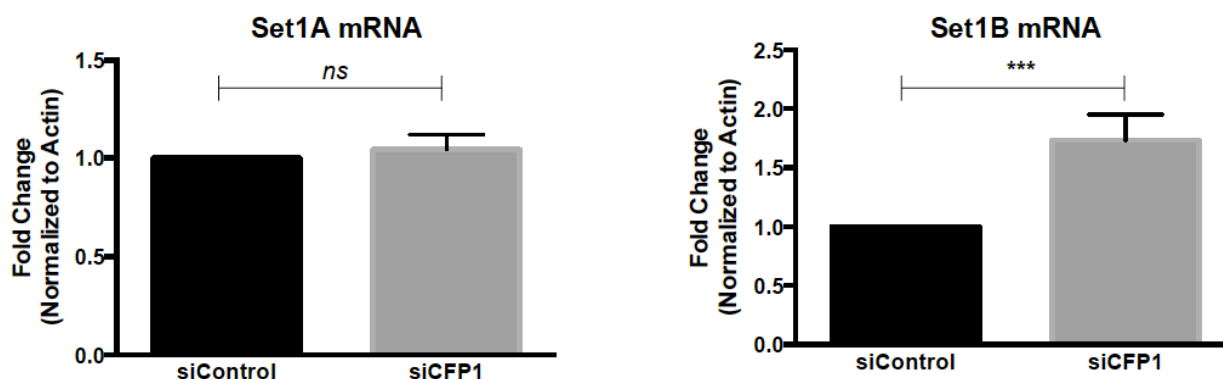
Figure 4-2: Knockdown of CFP1 increases Set1B mRNA, but not that of Set1A. HeLa cells were subjected to 72 hours of siRNA treatment with either control or CFP1-targeted oligonucleotides. Cells were then lysed with Trizol, and mRNA was extracted. RT-PCR was performed for the indicated target mRNAs. Quantitation was normalized against the control condition. Error bars show the standard deviation of three experiments. Statistical significance was assessed by Student's two-tailed t-test; *** = $p < 0.0005$.

regulate Set1A and Set1B methyltransferase complexes are limited (257). It is clear that they target distinct subsets of genes (210) but how and why regulation occurs remains an open question.

CFP1 T227V fails to support global H3K4 trimethylation

Our first priority was to assess the importance of residues that we identified as putative ERK1/2 substrate sites, S224 and T227, in CFP1 function. These two sites were the most highly predicted sites by Scansite, appear in multiple studies compiled on Phosphosite, and were high confidence sites in our mass spectrometry results. We reasoned that CFP1 phosphorylation by ERK1/2 might generally direct Set1A/B complex targeting.

We compared the impact of overexpression of wild type CFP1, as well as S224A and T227V, on global deposition of H3K4 trimethylation (Figure 4-3). Overexpression of Flag-CFP1 or the S224A mutant did not appreciably impact H3K4 trimethylation immunoreactivity. CFP1 T227V reduced detected trimethylated H3K4 by nearly 25%, a change strikingly similar to CFP1

knockdown. This suggested to us that T227 was important for the function of CFP1 in the Set1A/B complexes.
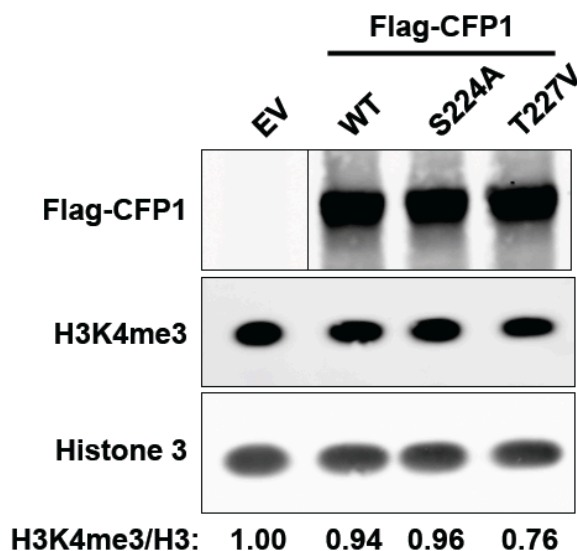
Figure 4-3: Overexpression of CFP1 T227V results in global H3K4me3 defects. HeLa cells were transfected with the indicated CFP1 constructs or an empty vector control (EV) for 48 hours. Cells were lysed and proteins resolved by SDS-PAGE prior to blotting with the indicated antibodies. Signal intensity of H3K4 trimethylated relative to total H3 was quantified using the LiCor imaging system, and relative modification normalized to total H3 was calculated. Representative of a single replicate.

CFP1 transactivation assays reveal a requirement for S224

In an effort to look at the general transcriptional effects of CFP1 compared to potential phosphorylation site point mutants, we turned to a transcriptional transactivation assay (245). This assay relies on the ability of CFP1 to specifically bind and transactivate the CpG-rich cytomegalovirus (CMV) promoter linked to a green fluorescent protein (GFP) reporter. Transactivation of the CMV promoter by CFP1 leading to GFP expression is dose-dependent. Mutants of CFP1 that are unable to bind to DNA fail to induce detectable expression of GFP in this experimental paradigm. Based on data indicating that CFP1 T227V introduction blocked global H3K4 trimethylation, we expected that it would show defects in transactivation of GFP expression.

HeLa cells were transfected with pCMV-GFP along with varying concentrations of Flag-CFP1, S224A or T227V mutants for 48 hours prior to lysis and blotting (Figure 4-4). The wild type protein induces GFP protein expression in a dose-dependent fashion, as previously reported.
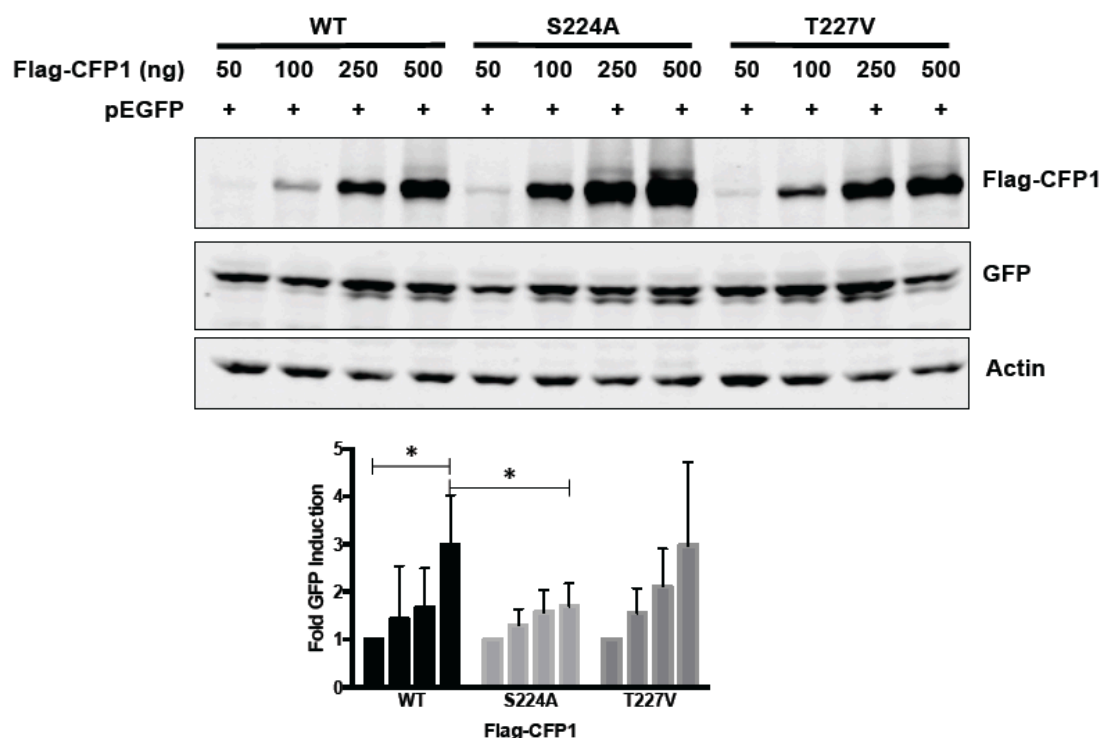
Figure 4-4: CFP1 S224A fails to robustly transactivate a CpG-rich promoter. HeLa cells were transfected with equivalent amount of pEGFP plasmid and a range of concentrations of the Flag-CFP1 constructs indicated. Lysates were resolved by SDS-PAGE and blotted with the indicated antibodies. Signal intensity of GFP normalized to actin was quantified using the LiCor imaging system, and relative modification normalized to total actin was plotted. Error bars indicate standard deviation of three separate experiments. Significance was quantified by Student's two-tailed t-test; * = $p < 0.05$.

To our surprise, the T227V mutant induces GFP expression to a similar extent. However, CFP1 S224A, in spite of its higher protein expression per transfected DNA, particularly at 250 and 500 ng, fails to induce GFP expression to a similar extent as wild type protein. These data indicated that S224 is important for transactivation of this construct, and that T227 is dispensable.

Flag-CFP1 point mutants localize to the nucleus

Phosphorylation of many transcriptional regulators impacts their subcellular distribution. We wanted to ensure that the phosphorylation point mutants that we had generated for CFP1, including S224A, S224D, T227V and T227D, localized correctly. We transfected equivalent amounts of Flag-CFP1 and mutants, into HeLa cells (Figure 4-5). After 48 hours, we fractionated
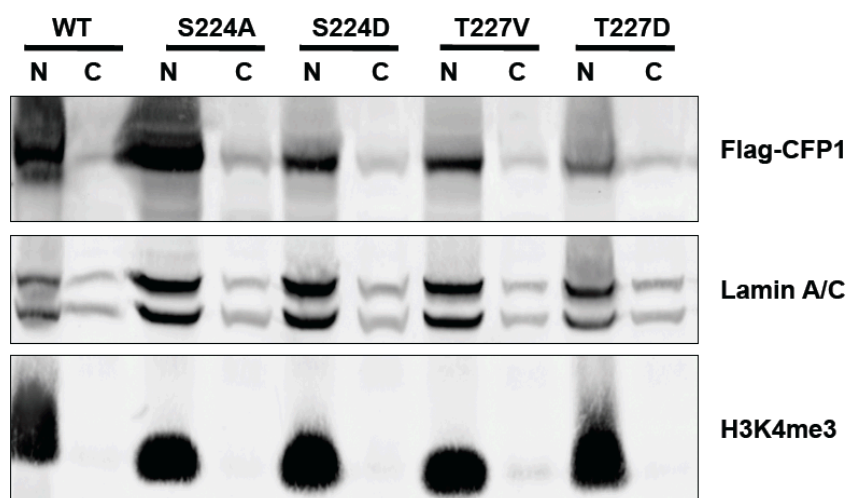
Figure 4-5: Potential phosphorylation and acetylation site mutants of CFP1 localize to the nucleus. Equivalent concentrations of Flag-CFP1 plasmid bearing the indicated point mutants were transfected into HeLa cells. Nuclei were isolated by mild detergent treatment and centrifugation, and nuclear and cytosolic fractions were brought to equal volumes prior to separation by SDS-PAGE. Membranes were immunoblotted for Flag-CFP1 and with antibodies against lamin A/C and H3K4me3 to roughly assess purity of the nuclear fraction. Representative of a single replicate.

nuclei from cytosol by differential centrifugation, and blotted the fractions for Flag-CFP1 and two nuclear markers, trimethylated H3K4 and lamin A/C.

Fractionations were relatively clean and consistent, as judged by marker immunoblotting. All point mutants predominantly co-fractionationated with nuclear markers, similar to the wild type overexpressed protein. Despite transfection of equivalent DNA concentrations into cells, we consistently observed that S224A expresses to a greater extent than the wild type protein, and that T227D was difficult to express at detectable levels. The exogenous expression of these mutants is under control of a CMV promoter, which wild type CFP1 itself can positively

regulate, so changes in expression could reflect transactivation capacity, translated protein instability, or both.
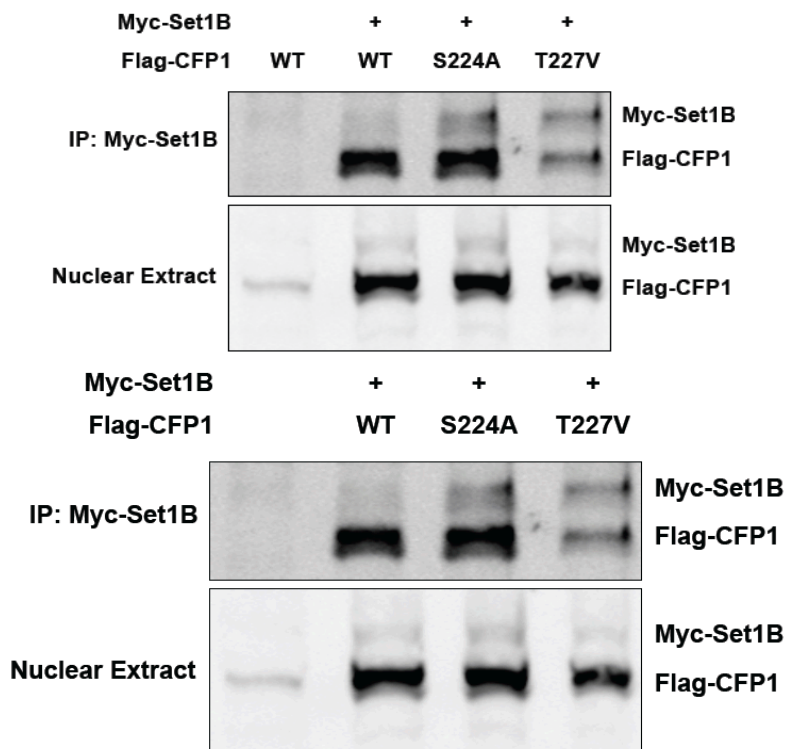
Figure 4-6: Myc-Set1B 1185-1985 and Flag-CFP1 mutants co-immunoprecipitate from HeLa cells. HeLa cells were transfected for 48 hours with the indicated constructs. Immunoprecipitations were performed on crude nuclear fractions with Myc-targeted antibody, and bound proteins were isolated with anti-mouse Dynabeads. Immunoprecipitates and nuclear lysates were resolved by SDS-PAGE and blotted with Myc and Flag antibodies. Representative of two separate experiments.

Interaction of Flag-CFP1 S224A and T227V mutants with a Myc-Set1B fragment

One possible consequence of CFP1 phosphorylation may be altered interaction with binding partners. Observed deficiencies in transactivation and methyltransferase assays could be a consequence of poor interaction of mutants with Set1A/B complexes. We next wanted to determine whether the S224A and T227V mutants interacted efficiently with Set1 complexes. To circumvent our difficulties with endogenous Set1A/B blotting, we subcloned mouse Set1B residues 1185-1985 into a pCMV5 vector incorporating an N-terminal Myc tag. This fragment of the human protein had previously been shown to be sufficient for interaction with CFP1 (23).

Due to differences in protein expression between various Flag-CFP1 point mutants, we attempted to titrate transfected DNA to achieve a similar extent of Flag-CFP1 overexpression while simultaneously maintaining the same relative amount of Myc-Set1B. After 48 hours, we isolated a crude nuclear fraction from transfected cells, then immunoprecipitated using a monoclonal antibody directed against Myc (Figure 4-6).

Wild type and both point mutants were pulled down to a similar extent by this Myc-Set1B fragment. The total amount of co-immunoprecipitated Flag-CFP1 T227V is less than either the wild type or S224A forms, which may reflect a weaker interaction, but also may be a consequence of rapid degradation, an effect we consistently observed with this point mutant. Similar experiments in which nuclear lysates are held at 4°C for the duration of immunoprecipitation exhibit a similar extent of signal loss for input and immunoprecipitated CFP T227V. Preliminary attempts to examine whether acute ERK1/2 pathway stimulation or treatment with PD0325901 impacted interaction did not reveal obvious stimulus-responsive binding.

CFP1 supports transcription of serum-inducible genes

We were curious about the interplay of CFP1 and ERK1/2 activities on gene expression, and sought to gain a broad view of the impact of CFP1 depletion on ERK1/2-dependent mRNA induction. To this end, we performed a microarray experiment either in control or CFP1-depleted HeLa cells with the following three conditions: serum-starved, stimulated with serum for 30 minutes, with or without pretreatment with PD0325901 (Figure 4-7).

With assistance from the UT Southwestern Microarray Core facility, we assayed gene expression profiles using the Affymetrix Human Transcriptome 2.0 array. Our goal was to identify serum-induced genes on a short time scale, then ask whether they were similarly regulated following CFP1 knockdown. We set a relatively low threshold of fold expression difference (1.45) to assign serum-induced change, and limited our initial analysis to protein-coding mRNAs. In cells treated with control oligonucleotides, 10% serum treatment for 30 minutes resulted in upregulation of 55 mRNAs (Selected regulated genes in Table 4-1; Appendix C contains full data table). Of these, 29 were reduced by at least 25% with PD0325901 pre-treatment. Many of these genes have been previously classified as IEGs.

Changes induced by serum treatment between control and CFP1-depleted cells were in good agreement by trend. Only five genes were more highly induced by serum in the CFP1-depleted condition than the control. However, the degree of induction between these two sets differed considerably. Induction of some highly expressed targets, including EGR1, FOSB and ZFP36, was reduced by 50% or more by CFP1 depletion. Other targets showed a more modestly reduced induction under these conditions.

We were struck by the extent of overlap between FBS-stimulated genes under control knockdown conditions, and genes that were upregulated in response to CFP1 knockdown in the absence of stimulus (Appendix C). Ten of the 87 most-induced genes by CFP1 knockdown were also on the list of serum-induced genes under control knockdown conditions. Upregulated targets

P1 Basal	siCFP1 PD03/Basal
9	1.01
2	1.27
2	1.19
9	1.33
6	1.22
	1.73
1	1.15
6	1.03
7	1.43
2	1.24
04	1.32
2	3
4	1.64
7	1.14
4	1.36
1	2.1
4	1.11
4	1.89
1	1.43
1	1.24
6	1.85
7	-1.07
8	1.07
8	1.43
2	1.37
4	-1.13
6	-1.01
9	-1.08
3	1.12

Table 4-1. mRNA-encoding genes most induced by serum treatment. HeLa cells were subjected to 72 hours of siRNA treatment with either control or CFP1-targeted oligonucleotides. Cells were serum starved, pretreated with 500 nM PD0325901 (PD03) or DMSO, then treated with 10% fetal bovine serum (FBS) for 30 minutes prior to lysis and RNA purification. Samples were submitted to the UT Southwestern Microarray Core for hybridization to an Affymetrix Human Transcriptome 2.0 chip and detection of relative RNA levels. Fold change was assessed using Affymetrix Transcriptome Analysis 2.0 software.

by CFP1 knockdown (including noncoding RNAs, totaling 254 targets) were analyzed using the Qiagen Ingenuity Pathway Analysis (IPA) software. The top hit for upstream regulatory proteins for this gene expression profile was MAPK1 (ERK2).

Microarray validation

To place these results in context, the outcome of our microarray experiment needed independent validation. We were specifically interested in identifying protein-coding genes whose mRNA increased in response to FBS stimulation under control conditions, but were hampered by CFP1 depletion for a more detailed dissection of the contribution that CFP1 phosphorylation might make to transactivation.

We selected several genes to test by qRT-PCR, including IEG transcription factors (EGR1, EGR3 and FOSB), inducible MAPK phosphatases that display activity towards ERK1/2 (DUSP1 and DUSP5), and two other transcription factors that were substantially downregulated in CFP1-depleted basal conditions (BHLHE40 and KLF10). Experiments were repeated in triplicate, following the same conditions as were used to prepare samples for the microarray. We analyzed the data as relative enrichment in two ways: first with all conditions normalized to the basal stimulation for the nontargeting control oligonucleotide knockdown (Figure 4-7), then by fold change over the basal condition by each knockdown individually (Figure 4-8).

By fold induction over the control Basal condition, CFP1 knockdown invoked subtle but significant changes on basal and serum-induced mRNA concentrations of many of the genes examined. The most consistent significant result was that the fold induction of mRNA following serum treatment was less in the siCFP1-treated cells than the control set, noted for all assayed transcripts but EGR1. To assess whether induced changes were a consequence of altered basal expression, we also looked at these genes, comparing fold induction over basal expression for the control and CFP1 depleted conditions separately. Two gene targets, FOSB and EGR3, were still significantly downregulated in the CFP1 knockdown condition. This second analysis also identified a strong reduction in inducible EGR1 transcript, not seen when all samples are normalized to basal control values. This is likely the consequence of increased basal transcription of EGR1 under siCFP1 conditions (where induced values with serum treatment are similar in the knockdown and control conditions). This particular analysis would benefit from comparison to a

Figure 4-7: Fold induction of ERK1/2-regulated genes with and without CFP1 depletion. HeLa cells were subjected to 72 hours of siRNA treatment with either control or CFP1-targeted oligonucleotides. Cells were lysed with Trizol and mRNA was extracted. qRT-PCR was performed for the indicated target mRNAs. All changes were normalized to the siControl basal condition. Error bars represent standard deviation of three separate experiments. Statistical significance was assessed by Student's two-tailed t-test; * = p < 0.05, ** = p < 0.005 and *** = p < 0.0005.

Figure 4-8: Fold induction of ERK1/2-regulated genes with and without CFP1 depletion. HeLa cells were subjected to 72 hours of siRNA treatment with either control or CFP1-targeted oligonucleotides. Cells were lysed with Trizol and mRNA was extracted. qRT-PCR was performed for the indicated target mRNAs. All changes were normalized to the knockdown-matched basal condition. Error bars represent standard deviation of three separate experiments. Statistical significance was assessed by Student's two-tailed t-test; ** = $p < 0.05$ and *** = $p < 0.005$.

standard curve to assess transcript copy number as an unbiased means of comparison between conditions.

DISCUSSION

The literature concerning the function that CFP1 performs in H3K4 trimethylation is limited mainly to mouse embryonic stem cells (mESCs; (210)), where loss has been reported to have no detectable effect, or even increase global H3K4 trimethylation (220). Differentiation of these cells, induced by cytokine removal, resulted in a four-fold increase in this marking. Further investigation in this knockout mESC line revealed widespread dysregulation of H3K4 trimethylation deposition throughout the genome that was interpreted to indicate altered regulation of Set1A/B complex localization (4). Therefore, we did not anticipate the decrease in H3K4 trimethylation levels that we observed following CFP1 knockdown in HeLa cells.

In contrast to data from mESCs, loss of the CFP1 homolog from budding and fission yeasts had a profound impact on H3K4 trimethylation with an 80% reduction found in *S. cerevisiae* (121) and a complete loss of detectable signal observed in *S. pombe* (218). The impact observed in HeLa cells compared to mESCs may reflect effects restricted to particular differentiation states. H3K4 trimethylation plays a special role in ESCs, where in addition to marking transcriptionally active promoters, it is present at bivalent domains. Bivalent domains are promoters where both H3K4 and inhibitory H3K27 trimethylation are present in undifferentiated cells, and represent chromatin primed to be turned either "on" or "off" for tissue-specific development (258). Although H3K4 trimethylation at bivalent promoters has been attributed to the MLL2 methyltransferase complex (127), it remains a possibility that global regulation of this marking is very different in ESCs than in terminally differentiated cells. Studies to date on CFP1 in differentiated cells have been extremely limited. Evidence to support this hypothesis is limited by the embryonic lethality of CFP1 loss in mice, and neither conditional knockouts nor the effects of knockdown in other cell lines have been reported.

We had hoped to bypass the technical limitations of phosphorylation site identification by assessing phosphorylation point mutant competency in functional assays. We chose two sites that we reasoned were most likely to be phosphorylated (S224 and T227) based on prediction from primary amino acid sequence and presence in our mass spectrometry analysis, anticipating that one or both might be important for global H3K4 trimethylation and transcriptional transactivation. Our data indicated that CFP1 T227V, but not S224A, blocks global H3K4 trimethylation to a similar extent as CFP1 knockdown in HeLa cells.

As a consequence, we expected that the T227V mutant would also fail to appropriately transactivate a CpG-rich target promoter. However, T227V functioned as well as the wild type overexpressed protein. In contrast, the other mutant that we assayed, S224A, did a poor job of inducing transactivation, even at very high concentrations. This was particularly unexpected, because CFP1 T227V typically expresses at approximately 50% the extent of wild type protein, and S224A expresses about twice as well. CFP1 transactivates its own expression, as the pCMV plasmids regulate protein expression by a CpG-rich promoter, so differential expression could be the result of transcriptional effects or protein stability changes. Neither site appears to result in a gross impact on nuclear localization or integration into a complex with Set1B, suggesting that

changes to function might be the result of targeting to chromatin. Both of these sites appear to positively contribute to the function of CFP1 in supporting transcriptional activation.

The discrepancy between functional outcomes of these two mutants may lie with differences in the specific functions of phosphorylation at these sites, but also may reflect limitations of the assays used to study them. Although we observe a loss in global H3K4 trimethylation after introduction of CFP1 T227V, further work is required to verify that this reflects decreased trimethylation at target promoters and that this mutant impairs transcription of CFP1-regulated target genes.

Problems of interpretation exist for the transactivation assay. Results are not only confounded by auto-regulation of Flag-CFP1 expression, but also CFP1 roles in translation initiation and protein turnover. In one report, CFP1 loss in murine ESCs potentiates phosphorylation of eIF2α, resulting in decreased translation initiation (23) This same report indicated that translated DNMT1 protein stability was impaired, although whether this extended to other proteins was unclear. This means that the outcome of the transactivation assay, which examined steady-state expression of GFP, is suspect without additional validation of CFP1 phosphorylation on transcriptional effects. Issues with self-regulated transcription of CFP1 constructs from the CMV promoter might be evaded by subcloning these constructs into plasmids bearing other promoters, by making genomically integrated inducible forms of the protein or by replacing the wild type endogenous protein using Cas/CRISPR technology (259).

Finally, we have not ruled out the possibility that the effects we have seen are not the consequence of CFP1 interaction with DNMT1. This seems less likely than effects mediated by the Set1A/B complexes, given that H3K4 trimethylation is typically associated with subtle changes in gene expression, whereas DNA methylation robustly silences gene expression. None of the phenotypes we observed suggested this type of regulation, and cytosine and H3K4 methylation in mammals appear to be generally uncoupled from one another (4).

We are currently generating point mutants of the other most probable phosphorylation sites based on a combination of bioinformatics data and our *in vitro* mass spectrometry results. For large stretches with multiple potential phosphorylation sites, we are making short deletions from the protein for use in functional assays. It is an additional possibility that several sites are

phosphorylated and important for functional effects. Whether they act in a concerted manner or oppose one another will have to be tested.

If ERK1/2 can direct accumulation of H3K4 trimethylation to target promoters by phosphorylating CFP1, this would parse well with the existing literature on ERK1/2-mediated immediate early gene expression. Phosphorylation of multiple components of the transcriptional machinery, transactivating proteins and histone modification complexes collectively aid in promoting target gene expression. H3K4 trimethylation would serve as another layer of regulation to positively reinforce transcription of targeted genes. Moving forward, we are interested in determining whether CFP1 phosphorylation is important for gene expression, what the mechanisms of this process may be, and whether the importance of this event extends to genes other than just those that are inducibly ERK1/2-regulated.

Evidence from other inducible transcriptional events has suggested that H3K4 trimethylation can be upregulated at target genes in response to pathway activation, though documented instances occur on longer time scales than we tested. p38-directed deposition on H3K4 trimethylation to muscle-specific genes during myogenesis occurred over a period of days (193). p53-dependent Set1 activity at stress-responsive loci was examined hours after stimulation (252). We may not see acute changes at ERK1/2-dependent immediate early gene targets in response to serum stimulation.

This would not necessarily preclude the hypothesis that CFP1 phosphorylation by ERK1/2 induces H3K4 trimethylation at serum-inducible genes. Many inducible genes have paused RNA polymerase II resident at their promoters, waiting for a strong transactivation signal to induce rapid activation. In such a case, we might find that H3K4 trimethylation is constitutively maintained at a high level, potentially maintained by basal ERK1/2 signaling. We hope to clarify the role that ERK1/2 phosphorylation plays by assaying promoter-specific H3K4 trimethylation levels by chromatin immunoprecipitation under various states of serum and inhibitor treatment. Furthermore, we plan to overexpress phosphorylation site point mutants of CFP1 to assess the role of specific phosphorylation events in gene transcription and histone methylation under both basal and serum-induced conditions.

To this point, our assays had examined the steady-state involvement of CFP1 and various point mutants in physiological readouts. In our first step towards looking at the role that CFP1

plays in inducible transcriptional responses, we sought to define genes rapidly upregulated by serum that were ERK1/2 dependent. Our rationale was that if CFP1 were an ERK1/2-responsive substrate, direct transcriptional effects of CFP1 phosphorylation might be most clear in early rounds of transcription prior to significant contribution of newly-translated proteins. We identified 29 genes for which ERK1/2 clearly contributes to gene expression. A majority of these were not induced to the same extent in CFP1-depleted cells. Seven genes were selected for validation by qRT-PCR, which highlighted the fact that CFP1 has an inconsistent effect on basal regulation of many genes, but that serum-induced expression changes can by significantly hampered by CFP1 loss.

In our analysis of the microarray data set, one startling observation we made was that there was substantial overlap between protein-coding genes upregulated by serum and transcripts basally increased by CFP1 knockdown, with both lists sharing the same top hit: EGR1. Nine other transcripts, including FOS, FOSB, JUNB, BTG2 and ZFP36, all substantially upregulated in the presence of serum, exhibited a high degree of basal upregulation in the absence of CFP1. This certainly suggests that both ERK1/2 and CFP1 are important for the expression of this subset of genes, though what it means for serum-dependent CFP1 function is murkier.

At face value, this observation appears to be at odds with the idea that CFP1 has a positive effect on transcription. We anticipated that CFP1 knockdown would result in diminution of induced gene expression, or perhaps decreased basal expression. Our observations suggest the opposite. The reasoning for this may underscore the requirement for basal expression of these genes for continued cell survival. It will be interesting to determine if overexpression of any of the ERK1/2 phosphorylation point mutants phenocopies this effect.

Basal upregulation of these genes may be a consequence of their general importance in cell survival, although what compensatory mechanisms are employed to achieve this are unclear. Importantly, three days are required to effectively knock down CFP1 with targeted siRNAs. One topic that is often overlooked is the role of kinase induction in the "physiological" function of ERK1/2 outside of the context of an acute stimulus. They are no less important to cellular functions, particularly functions that are essential to cell proliferation, but these outcomes are frequently overlooked in favor of defining large, stimulus-dependent changes. We propose that phosphorylation of CFP1 by ERK1/2 under basal conditions may play a key role in constitutive

basal transcription of these co-regulated transcripts. The next logical step will be to determine whether loss of promoter H3K4 trimethylation contributes to the diminished expression of immediate early gene mRNAs following CFP1 knockdown. This would further implicate enzymatic activities associated with CFP1 in supporting ERK1/2-dependent transcription. Furthermore, it will be informative to observe what effect point mutant overexpression will play in basal and inducible transcript concentrations of these genes.

CHAPTER FIVE
Conclusions and Recommendations

OVERVIEW

We have demonstrated that CFP1 interacts with ERK1/2 in the context of chromatin and is an ERK1/2 substrate *in vitro* and in cells (Chapter 3). CFP1 is instrumental in eliciting full ERK1/2-directed transcriptional responses to mitogenic stimulation at multiple serum-responsive genes. Overexpressed CFP1 bearing point mutants of putative ERK1/2 phosphorylation sites function abnormally in a transcriptional transactivation assay and in analysis of global H3K4 trimethylation (Chapter 4). Together, these results suggest that ERK1/2 phosphorylation of CFP1 may be important in supporting transcription of ERK1/2 target genes. In addition to other experiments outlined in the Discussion sections of Chapters 3 and 4, we put forth the following proposals for further examination of the consequences of post-translational modification to CFP1 function.

CFP1 acetylation

We were unable to determine the functional consequences of acetylation after we failed to define sites of modification or the lysine acetyltransferase (KAT) responsible for modification. Based on our broad hypothesis that collaborative post-translational modifications occur at ERK1/2 target promoters to support transcription, we suspected that p300 might be the KAT responsible for CFP1 modification. This was based on evidence that ERK1/2 phosphorylation of p300 helps to direct its activity and transactivational capacity at target genes (16). Our lab has unpublished chromatin immunoprecipitation evidence that ERK1/2 and p300 co-occupy a large subset of promoters in pancreatic beta cells, and other reports suggest coordinated binding of these factors in other cell types (110). Set1 complexes and p300 are both recruited to p53 target genes following DNA damage (253). It therefore seemed parsimonious that p300 might target CFP1 to coordinate its recruitment to target genes.

Testing requirements for CFP1 acetylation on protein function would take an approach similar to that for phosphorylation. Identification of acetylation site(s) would be validated by mutagenesis and immunoblotting. Ideally, any acetyltransferase responsible for this modification

would be identified by measuring relative acetylation levels of CFP1 in the presence of knockdown or pharmacological inhibition of the relevant KAT. The requirement for acetylation on CFP1 function at target genes would be assessed by introduction of lysine point mutants of the protein and measuring changes in gene expression. Finally, it would be interesting to determine the extent of collaboration between the KAT and H3K4 trimethylation on regulated genes by ChIP-sequencing analysis. If the KAT responsible for CFP1 acetylation is p300, it would be useful to determine whether activation of p300 by ERK1/2, is also essential for CFP1 recruitment to target genes and vice versa.

Other Set1A/B complex subunits as putative ERK1/2 substrates

We were interested in other subunits of the Set1A/B complexes following our identification of CFP1 as an ERK1/2 substrate. The Set1A/B complexes are distinguished from the MLL family of H3K4 trimethylases by unique enzymatic and accessory subunits, including CFP1 and Wdr82 (Figure 2-2). Presumably, unique subunit association contributes to targeting of the Set1A/B complexes to genomic loci distinct from the related MLL complexes. Targeting of Set1A and Set1B appears to be mutually exclusive, suggesting further regulatory differences between the two complexes (245). Recently, two new subunits unique to the Set1B complex were reported, Bod1 and Bod1L (128), which have been implicated in mitotic progression and may point to specific roles for the Set1B complex in this process (260).

Human Set1A and Set1B are large proteins at 1707 and 1966 amino acids respectively, and despite sharing a highly-conserved C-terminal methyltransferase domain, otherwise display limited identity throughout the rest of the primary amino acid sequence. Unique regulation and targeting may occur as a consequence of divergent post-translational modifications. Set1A and Set1B both represent good hypothetical ERK1/2 substrates, based on primary amino acid sequence. 23 Ser/Thr-Pro sites are present on Set1A and 38 are present on Set1B, with 5 and 8 predicted consensus sites (Pro-X-Ser/Thr-Pro) respectively.

Our preliminary *in vitro* studies suggest that Set1B and ERK2 bind directly, and that ERK2 is able to robustly phosphorylate Set1B. Notably, though perhaps not unsurprisingly, ERK2 phosphorylates a C-terminal Set1B fragment encompassing approximately 800 residues of the mouse protein to a greater extent than any other potential ERK1/2 substrate that we have

studied to date, with a molar incorporation of 4 phosphates per mole Set1B. Coupled with phosphorylation of CFP1 and potentially other complex subunits by ERK1/2, this could serve to fine-tune targeting or activity of methyltransferase activity towards substrate chromatin. It may also contribute to phenotypes observed in our studies on the transcriptional effects of CFP1 depletion.

CFP1 interactions with DNMT1

The role that CFP1 phosphorylation might play in directing DNMT1 function, particularly in the context of Ras-activated cancers, appealed to us for further study. In one model, mutation of KRas leading to sustained ERK1/2 pathway activation specifically silences a subset of genes, including the pro-apoptotic Fas gene (21,198). Silencing is the result of heterochromatin formation, and an siRNA screen was employed to define signaling and chromatin-modifying elements required for maintenance of the silenced state. This screen identified the maintenance DNA methyltransferase DNMT1, as well as ERK2, as requisite components of the signaling cascade that enforces silencing.

This was particularly intriguing, since CFP1 and DNMT1 interact both physically and functionally. We hypothesized that CFP1 might be a link between DNMT1 targeting and ERK2 signaling in this system. The initial screen was performed in KRas-mutated NIH3T3 cells (21), but tests in a human cell line carrying an activating mutation in HRas suggested that the effect was a general one. However, efforts in our own lab with several human Ras-activated cancer lines as well as in the NIH3T3 line employed in the initial screen failed to replicate changes in Fas protein or mRNA expression with knockdown or pharmacological inhibition of either DNMT1 or ERK2.

CFP1 phosphorylation and binding partner interactions

Although CFP1 exists in complex with Set1A and Set1B at stoichiometric concentrations (128), the extent of interaction between CFP1 and DNMT1 remains unclear. Attempts to co-immunoprecipitate other Set1A/B complex components with DNMT1 (128,209) have failed, and mutation of residues on CFP1 that ablate interaction with Set1 have no impact on DNMT1 binding (23) suggesting that CFP1 exists in separate complexes. Interaction studies using

truncation mutants of CFP1 and DNMT1 to determine minimal interacting domains suggest that interactions are complex and rely on multiple points of contact (23).

We hypothesized that CFP1 phosphorylation by ERK1/2 might dictate changes in interaction with its various binding partners, and perhaps mediate switching of CFP1 from one complex to another. In such a paradigm, stable activation of ERK1/2 might promote binding of CFP1 to DNMT1 instead of the Set1A/B complexes, resulting in mislocalization of this complex to genes containing unmethylated CpG islands that were otherwise actively transcribed and instead, silencing them. To date, neither of the CFP1 phosphorylation point mutants that we have tested have detectably impacted interaction with Set1B, but we have yet to test interactions with DNMT1, or with other point mutants. Figure 5-1 outlines experimentally-observed phosphorylation sites on CFP1 and where they lie in relation to regions responsible for binding with DNMT1 and Set1A/B. We are in the process of generating point and deletion mutants to test in binding and functional assays. In addition to interaction studies with point mutants, we are also interested in assaying pairwise interactions between CFP1 and its binding partners under conditions of ERK1/2 pathway activation and inhibition.
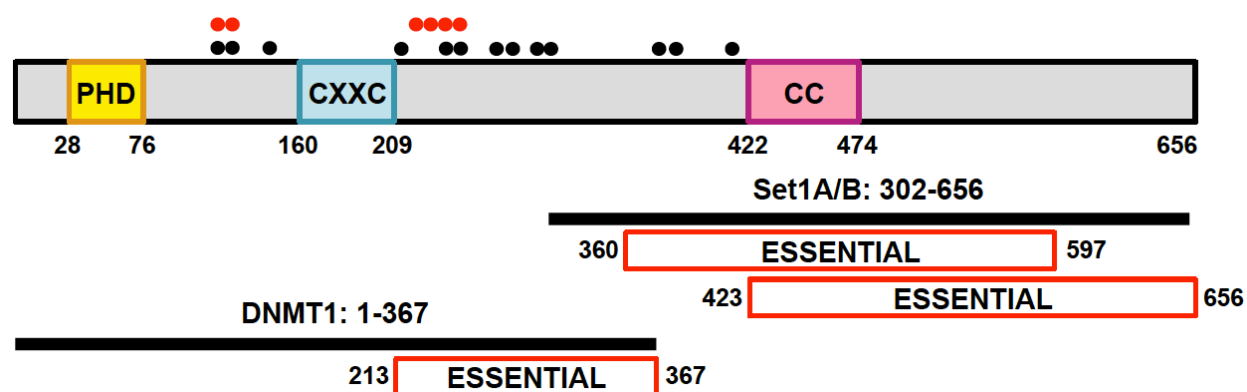
Figure 5-1: Observed phosphorylation sites on CFP1 by mass spectrometry with pERK2 (black circles) and from Phosphosite.org (red circles). Full site numbering, including sites that were predicted to be ERK2 targets and were not experimentally observed, is available in Table 3-2.

CFP1 and ERK1/2 interaction in development

Proliferation defects were consistently observed in CFP1 depleted systems (208,250), attributed to an increased rate of apoptosis. CFP1 knockout mice die early in embryogenesis, typically between embryonic days 4.5-6.5. Embryonic lethality associated with ERK2 knockout occurs at a somewhat later stage of development, around embryonic day 8.5 (38). CFP1 phosphorylation by ERK1/2 might exert developmentally-restricted phenotypic effects. In this case, loss of ERK1/2-directed CFP1 function might contribute to early lethality in these systems.

To address this, CFP1 constructs bearing point mutations at putative ERK1/2 phosphorylation sites would be introduced into CFP1-null mESCs. The competency of these cells to develop and proliferate similar to rescue with a wild type construct would be assessed and compared. DNA methylation would also be assayed by methyl acceptance assay. Based on previous work, we could also determine whether phosphorylation point mutant constructs could rescue aberrant upregulation of H3K4 trimethylation observed during induced differentiation by immunoblotting, as well as ectopic expression of this marking observed by ChIP-sequencing studies.

In closing, identification of CFP1 as an ERK1/2-associated epigenetic regulator and substrate was of great interest for further study. Although the complexity of CFP1 modification has limited our ability to fully dissect the functional consequences of phosphorylation, our work to date has served to direct future study on this topic. We anticipate that CFP1 phosphorylation by ERK1/2 may have broad impact on physiological and pathological outcomes of histone and DNA methylation, and look forward to performing a more detailed investigation into this subject.

APPENDIX A
Full Mass Spectrometry Results for Phosphorylated CFP1 1-481 by Trypsin Digestion

This report summarizes post-translational modifications in K7EQ21, that were assigned amongst peptide to spectrum matches (PSMs) with a q-value <= 0.01. Each site listed was the 'best' localization in at least one PSM considered by the ModLS localization scoring algorithm. For each site, details of the supporting PSM with the highest PTMScore (highest quality match) are given. Contact the Proteomics Core if you have queries about reading or understanding these results. Due to the limitations of site assignment from MS/MS data, conflicting PTM site assignments may be given for identical or overlapping peptides. These conflicts require knowledge of correct data interpretation practices to resolve.

K7EQ21 - K7EQ21_HUMAN CXXC finger 1 (PHD domain), isoform CRA_c OS=Homo sapiens GN=CXXC1 PE=2 SV=1

Percent Coverage: 63.0%

```
MEGIGSDPEP PLAGEISKSE NGENREIICI CRKPDINCEM IGCDNCHEWT SGDCIRITEK  80
MARAIREWYC RECRKKDPKL
KIRYRNKKSR KRKGNREDSS KRPKEGGGRK RFVEPEICRE RAGSGTGVGA MLARGSASEN 160
RSSPOPLVRT RSCHROOOOO
CIRRSARMCG RCKRCPRTED CGHCRRCRIM KKRGGRNKIR QKCRLRQCCL RARESYKYFP 240
SSISPVTPSE SIRRRRRPLP
TCCCRQPSQK LGRIRRIEGA VASSTVNRRP RRTRTRRFIS DRDLPLDPDL YQDFCAGAFD 320
DHGLPWMSDT EESPFLDPAL
RKPRYKYKHV RRRRNKSEKK KEERYKRHRQ KGNHKENNNR RRRRDRKLRR SLRQCLRRGC 400
RRIRRQRSSKY CRRRCGRRLR
RRRIRRILRQ RICCNCCGRC IRRRNGRKLL ERIRRRCCSA RTRLCRMRRR IRRLRRIILR 480
RRCQRRVRKDE ESNEGDSDDT
DLQIFCVSCG HPINPRVALR HMERCYAKYE SQTSFGSMYP TRIEGAKHTV SGSRCCAPST 560
HGTPKCQLTR YAGAPLYVMS
LSSRVTSAAC PSASAIAITA GRSCGVRKWT WSACVCGTSW TSCLSRSAMC AQP
```

PTM Summary – Top sites and best supporting peptide spectrum matches

Found in Group: All

Site	Modification	Best Supporting Peptide Spectrum Match		
		Peptide Spectrum Match	Scores	Peptide Localizations
S99	Phospho (ST)	DGNERDSSEPRDEGGGR 1 x Phospho (ST) QEX1_S03161.8827.8827.2 (HCD)	ModLS: 42 PTMScore: 431 ID Prob: 1 q-value: 0.000	S99 1.000 Phospho (ST)
S100	Phospho (ST)	ERDGNERDSSEPRDEGGGR 1 x Phospho (ST) QEX1_S03161.8486.8486.3 (HCD)	ModLS: 1 PTMScore: 294 ID Prob: 1 q-value: 0.000	S99 0.443 Phospho (ST) S100 0.557 Phospho (ST)
S124	Phospho (ST)	AGSGTGVGAMLAR 1 x Oxidation (M) 1 x Phospho (ST) QEX1_S03161.14165.14165.2 (HCD)	ModLS: 138 PTMScore: 414 ID Prob: 1 q-value: 0.000	S124 1.000 Phospho (ST) M131 1.000 Oxidation (M)
	Phospho	AGSGTGVGAMLAR	ModLS: 79 PTMScore: 379	T126 1.000 Phospho (ST)

T126 (ST)		1 x Oxidation (M) 1 x Phospho (ST) QEX1_S03161.13879.13879.2 (HCD)	ID Prob: 1 q-value: 0.000	M131 1.000 Oxidation (M)
S136	Phospho (ST)	GSASPHKSSPQPLVATPSQHHQQQQQQIK 2 x Phospho (ST) QEX1_S03161.11327.11327.3 (HCD)	ModLS: 16 PTMScore: 407 ID Prob: 1 q-value: 0.000	S136 0.975 Phospho (ST) S138 0.025 Phospho (ST) T150 0.988 Phospho (ST) S152 0.012 Phospho (ST)
S138	Phospho (ST)	GSASPHKSSPQPLVATPSQHHQQQQQQIK 2 x Phospho (ST) QEX1_S03161.11356.11356.4 (HCD)	ModLS: 1 PTMScore: 409 ID Prob: 1 q-value: 0.000	S136 0.442 Phospho (ST) S138 0.556 Phospho (ST) S143 0.002 Phospho (ST) T150 1.000 Phospho (ST)
S142	Phospho (ST)	GSASPHKSSPQPLVATPSQHHQQQQQQIKR 2 x Phospho (ST) QEX1_S03161.10769.10769.3 (HCD)	ModLS: 30 PTMScore: 392 ID Prob: 1 q-value: 0.000	S142 0.999 Phospho (ST) T150 1.000 Phospho (ST)
S143	Phospho (ST)	GSASPHKSSPQPLVATPSQHHQQQQQQIK 1 x Phospho (ST) QEX1_S03161.10872.10872.4 (HCD)	ModLS: 1 PTMScore: 341 ID Prob: 1 q-value: 0.000	S142 0.443 Phospho (ST) S143 0.557 Phospho (ST)
T150	Phospho (ST)	SSPQPLVATPSQHHQQQQQQIK 1 x Phospho (ST) QEX1_S03161.12268.12268.2 (HCD)	ModLS: 80 PTMScore: 471 ID Prob: 1 q-value: 0.000	T150 1.000 Phospho (ST)
S152	Phospho (ST)	GSASPHKSSPQPLVATPSQHHQQQQQQIKR 3 x Phospho (ST) QEX1_S03161.11177.11177.4 (HCD)	ModLS: 0 PTMScore: 258 ID Prob: 1 q-value: 0.000	S136 0.443 Phospho (ST) S138 0.557 Phospho (ST) S142 0.500 Phospho (ST) S143 0.500 Phospho (ST) T150 0.008 Phospho (ST) S152 0.992 Phospho (ST)
T178	Phospho (ST)	MCGECEACRRTEDCGHCDFCR 6 x Carbamidomethyl (C) 1 x Phospho (ST) QEX1_S03161.11347.11347.3 (HCD)	ModLS: Unambiguous PTMScore: 245 ID Prob: 1 q-value: 0.000	C169 1.000 Carbamidomethyl (C) C172 1.000 Carbamidomethyl (C) C175 1.000 Carbamidomethyl (C) T178 1.000 Phospho (ST) C181 1.000 Carbamidomethyl (C) C184 1.000 Carbamidomethyl (C) C187 1.000 Carbamidomethyl (C)
S215	Phospho (ST)	ESYKYFPSSLSPVTPSESLPRPR 2 x Phospho (ST) QEX1_S03161.19237.19237.3 (HCD)	ModLS: 21 PTMScore: 352 ID Prob: 1 q-value: 0.000	S215 0.988 Phospho (ST) S221 0.005 Phospho (ST) S222 0.007 Phospho (ST) T227 0.016 Phospho (ST) S229 0.984 Phospho (ST)
Y218 Phospho (Y)		YFPSSLSPVTPSESLPRPR 1 x Phospho (Y) QEX1_S03161.17714.17714.3 (HCD)	ModLS: 40 PTMScore: 210 ID Prob: 1 q-value: 0.000	Y218 1.000 Phospho (Y)
S221	Phospho (ST)	YFPSSLSPVTPSESLPRPR 2 x Phospho (ST) QEX1_S03161.21862.21862.2 (HCD)	ModLS: 8 PTMScore: 244 ID Prob: 1 q-value: 0.000	S221 0.845 Phospho (ST) S222 0.134 Phospho (ST) S224 0.021 Phospho (ST) T227 1.000 Phospho (ST)
S222	Phospho (ST)	YFPSSLSPVTPSESLPRPR 2 x Phospho (ST)	ModLS: 1 PTMScore: 363 ID Prob: 1	S221 0.443 Phospho (ST) S222 0.557 Phospho (ST)

		QEX1_S03161.20015.20015.2 (HCD)	q-value: 0.000	T227 1.000 Phospho (ST)

S224	Phospho (ST)	YFPSSLSPVTPSESLPRPR 1 x Phospho (ST) QEX1_S03161.19094.19094.2 (HCD)	ModLS: 53 PTMScore: 332 ID Prob: 1 q-value: 0.000	S224 1.000 Phospho (ST)
T227	Phospho (ST)	YFPSSLSPVTPSESLPRPR 1 x Phospho (ST) QEX1_S03161.18841.18841.2 (HCD)	ModLS: 79 PTMScore: 492 ID Prob: 1 q-value: 0.000	T227 1.000 Phospho (ST)
S229	Phospho (ST)	ESYKYFPSSLSPVTPSESLPRPR 2 x Phospho (ST) QEX1_S03161.19237.19237.3 (HCD)	ModLS: 21 PTMScore: 352 ID Prob: 1 q-value: 0.000	S215 0.988 Phospho (ST) S221 0.005 Phospho (ST) S222 0.007 Phospho (ST) T227 0.016 Phospho (ST) S229 0.984 Phospho (ST)
S231	Phospho (ST)	YFPSSLSPVTPSESLPRPR 1 x Phospho (ST) QEX1_S03161.19527.19527.3 (HCD)	ModLS: 34 PTMScore: 248 ID Prob: 1 q-value: 0.000	S231 1.000 Phospho (ST)
T241	Phospho (ST)	RPLPTQQQPQPSQK 1 x Phospho (ST) QEX1_S03161.11247.11247.2 (HCD)	ModLS: 165 PTMScore: 315 ID Prob: 1 q-value: 0.000	T241 1.000 Phospho (ST)
S248	Phospho (ST)	RPLPTQQQPQPSQK 1 x Phospho (ST) QEX1_S03161.11053.11053.2 (HCD)	ModLS: 148 PTMScore: 349 ID Prob: 1 q-value: 0.000	S248 1.000 Phospho (ST)
S263	Phospho (ST)	IREDEGAVASSTVK 1 x Phospho (ST) QEX1_S03161.11798.11798.2 (HCD)	ModLS: 22 PTMScore: 395 ID Prob: 1 q-value: 0.000	S263 0.994 Phospho (ST) S264 0.006 Phospho (ST)
S264	Phospho (ST)	IREDEGAVASSTVK 1 x Phospho (ST) QEX1_S03161.11504.11504.2 (HCD)	ModLS: 118 PTMScore: 298 ID Prob: 1 q-value: 0.000	S264 1.000 Phospho (ST)
T275	Phospho (ST)	IREDEGAVASSTVKEPPEATATPEPLSDEDLPLD 1 x Phospho (ST) QEX1_S03161.18780.18780.3 (HCD)	ModLS: 75 PTMScore: 324 ID Prob: 0.2899 q-value: 0.013	T275 1.000 Phospho (ST)
S371	Phospho (ST)	DPASLPQCLGPGCVRPAQPSSKYCSDDCGMK 4 x Carbamidomethyl (C) 1 x Oxidation (M) 1 x Phospho (ST) QEX1_S03161.16181.16181.3 (HCD)	ModLS: 5 PTMScore: 157 ID Prob: 1 q-value: 0.000	S371 0.513 Phospho (ST) C375 1.000 Carbamidomethyl (C) C380 1.000 Carbamidomethyl (C) S387 0.162 Phospho (ST) S388 0.162 Phospho (ST) Y390 0.162 Phospho (Y) C391 1.000 Carbamidomethyl (C) C395 1.000 Carbamidomethyl (C) M397 1.000 Oxidation (M)
S387	Phospho (ST)	ADAKDPASLPQCLGPGCVRPAQPSSK 2 x Carbamidomethyl (C) 1 x Phospho (ST) QEX1_S03161.15695.15695.3 (HCD)	ModLS: 16 PTMScore: 364 ID Prob: 1 q-value: 0.000	C375 1.000 Carbamidomethyl (C) C380 1.000 Carbamidomethyl (C) S387 0.975 Phospho (ST) S388 0.025 Phospho (ST)

S388	Phospho (ST)	ADAKDPASLPQCLGPGCVRPAQPSSK 2 x Carbamidomethyl (C) 1 x Phospho (ST) QEX1_S03161.15871.15871.3 (HCD)	ModLS: 1 PTMScore: 392 ID Prob: 1 q-value: 0.000	C375 1.000 Carbamidomethyl (C) C380 1.000 Carbamidomethyl (C) S387 0.443 Phospho (ST) S388 0.557 Phospho (ST)
Y390 Phospho (Y)		YCSDDCGMK 1 x Phospho (Y) 2 x Carbamidomethyl (C) 1 x Oxidation (M) QEX1_S03161.8852.8852.2 (HCD)	ModLS: 91 PTMScore: 309 ID Prob: 1 q-value: 0.000	Y390 1.000 Phospho (Y) C391 1.000 Carbamidomethyl (C) C395 1.000 Carbamidomethyl (C) M397 1.000 Oxidation (M)
S392	Phospho (ST)	ADAKDPASLPQCLGPGCVRPAQPSSKYCSDDCGMK 4 x Carbamidomethyl (C) 1 x Phospho (ST) QEX1_S03161.15489.15489.4 (HCD)	ModLS: 1 PTMScore: 131 ID Prob: 1 q-value: 0.000	S371 0.008 Phospho (ST) C375 1.000 Carbamidomethyl (C) C380 1.000 Carbamidomethyl (C) S387 0.304 Phospho (ST) S388 0.304 Phospho (ST) C391 1.000 Carbamidomethyl (C) S392 0.383 Phospho (ST) C395 1.000 Carbamidomethyl (C)
S418	Phospho (ST)	IQQWQQSPCIAEEHGK 1 x Carbamidomethyl (C) 1 x Phospho (ST) QEX1_S03161.14342.14342.3 (HCD)	ModLS: Unambiguous PTMScore: 314 ID Prob: 1 q-value: 0.000	S418 1.000 Phospho (ST) C420 1.000 Carbamidomethyl (C)

APPENDIX B
Full Mass Spectrometry Results for Phosphorylated
CFP1 1-481 by Elastase Digestion

This report summarizes post-translational modifications in K7EQ21, that were assigned amongst peptide to spectrum matches (PSMs) with a q-value <= 0.01. Each site listed was the 'best' localization in at least one PSM considered by the ModLS localization scoring algorithm. For each site, details of the supporting PSM with the highest PTMScore (highest quality match) are given. Contact the Proteomics Core if you have queries about reading or understanding these results. Due to the limitations of site assignment from MS/MS data, conflicting PTM site assignments may be given for identical or overlapping peptides. These conflicts require knowledge of correct data interpretation practices to resolve.

K7EQ21 - K7EQ21_HUMAN CXXC finger 1 (PHD domain), isoform CRA_c OS=Homo sapiens GN=CXXC1 PE=2 SV=1

Percent Coverage: 77.7%

```
MEGDGSIFEP PDAGEDSKSE NGENAFIYCI CRKPDINCFM IGCDNCNEWF HGDCIRITEK MAKAIREWYC RECREKDPKL 80
EIRYRHKKSR ERDGNERDSS EPRDEGGGRK RPVPDFILCR RAGSGTGVGA MLARGSASPH KSSPCFIVAT PSQHHCCCCQ 160
CIKRSARMCG ECEACRRTED CGHCDFCRDM KKFGGPNKIR QKCRLRCCQL RARESYKYFP SSLSFVTPSE SLPRFRRPLP 240
TGQQPQPSCK LGRIRELEGA VASSTVKEPP EATATEEPLS DEDLELLPDL YQDFCAGAFD DHGLPWMSDT EESEELDEAL 320
RKRAVKVKHV KRREKKSEKK KEERYKRHRQ KQKHKIKWKH PERADAKDPA SLPQCLGPGC VRPACFSSKY CSDICGMKLA 400
ANRIYEILEQ RIQQWCCSPC IAEEHGKKLL ERIRRECQSA RTRLCEMERR FHELEAIILR AKQCAVREDE ESNEGISDDT 480
DLQIFCVSCG HPINPRVALR HMERCYAKYE SQTSFGSMYP TRIEGAKHTV SGSRCCAPST HGTPKCQLTR YAGAPLYVMS 560
LSSRVTSAAC PSASAIAITA GRSCGVRKWT WSACVCGTSW TSCLSRSAMC AQP
```

PTM Summary - Top sites and best supporting peptide spectrum matches

Found in Group: All

Site	Modification	Best Supporting Peptide Spectrum Match		
		Peptide Spectrum Match	Scores	Peptide Localizations
T126	Phospho (ST)	RAGSG**T**GVGA 1 x Phospho (ST) QEX1_S03162.9124.9124.2 (HCD)	ModLS: 25 PTMScore: 157 ID Prob: 0.2313 q-value: 0.028	S124 0.003 Phospho (ST) T126 0.997 Phospho (ST)
S136	Phospho (ST)	RG**S**ASPHK**S**SPQPL 2 x Phospho (ST) QEX1_S03162.10840.10840.2 (HCD)	ModLS: 1 PTMScore: 243 ID Prob: 0.3942 q-value: 0.012	S136 1.000 Phospho (ST) S142 0.557 Phospho (ST) S143 0.443 Phospho (ST)
S138	Phospho (ST)	RGSA**S**PHKSSPQPL 1 x Phospho (ST) QEX1_S03162.10024.10024.2 (HCD)	ModLS: 22 PTMScore: 455 ID Prob: 1 q-value: 0.000	S136 0.006 Phospho (ST) S138 0.994 Phospho (ST)
	Phospho	RG**S**ASPHK**S**SPQPL	ModLS: 1 PTMScore: 243	S136 1.000 Phospho (ST)

S142 (ST)		2 x Phospho (ST) QEX1_S03162.10840.10840.2 (HCD)	ID Prob: 0.3942 q-value: 0.012	S142 0.557 Phospho (ST) S143 0.443 Phospho (ST)
S143	Phospho (ST)	RGSASPHKSSPQPL 2 x Phospho (ST) QEX1_S03162.10670.10670.2 (HCD)	ModLS: 19 PTMScore: 326 ID Prob: 0.7966 q-value: 0.012	S138 1.000 Phospho (ST) S142 0.012 Phospho (ST) S143 0.988 Phospho (ST)
T150	Phospho (ST)	RGSASPHKSSPQPLVATPS 2 x Phospho (ST) QEX1_S03162.12483.12483.3 (HCD)	ModLS: 22 PTMScore: 415 ID Prob: 0.9248 q-value: 0.012	S136 0.006 Phospho (ST) S138 0.994 Phospho (ST) T150 1.000 Phospho (ST)
S165	Phospho (ST)	QHHQQQQQQJKRSAR 1 x Phospho (ST) QEX1_S03162.7718.7718.3 (HCD)	ModLS: Unambiguous PTMScore: 95 ID Prob: 0.0662 q-value: 0.045	S165 1.000 Phospho (ST)
S215	Phospho (ST)	RESYKYFPS 1 x Phospho (ST) QEX1_S03162.14812.14812.2 (HCD)	ModLS: 156 PTMScore: 176 ID Prob: 0.6395 q-value: 0.013	S215 1.000 Phospho (ST)
Y218	Phospho (Y)	YFPSSLSPVTPSESLPRPR 1 x Phospho (Y) 1 x Phospho (ST) QEX1_S03162.19178.19178.3 (HCD)	ModLS: 1 PTMScore: 108 ID Prob: 0.887 q-value: 0.012	Y218 0.327 Phospho (Y) S221 0.206 Phospho (ST) S222 0.206 Phospho (ST) S224 0.260 Phospho (ST) T227 1.000 Phospho (ST)
S222	Phospho (ST)	SLSPVTPSESLPRPR 2 x Phospho (ST) QEX1_S03162.17039.17039.2 (HCD)	ModLS: 3 PTMScore: 288 ID Prob: 0.7905 q-value: 0.012	S222 0.923 Phospho (ST) S224 0.077 Phospho (ST) T227 0.692 Phospho (ST) S229 0.308 Phospho (ST)
S224	Phospho (ST)	SLSPVTPSESLPRPR 1 x Phospho (ST) QEX1_S03162.15876.15876.2 (HCD)	ModLS: 29 PTMScore: 419 ID Prob: 0.789 q-value: 0.012	S224 0.999 Phospho (ST) T227 0.001 Phospho (ST)
T227	Phospho (ST)	SPVTPSESLPRPR 1 x Phospho (ST) QEX1_S03162.14361.14361.2 (HCD)	ModLS: 71 PTMScore: 389 ID Prob: 0.8274 q-value: 0.012	T227 1.000 Phospho (ST)
S229	Phospho (ST)	SLSPVTPSESLPRPR 2 x Phospho (ST) QEX1_S03162.16676.16676.2 (HCD)	ModLS: 32 PTMScore: 315 ID Prob: 0.7226 q-value: 0.012	S224 0.999 Phospho (ST) S229 1.000 Phospho (ST)
S231	Phospho (ST)	SLSPVTPSESLPR 1 x Phospho (ST) QEX1_S03162.16983.16983.2 (HCD)	ModLS: 35 PTMScore: 273 ID Prob: 1 q-value: 0.000	S231 0.999 Phospho (ST)
T241	Phospho (ST)	RPLFTQQQPQPS 1 x Phospho (ST) QEX1_S03162.12777.12777.2 (HCD)	ModLS: 312 PTMScore: 331 ID Prob: 0.631 q-value: 0.012	T241 1.000 Phospho (ST)
S248	Phospho (ST)	QQQPQPSQKLGR 1 x Phospho (ST) QEX1_S03162.10513.10513.2 (HCD)	ModLS: Unambiguous PTMScore: 262 ID Prob: 0.8853 q-value: 0.012	S248 1.000 Phospho (ST)
S263	Phospho (ST)	IREDEGAVASSTVKEPPEAT 1 x Phospho (ST)	ModLS: 22 PTMScore: 435 ID Prob: 1	S263 0.988 Phospho (ST) S264 0.006 Phospho (ST)

		QEX1_S03162.13163.13163.2 (HCD)	q-value: 0.000	T265 0.006 Phospho (ST)
S264	Phospho (ST)	IREDEGAVASSTVKEPPEAT 1 x Phospho (ST) QEX1_S03162.13217.13217.2 (HCD)	ModLS: 0 PTMScore: 385 ID Prob: 1 q-value: 0.000	S263 0.333 Phospho (ST) S264 0.333 Phospho (ST) T265 0.333 Phospho (ST)
T265	Phospho (ST)	SSTVKEPPEATATPEPL 2 x Phospho (ST) QEX1_S03162.17218.17218.2 (HCD)	ModLS: 1 PTMScore: 183 ID Prob: 0.737 q-value: 0.012	S263 0.307 Phospho (ST) S264 0.307 Phospho (ST) T265 0.386 Phospho (ST) T273 0.008 Phospho (ST) T275 0.992 Phospho (ST)
T273	Phospho (ST)	SSTVKEPPEATATPEPL 1 x Phospho (ST) QEX1_S03162.16072.16072.2 (HCD)	ModLS: 91 PTMScore: 356 ID Prob: 1 q-value: 0.000	T273 1.000 Phospho (ST)
T275	Phospho (ST)	SSTVKEPPEATATPEPL 1 x Phospho (ST) QEX1_S03162.15802.15802.2 (HCD)	ModLS: 48 PTMScore: 448 ID Prob: 1 q-value: 0.000	T275 1.000 Phospho (ST)
S280	Phospho (ST)	SSTVKEPPEATATPEPLSDEDLPLD 1 x Phospho (ST) QEX1_S03162.19057.19057.2 (HCD)	ModLS: 21 PTMScore: 138 ID Prob: 0.6065 q-value: 0.015	T275 0.008 Phospho (ST) S280 0.992 Phospho (ST)
T310	Phospho (ST)	AFDDHGLPWMSDTEESPFLDPALRKR 1 x Oxidation (M) 1 x Phospho (ST) QEX1_S03162.19292.19292.3 (HCD)	ModLS: 0 PTMScore: 110 ID Prob: 1 q-value: 0.000	M307 1.000 Oxidation (M) S308 0.333 Phospho (ST) T310 0.333 Phospho (ST) S313 0.333 Phospho (ST)
S313	Phospho (ST)	AFDDHGLPWMSDTEESPFLDPALRKR 1 x Oxidation (M) 1 x Phospho (ST) QEX1_S03162.19243.19243.4 (HCD)	ModLS: 20 PTMScore: 238 ID Prob: 1 q-value: 0.000	M307 1.000 Oxidation (M) T310 0.010 Phospho (ST) S313 0.990 Phospho (ST)
S371	Phospho (ST)	DAKDPASLPQCL 1 x Carbamidomethyl (C) 1 x Phospho (ST) QEX1_S03162.18214.18214.2 (HCD)	ModLS: Unambiguous PTMScore: 114 ID Prob: 0.1033 q-value: 0.035	S371 1.000 Phospho (ST) C375 1.000 Carbamidomethyl (C)
S387	Phospho (ST)	GPGCVRPAQPSSKYCSDDCGMKL 3 x Carbamidomethyl (C) 1 x Oxidation (M) 1 x Phospho (ST) QEX1_S03162.12802.12802.3 (HCD)	ModLS: 36 PTMScore: 471 ID Prob: 0.9941 q-value: 0.012	C380 1.000 Carbamidomethyl (C) S387 1.000 Phospho (ST) C391 1.000 Carbamidomethyl (C) C395 1.000 Carbamidomethyl (C) M397 1.000 Oxidation (M)
S388	Phospho (ST)	GPGCVRPAQPSSKYCSDDCGMKL 3 x Carbamidomethyl (C) 1 x Phospho (ST) QEX1_S03162.14310.14310.3 (HCD)	ModLS: 17 PTMScore: 355 ID Prob: 0.9486 q-value: 0.012	C380 1.000 Carbamidomethyl (C) S387 0.020 Phospho (ST) S388 0.980 Phospho (ST) C391 1.000 Carbamidomethyl (C) C395 1.000 Carbamidomethyl (C)
Y390	Phospho (Y)	QPSSKYCSDDCGMKL 1 x Phospho (Y) 2 x Carbamidomethyl (C) 1 x Oxidation (M) QEX1_S03162.12288.12288.3 (HCD)	ModLS: 16 PTMScore: 213 ID Prob: 0.8209 q-value: 0.012	S387 0.024 Phospho (ST) S388 0.019 Phospho (ST) Y390 0.957 Phospho (Y) C391 1.000 Carbamidomethyl (C) C395 1.000 Carbamidomethyl

				(C)
				M397 1.000 Oxidation (M)
S418	Phospho (ST)	QQWQQSPCIAEEHGKKLLERI 1 x Carbamidomethyl (C) 1 x Gln->pyro-Glu (N-term Q) 1 x Phospho (ST) QEX1_S03162.18406.18406.3 (HCD)	ModLS: Unambiguous PTMScore: 293 ID Prob: 0.9701 q-value: 0.012	Q413 1.000 Gln->pyro-Glu (N-term Q) S418 1.000 Phospho (ST) C420 1.000 Carbamidomethyl (C)

CPFP version 2.1.0 @ CPFP UTSW Live (prcvlx-cpfp-01.swmed.edu)

Page requested 13:26:28 22-May-2014 by prcvlx-web-01.swmed.edu

CPFP is built on the ISB Trans-Proteomic Pipeline, GPM X!Tandem, and NCBI OMSSA.

APPENDIX C
Microarray Results
mRNA-Encoding Genes Most Changed by FBS Treatment, Dependent on ERK1/2

				Fold Change Relationships			
Gene Name	Description	Public Gene IDs	siCtl FBS/ siCtl Basal	siCtl PD03/ siCtl Basal	siCFP1 FBS/ siCFP1 Basal	siCFP1 PD03/ siCFP1 Basal	siCFP1 Basal/ siFBS Basal
EGR1	early growth response 1	NM_001964	23.19	1.05	7.09	1.01	3.21
CYR61	cysteine-rich, angiogenic inducer, 61	NM_001554	6.65	6.6	5.63	5.41	1.04
FOSB	FBJ murine osteosarcoma viral oncogen	NM_001114171	5.18	2.2	2.62	1.27	1.79
FOS	FBJ murine osteosarcoma viral oncogen	NM_005252	5.05	3.72	3.16	1.85	1.51
BTG2	BTG family, member 2	NM_006763	4.7	2.1	2.52	1.19	1.69
ZFP36	ZFP36 ring finger protein	NM_003407	4.5	2.12	2.19	1.33	1.88
CTGF	connective tissue growth factor	NM_001901	4.39	2.89	4.82	3	-1.54
DUSP1	dual specificity phosphatase 1	NM_004417	3.7	2.54	3.41	2.1	-1.03
IL6	interleukin 6 (interferon, beta 2)	NM_000600	3.44	2.36	1.84	1.36	-1.33
DUSP5	dual specificity phosphatase 5	NM_004419	3.27	1.57	3	1.73	-1.21
NR4A2	nuclear receptor subfamily 4, group A, n	NM_006186	3.18	1.92	2.17	1.43	1.35
NR4A1	nuclear receptor subfamily 4, group A, n	NM_001202233	3.08	2.19	2.64	1.89	1.12
JUNB	jun B proto-oncogene	BC009465	2.96	1.41	2.46	1.22	1.46
ATF3	activating transcription factor 3	NM_001030287	2.56	3.52	2.06	2.39	1.09
NR4A3	nuclear receptor subfamily 4, group A, n	NM_173199	2.52	1.67	2.34	1.64	-1.36
ADAMTS1	ADAM metallopeptidase with thrombos	NM_006988	2.25	2.57	2.13	2.38	1.36
DUSP2	dual specificity phosphatase 2	NM_004418	2.2	1.35	2.32	1.24	1.31
TNFAIP3	tumor necrosis factor, alpha-induced pro	NM_006290	2.2	1.97	1.63	1.53	-1.19
PTGS2	prostaglandin-endoperoxide synthase 2 (	NM_000963	2.09	1.64	1.92	1.37	-1.06
C8orf4	chromosome 8 open reading frame 4	NM_020130	2.03	1.92	1.33	1.37	-1.07
IER2	immediate early response 2	NM_004907	1.98	1.14	1.71	1.15	1.33
EDN1	endothelin 1	NM_001168319	1.9	1.39	1.81	1.43	1.08
ZC3H12A	zinc finger CCCH-type containing 12A	NM_025079	1.74	1.18	1.57	1.14	1.11
EGR2	early growth response 2	NM_000399	1.71	1.03	1.76	1.03	1.05
IER3	immediate early response 3	NM_003897	1.68	-1.14	1.79	-1.08	1.09
KRT17	keratin 17	NM_000422	1.66	1.43	1.93	1.47	-1.69
THBS1	thrombospondin 1	NM_003246	1.66	1.55	1.78	1.54	-1.71
NUAK2	NUAK family, SNF1-like kinase, 2	NM_030952	1.65	1.21	1.51	1.24	-1.1
RHOB	ras homolog family member B	NM_004040	1.62	2.87	1.42	1.94	1.5
KLF10	Kruppel-like factor 10	NM_001032282	1.57	1.08	1.54	1.11	-1.2
GPR183	G protein-coupled receptor 183	NM_004951	1.57	1.27	1.39	1.33	-1.72
CITED2	Cbp/p300-interacting transactivator, with	NM_001168388	1.56	1.79	1.19	1.39	1.3
KIR2DL5B	killer cell immunoglobulin-like receptor	NM_001018081	1.56	1.02	-1.04	1.32	1.21
ARC	activity-regulated cytoskeleton-associate	NM_015193	1.55	1.36	1.43	1.08	1.75
SOCS3	suppressor of cytokine signaling 3	NM_003955	1.55	1.2	1.28	1.07	1.21
TRIB1	tribbles homolog 1 (Drosophila)	NM_025195	1.55	1.52	1.34	1.39	1.04
EGR3	early growth response 3	NM_001199880	1.54	-1.01	1.63	1.12	1.03
TMEM88	transmembrane protein 88	NM_203411	1.54	1.46	1.09	1.31	1.19
RASD1	RAS, dexamethasone-induced 1	NM_001199989	1.53	1.28	1.13	1.15	2.84
CCIN	calicin	NM_005893	1.52	1.46	-1.54	-1.49	2.09
BHLHE40	basic helix-loop-helix family, member e	NM_003670	1.52	-1.08	1.34	-1.13	-1.34
FOSL1	FOS-like antigen 1	NM_005438	1.49	1.15	1.37	-1.07	1.14
MYADM	myeloid-associated differentiation marke	NM_001020821	1.48	1.31	1.25	1.2	-1.25
DEFB115	defensin, beta 115	NM_001037730	1.47	1.24	1.22	1.38	1.14
HES1	hairy and enhancer of split 1, (Drosophil	NM_005524	1.46	-1.01	1.26	-1.01	1.25
MAGEL2	MAGE-like 2	NM_019066	1.46	1.22	1.3	1.45	-1.03
NFKBIZ	nuclear factor of kappa light polypeptide	NM_001005474	1.46	1.66	1.14	1.24	1.03
LCE3E	late cornified envelope 3E	NM_178435	1.45	1.31	1.13	-1.09	1.35
SRF	serum response factor (c-fos serum resp	NM_003131	1.45	1.33	1.38	1.42	-1.05
FLJ44838	uncharacterized LOC644767	BC140939	1.45	1.13	1.18	1.43	-1.08

mRNA-encoding Genes Most Upregulated with CFP1 Knockdown			
Gene Name	**Description**	**Public Gene IDs**	**Fold Change siCFP1/siCtl**
EGR1	early growth response 1	NM_001964	3.21
RASD1	RAS, dexamethasone-induced 1	NM_001199989	2.84
METTL12	methyltransferase like 12; small nucleolar RNA, H/ACA box 57	NM_001043229	2.29
SPINK5	serine peptidase inhibitor, Kazal type 5	NM_001127699	2.26
ADRB2	adrenoceptor beta 2, surface	NM_000024	2.14
CCIN	calicin	NM_005893	2.09
GADD45B	growth arrest and DNA-damage-inducible, beta	NM_015675	1.98
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	NM_005461	1.91
MAFA	v-maf musculoaponeurotic fibrosarcoma oncogene homolog A (avian)	NM_201589	1.89
ZFP36	ZFP36 ring finger protein	NM_003407	1.88
CABYR	calcium binding tyrosine-(Y)-phosphorylation regulated	NM_153769	1.87
JUN	jun proto-oncogene	NM_002228	1.86
SNAI1	snail family zinc finger 1	NM_005985	1.83
FOSB	FBJ murine osteosarcoma viral oncogene homolog B	NM_001114171	1.79
FRG2	FSHD region gene 2; FSHD region gene 2-like	NM_001005217	1.78
FRG2	FSHD region gene 2; FSHD region gene 2-like	NM_001005217	1.78
SLC7A2	solute carrier family 7 (cationic amino acid transporter, y+ system), member 2	NM_001008539	1.78
MBOAT4	membrane bound O-acyltransferase domain containing 4	NM_001100916	1.78
KLLN	killin, p53-regulated DNA replication inhibitor	NM_001126049	1.78
ARC	activity-regulated cytoskeleton-associated protein	NM_015193	1.75
OR51A4	olfactory receptor, family 51, subfamily A, member 4	NM_001005329	1.71
ADRA2A	adrenoceptor alpha 2A	NM_000681	1.7
BTG2	BTG family, member 2	NM_006763	1.69
CKMT1A	creatine kinase, mitochondrial 1A; creatine kinase, mitochondrial 1B	NM_001015001	1.68
FBXO48	F-box protein 48	NM_001024680	1.68
TMEM133	transmembrane protein 133	NM_032021	1.65
ZNHIT6	zinc finger, HIT-type containing 6	NM_001170670	1.63
RGS2	regulator of G-protein signaling 2, 24kDa	NM_002923	1.62
YRDC	yrdC domain containing (E. coli)	NM_024640	1.62
MET	met proto-oncogene (hepatocyte growth factor receptor)	NM_000245	1.61
SLC27A2	solute carrier family 27 (fatty acid transporter), member 2	NM_001159629	1.61
CD70	CD70 molecule	NM_001252	1.61
DDX10	DEAD (Asp-Glu-Ala-Asp) box polypeptide 10	NM_004398	1.61
IER5L	immediate early response 5-like	NM_203434	1.61
FKBP14	FK506 binding protein 14, 22 kDa	NM_017946	1.6
OR2A2	olfactory receptor, family 2, subfamily A, member 2	NM_001005480	1.59
MUC13	mucin 13, cell surface associated	NM_033049	1.59
DOCK10	dedicator of cytokinesis 10	NM_014689	1.58
MARS2	methionyl-tRNA synthetase 2, mitochondrial	NM_138395	1.58
CD68	CD68 molecule	NM_001040059	1.57
USP17	ubiquitin specific peptidase 17	NM_001105662	1.57
CKMT1B	creatine kinase, mitochondrial 1B	NM_020990	1.57
CCDC86	coiled-coil domain containing 86	NM_024098	1.57
CPM	carboxypeptidase M	NM_001005502	1.56
TOMM20	translocase of outer mitochondrial membrane 20 homolog (yeast)	NM_014765	1.56
RFESD	Rieske (Fe-S) domain containing	NM_001131065	1.55
PPP2R1B	protein phosphatase 2, regulatory subunit A, beta	NM_001177562	1.55
EXOSC5	exosome component 5	NM_020158	1.55
DPH2	DPH2 homolog (S. cerevisiae)	NM_001039589	1.54
TAF4B	TAF4b RNA polymerase II, TATA box binding protein (TBP)-associated factor, 10	NM_005640	1.54
HOXA13	homeobox A13	NM_000522	1.53
HBA1	hemoglobin, alpha 1; hemoglobin, alpha 2	NM_000558	1.53
DIO3	deiodinase, iodothyronine, type II	NM_001362	1.53
NOP16	NOP16 nucleolar protein; NOP16 nucleolar protein homolog (yeast)	NM_016391	1.53

	mRNA-encoding Genes Most Upregulated with CFP1 Knockdown, cont.		
Gene Name	**Description**	**Public Gene IDs**	**Fold Change siCFP1/siCtl**
PNO1	partner of NOB1 homolog (S. cerevisiae)	NM_020143	1.52
NRARP	NOTCH-regulated ankyrin repeat protein	NM_001004354	1.51
ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	NM_002166	1.51
FOS	FBJ murine osteosarcoma viral oncogene homolog	NM_005252	1.51
RHOB	ras homolog family member B	NM_004040	1.5
OR10Z1	olfactory receptor, family 10, subfamily Z, member 1	NM_001004478	1.49
GEMIN5	gem (nuclear organelle) associated protein 5	NM_001252156	1.49
SAT1	spermidine/spermine N1-acetyltransferase 1	NM_002970	1.49
TNFSF9	tumor necrosis factor (ligand) superfamily, member 9	NM_003811	1.49
C12orf66	chromosome 12 open reading frame 66	NM_152440	1.49
RNF151	ring finger protein 151	NM_174903	1.49
LOC554207	uncharacterized LOC554207	BC031469	1.48
ABCC2	ATP-binding cassette, sub-family C (CFTR/MRP), member 2	NM_000392	1.48
HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)	NM_000860	1.48
ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	NM_002203	1.48
RRS1	RRS1 ribosome biogenesis regulator homolog (S. cerevisiae)	NM_015169	1.48
TCTA	T-cell leukemia translocation altered	NM_022171	1.48
HORMAD1	HORMA domain containing 1	NM_001199829	1.47
RGS16	regulator of G-protein signaling 16	NM_002928	1.47
JUNB	jun B proto-oncogene	BC009465	1.46
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	NM_002467	1.46
UBXN8	UBX domain protein 8	NM_005671	1.46
SERTAD1	SERTA domain containing 1	NM_013376	1.46
PDSS1	prenyl (decaprenyl) diphosphate synthase, subunit 1	NM_014317	1.46
DZIP3	DAZ interacting zinc finger protein 3; DAZ interacting protein 3, zinc finger	NM_014648	1.46
C11orf1	chromosome 11 open reading frame 1	NM_022761	1.46
MAK16	MAK16 homolog (S. cerevisiae)	NM_032509	1.46
TMEM145	transmembrane protein 145	NM_173633	1.46
ARRDC4	arrestin domain containing 4	NM_183376	1.46
ANKRD18B	ankyrin repeat domain 18B	NM_001244752	1.45
PI3	peptidase inhibitor 3, skin-derived	NM_002638	1.45
POLG2	polymerase (DNA directed), gamma 2, accessory subunit	NM_007215	1.45
PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1	NM_021127	1.45

mRNA-encoding Genes Most Downregulated with CFP1 Knockdown			
Gene Name	Description	Public Gene IDs	Fold Change siCFP1/siCtl
TULP3	tubby like protein 3	NM_001160408	-2.37
NCEH1	neutral cholesterol ester hydrolase 1	NM_020792	-2.12
TGM2	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase	NM_004613	-2.08
CXXC1	CXXC finger protein 1	NM_001101654	-2.07
NXT2	nuclear transport factor 2-like export factor 2	NM_001242617	-1.95
SLC9A6	solute carrier family 9, subfamily A (NHE6, cation proton antiporter 6), member 6	NM_001042537	-1.92
ANXA8L1	annexin A8-like 1; annexin A8	NM_001098845	-1.89
B4GALT1	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1	NM_001497	-1.83
LPGAT1	lysophosphatidylglycerol acyltransferase 1	NM_014873	-1.83
ANXA8L2	annexin A8-like 2	NM_001630	-1.81
MAPK6	mitogen-activated protein kinase 6	NM_002748	-1.77
LGALS1	lectin, galactoside-binding, soluble, 1	NM_002305	-1.75
SLFN5	schlafen family member 5	NM_144975	-1.74
CPA4	carboxypeptidase A4	NM_001163446	-1.72
GPR183	G protein-coupled receptor 183	NM_004951	-1.72
ANXA8L1	annexin A8-like 1; annexin A8	NM_001098845	-1.71
THBS1	thrombospondin 1; NULL	NM_003246	-1.71
FAM111B	family with sequence similarity 111, member B	NM_001142703	-1.7
KRT17	keratin 17	NM_000422	-1.69
SLC4A4	solute carrier family 4, sodium bicarbonate cotransporter, member 4	NM_003759	-1.69
PPP1R3C	protein phosphatase 1, regulatory subunit 3C	NM_005398	-1.69
ABHD4	abhydrolase domain containing 4	NM_022060	-1.69
ASS1	argininosuccinate synthase 1	NM_000050	-1.67
LAPTM4A	lysosomal protein transmembrane 4 alpha	NM_014713	-1.67
SLC44A1	solute carrier family 44, member 1	NM_080546	-1.67
LGALS9B	lectin, galactoside-binding, soluble, 9B	NM_001042685	-1.66
LGALS9C	lectin, galactoside-binding, soluble, 9C; lectin, galactoside-binding, soluble, 9B	NM_001040078	-1.65
ATP13A3	ATPase type 13A3	NM_024524	-1.65
CEP112	centrosomal protein 112kDa	NM_001037325	-1.61
ANO6	anoctamin 6	NM_001142680	-1.61
CSTF2T	cleavage stimulation factor, 3' pre-RNA, subunit 2, 64kDa, tau variant	NM_015235	-1.61
C15orf38	chromosome 15 open reading frame 38	NM_182616	-1.61
VWA9	von Willebrand factor A domain containing 9	NM_001207058	-1.59
LIF	leukemia inhibitory factor	NM_002309	-1.59
DAB2	Dab, mitogen-responsive phosphoprotein, homolog 2 (Drosophila)	NM_001244871	-1.58
ST6GALNAC2	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide	NM_006456	-1.58
LUM	lumican	NM_002345	-1.57
ITGB4	integrin, beta 4	NM_000213	-1.56
ZSCAN31	zinc finger and SCAN domain containing 31	NM_001135216	-1.56
PTPRU	protein tyrosine phosphatase, receptor type, U	NM_001195001	-1.56
CD9	CD9 molecule; uncharacterized LOC100653288; uncharacterized LOC100652804	NM_001769	-1.56
ANPEP	alanyl (membrane) aminopeptidase	NM_001150	-1.55
PIEZO2	piezo-type mechanosensitive ion channel component 2	NM_022068	-1.55
CTGF	connective tissue growth factor	NM_001901	-1.54
STRA6	stimulated by retinoic acid 6; stimulated by retinoic acid gene 6 homolog (mouse)	NM_001142617	-1.53
PSME1	proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	NM_006263	-1.53
SPA17	sperm autoantigenic protein 17	NM_017425	-1.53
RAP2A	RAP2A, member of RAS oncogene family	NM_021033	-1.53
EMC4	ER membrane protein complex subunit 4	NM_016454	-1.52
AHNAK2	AHNAK nucleoprotein 2	NM_138420	-1.52
OLFML2A	olfactomedin-like 2A	NM_182487	-1.52
IFNAR2	interferon (alpha, beta and omega) receptor 2	NM_000874	-1.51
IFIT2	interferon-induced protein with tetratricopeptide repeats 2	NM_001547	-1.51
KLRC3	killer cell lectin-like receptor subfamily C, member 3	NM_002261	-1.51
AMMECR1	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chrom	NM_001025580	-1.5
IFIT1	interferon-induced protein with tetratricopeptide repeats 1	NM_001548	-1.5

mRNA-encoding Genes Most Downregulated with CFP1 Knockdown, cont.			
Gene Name	Description	Public Gene IDs	Fold Change siCFP1/siCtl
ARL2BP	ADP-ribosylation factor-like 2 binding protein	NM_012106	-1.5
OR14A16	olfactory receptor, family 14, subfamily A, member 16	NM_001001966	-1.49
HEY1	hairy/enhancer-of-split related with YRPW motif 1	NM_012258	-1.49
FAM114A1	family with sequence similarity 114, member A1	NM_138389	-1.49
TMEM205	transmembrane protein 205	NM_198536	-1.49
UEVLD	UEV and lactate/malate dehyrogenase domains	NM_001040697	-1.48
SYNPO	synaptopodin	NM_001109974	-1.48
RAB3D	RAB3D, member RAS oncogene family	NM_004283	-1.48
SPTLC1	serine palmitoyltransferase, long chain base subunit 1	NM_006415	-1.48
DEFA5	defensin, alpha 5, Paneth cell-specific	NM_021010	-1.48
SLC2A12	solute carrier family 2 (facilitated glucose transporter), member 12	NM_145176	-1.48
IGSF3	immunoglobulin superfamily, member 3	NM_001542	-1.47
TNFAIP2	tumor necrosis factor, alpha-induced protein 2	NM_006291	-1.47
CD24	CD24 molecule	NM_013230	-1.47
PAX6	paired box 6	NM_000280	-1.46
GPD2	glycerol-3-phosphate dehydrogenase 2 (mitochondrial)	NM_000408	-1.46
MGP	matrix Gla protein	NM_000900	-1.46
RNASE13	ribonuclease, RNase A family, 13 (non-active)	NM_001012264	-1.46
KRT19	keratin 19	NM_002276	-1.46
TMEM2	transmembrane protein 2	NM_013390	-1.46
LGSN	lengsin, lens protein with glutamine synthetase domain	NM_001143940	-1.45
SH3RF1	SH3 domain containing ring finger 1	NM_02087	-1.45
SPZ1	spermatogenic leucine zipper 1	NM_032567	-1.45

BIBLIOGRAPHY

1. Raman, M., Chen, W., and Cobb, M. H. (2007) Differential regulation and properties of MAPKs. *Oncogene* **26**, 3100-3112

2. Long, H. K., Blackledge, N. P., and Klose, R. J. (2013) ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochemical Society transactions* **41**, 727-740

3. Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R., Deaton, A., Andrews, R., James, K. D., Turner, D. J., Illingworth, R., and Bird, A. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082-1086

4. Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J. H., Skalnik, D., and Bird, A. (2012) Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & development* **26**, 1714-1728

5. Wu, M., Wang, P. F., Lee, J. S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008) Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol Cell Biol* **28**, 7337-7344

6. Karlic, R., Chung, H. R., Lasserre, J., Vlahovicek, K., and Vingron, M. (2010) Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 2926-2931

7. Koch, C. M., Andrews, R. M., Flicek, P., Dillon, S. C., Karaoz, U., Clelland, G. K., Wilcox, S., Beare, D. M., Fowler, J. C., Couttet, P., James, K. D., Lefebvre, G. C., Bruce, A. W., Dovey, O. M., Ellis, P. D., Dhami, P., Langford, C. F., Weng, Z., Birney, E., Carter, N. P., Vetrie, D., and Dunham, I. (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome research* **17**, 691-707

8. Eberl, H. C., Spruijt, C. G., Kelstrup, C. D., Vermeulen, M., and Mann, M. (2013) A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. *Mol Cell* **49**, 368-378

9. Deaton, A. M., and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes & development* **25**, 1010-1022

10. Jones, P. A., and Baylin, S. B. (2002) The fundamental role of epigenetic events in cancer. *Nature reviews. Genetics* **3**, 415-428

11. Li, E., Bestor, T. H., and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926

12. Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D. J. (2012) Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* **335**, 709-712

13. Fowler, T., Sen, R., and Roy, A. L. (2011) Regulation of primary response genes. *Mol Cell* **44**, 348-360

14. Whitmarsh, A. J. (2007) Regulation of gene transcription by mitogen-activated protein kinase signaling pathways. *Biochim Biophys Acta* **1773**, 1285-1298

15. Edmunds, J. W., and Mahadevan, L. C. (2004) MAP kinases as structural adaptors and enzymatic activators in transcription complexes. *Journal of cell science* **117**, 3715-3723

16. Chen, Y. J., Wang, Y. N., and Chang, W. C. (2007) ERK2-mediated C-terminal serine phosphorylation of p300 is vital to the regulation of epidermal growth factor-induced keratin 16 gene expression. *J Biol Chem* **282**, 27215-27228

17. Galbraith, M. D., Saxton, J., Li, L., Shelton, S. J., Zhang, H., Espinosa, J. M., and Shaw, P. E. (2013) ERK phosphorylation of MED14 in promoter complexes during mitogen-induced gene activation by Elk-1. *Nucleic acids research* **41**, 10241-10253

18. Stevens, J. L., Cantin, G. T., Wang, G., Shevchenko, A., Shevchenko, A., and Berk, A. J. (2002) Transcription control by E1A and MAP kinase pathway via Sur2 mediator subunit. *Science* **296**, 755-758

19. Serra, R. W., Fang, M., Park, S. M., Hutchinson, L., and Green, M. R. (2014) A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype. *eLife* **3**, e02313

20. Dunn, K. L., Espino, P. S., Drobic, B., He, S., and Davie, J. R. (2005) The Ras-MAPK signal transduction pathway, cancer and chromatin remodeling. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **83**, 1-14

21. Gazin, C., Wajapeyee, N., Gobeil, S., Virbasius, C. M., and Green, M. R. (2007) An elaborate pathway required for Ras-mediated epigenetic silencing. *Nature* **449**, 1073-1077

22. Wajapeyee, N., Malonia, S. K., Palakurthy, R. K., and Green, M. R. (2013) Oncogenic RAS directs silencing of tumor suppressor genes through ordered recruitment of transcriptional repressors. *Genes & development* **27**, 2221-2226

23. Butler, J. S., Lee, J. H., and Skalnik, D. G. (2008) CFP1 interacts with DNMT1 independently of association with the Setd1 Histone H3K4 methyltransferase complexes. *DNA Cell Biol* **27**, 533-543

24. Carlone, D. L., Lee, J. H., Young, S. R., Dobrota, E., Butler, J. S., Ruiz, J., and Skalnik, D. G. (2005) Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. *Mol Cell Biol* **25**, 4881-4891

25. Young, S. R., Mumaw, C., Marrs, J. A., and Skalnik, D. G. (2006) Antisense targeting of CXXC finger protein 1 inhibits genomic cytosine methylation and primitive hematopoiesis in zebrafish. *J Biol Chem* **281**, 37034-37044

26. Boulton, T. G., Yancopoulos, G. D., Gregory, J. S., Slaughter, C., Moomaw, C., Hsu, J., and Cobb, M. H. (1990) An insulin-stimulated protein kinase similar to yeast kinases involved in cell cycle control. *Science* **249**, 64-67

27. Boulton, T. G., Nye, S. H., Robbins, D. J., Ip, N. Y., Radziejewska, E., Morgenbesser, S. D., DePinho, R. A., Panayotatos, N., Cobb, M. H., and Yancopoulos, G. D. (1991) ERKs: a family of protein-serine/threonine kinases that are activated and tyrosine phosphorylated in response to insulin and NGF. *Cell* **65**, 663-675

28. Sturgill, T. W., Ray, L. B., Erikson, E., and Maller, J. L. (1988) Insulin-stimulated MAP-2 kinase phosphorylates and activates ribosomal protein S6 kinase II. *Nature* **334**, 715-718

29. Ray, L. B., and Sturgill, T. W. (1987) Rapid stimulation by insulin of a serine/threonine kinase in 3T3-L1 adipocytes that phosphorylates microtubule-associated protein 2 in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 1502-1506

30. Ray, L. B., and Sturgill, T. W. (1988) Characterization of insulin-stimulated microtubule-associated protein kinase. Rapid isolation and stabilization of a novel serine/threonine kinase from 3T3-L1 cells. *J Biol Chem* **263**, 12721-12727

31. Ray, L. B., and Sturgill, T. W. (1988) Insulin-stimulated microtubule-associated protein kinase is phosphorylated on tyrosine and threonine in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 3753-3757

32. Meloche, S., and Pouyssegur, J. (2007) The ERK1/2 mitogen-activated protein kinase pathway as a master regulator of the G1- to S-phase transition. *Oncogene* **26**, 3227-3239

33. Kunath, T., Saba-El-Leil, M. K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007) FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895-2902

34. Reth, M., and Nielsen, P. (2014) Signaling circuits in early B-cell development. *Advances in immunology* **122**, 129-175

35. Huang, C., Jacobson, K., and Schaller, M. D. (2004) MAP kinases and cell migration. *Journal of cell science* **117**, 4619-4628

36. Kyriakis, J. M., and Avruch, J. (2012) Mammalian MAPK signal transduction pathways activated by stress and inflammation: a 10-year update. *Physiological reviews* **92**, 689-737

37. Lawrence, M. C., Jivan, A., Shao, C., Duan, L., Goad, D., Zaganjor, E., Osborne, J., McGlynn, K., Stippec, S., Earnest, S., Chen, W., and Cobb, M. H. (2008) The roles of MAPKs in disease. *Cell research* **18**, 436-442

38. Saba-El-Leil, M. K., Vella, F. D., Vernay, B., Voisin, L., Chen, L., Labrecque, N., Ang, S. L., and Meloche, S. (2003) An essential function of the mitogen-activated protein kinase Erk2 in mouse trophoblast development. *EMBO Rep* **4**, 964-968

39. Pages, G., Guerin, S., Grall, D., Bonino, F., Smith, A., Anjuere, F., Auberger, P., and Pouyssegur, J. (1999) Defective thymocyte maturation in p44 MAP kinase (Erk 1) knockout mice. *Science* **286**, 1374-1377

40. Bourcier, C., Jacquel, A., Hess, J., Peyrottes, I., Angel, P., Hofman, P., Auberger, P., Pouyssegur, J., and Pages, G. (2006) p44 mitogen-activated protein kinase (extracellular signal-regulated kinase 1)-dependent signaling contributes to epithelial skin carcinogenesis. *Cancer research* **66**, 2700-2707

41. Bost, F., Aouadi, M., Caron, L., Even, P., Belmonte, N., Prot, M., Dani, C., Hofman, P., Pages, G., Pouyssegur, J., Le Marchand-Brustel, Y., and Binetruy, B. (2005) The extracellular signal-regulated kinase isoform ERK1 is specifically required for in vitro and in vivo adipogenesis. *Diabetes* **54**, 402-411

42. Satoh, Y., Endo, S., Ikeda, T., Yamada, K., Ito, M., Kuroki, M., Hiramoto, T., Imamura, O., Kobayashi, Y., Watanabe, Y., Itohara, S., and Takishima, K. (2007) Extracellular signal-regulated kinase 2 (ERK2) knockdown mice show deficits in long-term memory; ERK2 has a specific function in learning and memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**, 10765-10776

43. Payne, D. M., Rossomando, A. J., Martino, P., Erickson, A. K., Her, J. H., Shabanowitz, J., Hunt, D. F., Weber, M. J., and Sturgill, T. W. (1991) Identification of the regulatory

phosphorylation sites in pp42/mitogen-activated protein kinase (MAP kinase). *EMBO J* **10**, 885-892

44. Canagarajah, B. J., Khokhlatchev, A., Cobb, M. H., and Goldsmith, E. J. (1997) Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. *Cell* **90**, 859-869

45. Sheridan, D. L., Kong, Y., Parker, S. A., Dalby, K. N., and Turk, B. E. (2008) Substrate discrimination among mitogen-activated protein kinases through distinct docking sequence motifs. *J Biol Chem* **283**, 19511-19520

46. Bain, J., Plater, L., Elliott, M., Shpiro, N., Hastie, C. J., McLauchlan, H., Klevernic, I., Arthur, J. S., Alessi, D. R., and Cohen, P. (2007) The selectivity of protein kinase inhibitors: a further update. *The Biochemical journal* **408**, 297-315

47. Karnoub, A. E., and Weinberg, R. A. (2008) Ras oncogenes: split personalities. *Nature reviews. Molecular cell biology* **9**, 517-531

48. Lito, P., Rosen, N., and Solit, D. B. (2013) Tumor adaptation and resistance to RAF inhibitors. *Nature medicine* **19**, 1401-1409

49. Osborne, J. K., Zaganjor, E., and Cobb, M. H. (2012) Signal control through Raf: in sickness and in health. *Cell research* **22**, 14-22

50. Dhomen, N., and Marais, R. (2007) New insight into BRAF mutations in cancer. *Current opinion in genetics & development* **17**, 31-39

51. Kolch, W. (2005) Coordinating ERK/MAPK signalling through scaffolds and inhibitors. *Nature reviews. Molecular cell biology* **6**, 827-837

52. Traverse, S., Gomez, N., Paterson, H., Marshall, C., and Cohen, P. (1992) Sustained activation of the mitogen-activated protein (MAP) kinase cascade may be required for differentiation of PC12 cells. Comparison of the effects of nerve growth factor and epidermal growth factor. *The Biochemical journal* **288 (Pt 2)**, 351-355

53. Gallego, C., Gupta, S. K., Heasley, L. E., Qian, N. X., and Johnson, G. L. (1992) Mitogen-activated protein kinase activation resulting from selective oncogene expression in NIH 3T3 and rat 1a cells. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 7355-7359

54. Robbins, D. J., Cheng, M., Zhen, E., Vanderbilt, C. A., Feig, L. A., and Cobb, M. H. (1992) Evidence for a Ras-dependent extracellular signal-regulated protein kinase (ERK) cascade. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 6924-6928

55. Marshall, C. J. (1995) Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80**, 179-185

56. Tan, P. B., and Kim, S. K. (1999) Signaling specificity: the RTK/RAS/MAP kinase pathway in metazoans. *Trends in genetics : TIG* **15**, 145-149

57. York, R. D., Yao, H., Dillon, T., Ellig, C. L., Eckert, S. P., McCleskey, E. W., and Stork, P. J. (1998) Rap1 mediates sustained MAP kinase activation induced by nerve growth factor. *Nature* **392**, 622-626

58. Courcelles, M., Bridon, G., Lemieux, S., and Thibault, P. (2012) Occurrence and detection of phosphopeptide isomers in large-scale phosphoproteomics experiments. *Journal of proteome research* **11**, 3753-3765

59. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648

60. Kosako, H., Yamaguchi, N., Aranami, C., Ushiyama, M., Kose, S., Imamoto, N., Taniguchi, H., Nishida, E., and Hattori, S. (2009) Phosphoproteomics reveals new ERK MAP kinase targets and links ERK to nucleoporin-mediated nuclear transport. *Nature structural & molecular biology* **16**, 1026-1035

61. Zanivan, S., Meves, A., Behrendt, K., Schoof, E. M., Neilson, L. J., Cox, J., Tang, H. R., Kalna, G., van Ree, J. H., van Deursen, J. M., Trempus, C. S., Machesky, L. M., Linding, R., Wickstrom, S. A., Fassler, R., and Mann, M. (2013) In vivo SILAC-based proteomics reveals phosphoproteome changes during mouse skin carcinogenesis. *Cell reports* **3**, 552-566

62. Carlson, S. M., and White, F. M. (2012) Labeling and identification of direct kinase substrates. *Science signaling* **5**, pl3

63. Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M. F., Piwnica-Worms, H., and Cantley, L. C. (1994) Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr Biol* **4**, 973-982

64. Northwood, I. C., Gonzalez, F. A., Wartmann, M., Raden, D. L., and Davis, R. J. (1991) Isolation and characterization of two growth factor-stimulated protein kinases that phosphorylate the epidermal growth factor receptor at threonine 669. *J Biol Chem* **266**, 15266-15276

65. Gonzalez, F. A., Raden, D. L., and Davis, R. J. (1991) Identification of substrate recognition determinants for human ERK1 and ERK2 protein kinases. *J Biol Chem* **266**, 22159-22163

66. Eblen, S. T., Kumar, N. V., Shah, K., Henderson, M. J., Watts, C. K., Shokat, K. M., and Weber, M. J. (2003) Identification of novel ERK2 substrates through use of an engineered kinase and ATP analogs. *J Biol Chem* **278**, 14926-14935

67. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12130-12135

68. Jacobs, D., Glossip, D., Xing, H., Muslin, A. J., and Kornfeld, K. (1999) Multiple docking sites on substrate proteins form a modular system that mediates recognition by ERK MAP kinase. *Genes & development* **13**, 163-175

69. Sharrocks, A. D., Yang, S. H., and Galanis, A. (2000) Docking domains and substrate-specificity determination for MAP kinases. *Trends in biochemical sciences* **25**, 448-453

70. Tanoue, T., Adachi, M., Moriguchi, T., and Nishida, E. (2000) A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nat Cell Biol* **2**, 110-116

71. Reszka, A. A., Seger, R., Diltz, C. D., Krebs, E. G., and Fischer, E. H. (1995) Association of mitogen-activated protein kinase with the microtubule cytoskeleton. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 8881-8885

72. Cheng, M., Boulton, T. G., and Cobb, M. H. (1996) ERK3 is a constitutively nuclear protein kinase. *J Biol Chem* **271**, 8951-8958

73. Ranganathan, A., Yazicioglu, M. N., and Cobb, M. H. (2006) The nuclear localization of ERK2 occurs by mechanisms both independent of and dependent on energy. *J Biol Chem* **281**, 15645-15652

74. Chuderland, D., Konson, A., and Seger, R. (2008) Identification and characterization of a general nuclear translocation signal in signaling proteins. *Mol Cell* **31**, 850-861

75. Whitehurst, A. W., Wilsbacher, J. L., You, Y., Luby-Phelps, K., Moore, M. S., and Cobb, M. H. (2002) ERK2 enters the nucleus by a carrier-independent mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7496-7501

76. Matsubayashi, Y., Fukuda, M., and Nishida, E. (2001) Evidence for existence of a nuclear pore complex-mediated, cytosol-independent pathway of nuclear translocation of ERK MAP kinase in permeabilized cells. *J Biol Chem* **276**, 41755-41760

77. Rubinfeld, H., Hanoch, T., and Seger, R. (1999) Identification of a cytoplasmic-retention sequence in ERK2. *J Biol Chem* **274**, 30349-30352

78. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260

79. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W., and Richmond, T. J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *Journal of molecular biology* **319**, 1097-1113

80. Bannister, A. J., and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell research* **21**, 381-395

81. Lorch, Y., LaPointe, J. W., and Kornberg, R. D. (1987) Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**, 203-210

82. Han, M., and Grunstein, M. (1988) Nucleosome loss activates yeast downstream promoters in vivo. *Cell* **55**, 1137-1145

83. Kornberg, R. D., and Lorch, Y. (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285-294

84. Mersfelder, E. L., and Parthun, M. R. (2006) The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. *Nucleic acids research* **34**, 2653-2662

85. Talbert, P. B., and Henikoff, S. (2010) Histone variants--ancient wrap artists of the epigenome. *Nature reviews. Molecular cell biology* **11**, 264-275

86. Grewal, S. I., and Moazed, D. (2003) Heterochromatin and epigenetic control of gene expression. *Science* **301**, 798-802

87. Sanyal, A., Lajoie, B. R., Jain, G., and Dekker, J. (2012) The long-range interaction landscape of gene promoters. *Nature* **489**, 109-113

88. Thomas, M. C., and Chiang, C. M. (2006) The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology* **41**, 105-178

89. Juven-Gershon, T., and Kadonaga, J. T. (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental biology* **339**, 225-229

90. Cluss, P. A., Epstein, L. H., Galvis, S. A., Fireman, P., and Friday, G. (1984) Effect of compliance for chronic asthmatic children. *Journal of consulting and clinical psychology* **52**, 909-910

91. Hawley, D. K., and Roeder, R. G. (1985) Separation and partial characterization of three functional steps in transcription initiation by human RNA polymerase II. *J Biol Chem* **260**, 8163-8172

92. Adelman, K., and Lis, J. T. (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews. Genetics* **13**, 720-731

93. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., and Handa, H. (1998) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & development* **12**, 343-356

94. Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D., and Ren, B. (2005) A high-resolution map of active promoters in the human genome. *Nature* **436**, 876-880

95. Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999) NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41-51

96. Peterlin, B. M., and Price, D. H. (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23**, 297-305

97. Core, L. J., and Lis, J. T. (2008) Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* **319**, 1791-1792

98. Gilchrist, D. A., Fromm, G., dos Santos, G., Pham, L. N., McDaniel, I. E., Burkholder, A., Fargo, D. C., and Adelman, K. (2012) Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. *Genes & development* **26**, 933-944

99. Hargreaves, D. C., Horng, T., and Medzhitov, R. (2009) Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138**, 129-145

100. Dahmus, M. E. (1995) Phosphorylation of the C-terminal domain of RNA polymerase II. *Biochim Biophys Acta* **1261**, 171-182

101. Lu, H., Flores, O., Weinmann, R., and Reinberg, D. (1991) The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 10004-10008

102. Jeronimo, C., Bataille, A. R., and Robert, F. (2013) The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chemical reviews* **113**, 8491-8522

103. Markowitz, R. B., Hermann, A. S., Taylor, D. F., He, L., Anthony-Cahill, S., Ahn, N. G., and Dynan, W. S. (1995) Phosphorylation of the C-terminal domain of RNA polymerase II by the extracellular-signal-regulated protein kinase ERK2. *Biochemical and biophysical research communications* **207**, 1051-1057

104. Tee, W. W., Shen, S. S., Oksuz, O., Narendra, V., and Reinberg, D. (2014) Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs. *Cell* **156**, 678-690

105. Muller, F., Zaucker, A., and Tora, L. (2010) Developmental regulation of transcription initiation: more than just changing the actors. *Current opinion in genetics & development* **20**, 533-540

106. Poss, Z. C., Ebmeier, C. C., and Taatjes, D. J. (2013) The Mediator complex and transcription regulation. *Critical reviews in biochemistry and molecular biology* **48**, 575-608

107. Saunders, A., Core, L. J., and Lis, J. T. (2006) Breaking barriers to transcription elongation. *Nature reviews. Molecular cell biology* **7**, 557-567

108. Kuehner, J. N., Pearson, E. L., and Moore, C. (2011) Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews. Molecular cell biology* **12**, 283-294

109. Lee, J. S., Smith, E., and Shilatifard, A. (2010) The language of histone crosstalk. *Cell* **142**, 682-685

110. Goke, J., Chan, Y. S., Yan, J., Vingron, M., and Ng, H. H. (2013) Genome-wide kinase-chromatin interactions reveal the regulatory network of ERK signaling in human embryonic stem cells. *Mol Cell* **50**, 844-855

111. Henikoff, S., and Shilatifard, A. (2011) Histone modification: cause or cog? *Trends in genetics : TIG* **27**, 389-396

112. Stassen, M. J., Bailey, D., Nelson, S., Chinwalla, V., and Harte, P. J. (1995) The Drosophila trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. *Mechanisms of development* **52**, 209-223

113. Jenuwein, T., Laible, G., Dorn, R., and Reuter, G. (1998) SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cellular and molecular life sciences : CMLS* **54**, 80-93

114. Herz, H. M., Mohan, M., Garruss, A. S., Liang, K., Takahashi, Y. H., Mickey, K., Voets, O., Verrijzer, C. P., and Shilatifard, A. (2012) Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & development* **26**, 2604-2620

115. Herz, H. M., Hu, D., and Shilatifard, A. (2014) Enhancer malfunction in cancer. *Mol Cell* **53**, 859-866

116. Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D. K., and Young, R. A. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517-527

117. Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-411

118. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., and Young, R. A. (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88

119. Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F., and Shilatifard, A. (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12902-12907

120. Krogan, N. J., Dover, J., Khorrami, S., Greenblatt, J. F., Schneider, J., Johnston, M., and Shilatifard, A. (2002) COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem* **277**, 10753-10755

121. Schneider, J., Wood, A., Lee, J. S., Schuster, R., Dueker, J., Maguire, C., Swanson, S. K., Florens, L., Washburn, M. P., and Shilatifard, A. (2005) Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol Cell* **19**, 849-856

122. Ng, H. H., Robert, F., Young, R. A., and Struhl, K. (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* **11**, 709-719

123. Vermeulen, M., and Timmers, H. T. (2010) Grasping trimethylation of histone H3 at lysine 4. *Epigenomics* **2**, 395-406

124. Shi, X., Kachirskaia, I., Walter, K. L., Kuo, J. H., Lake, A., Davrazou, F., Chan, S. M., Martin, D. G., Fingerman, I. M., Briggs, S. D., Howe, L., Utz, P. J., Kutateladze, T. G., Lugovskoy, A. A., Bedford, M. T., and Gozani, O. (2007) Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J Biol Chem* **282**, 2450-2455

125. Eissenberg, J. C., and Shilatifard, A. (2010) Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental biology* **339**, 240-249

126. Milne, T. A., Dou, Y., Martin, M. E., Brock, H. W., Roeder, R. G., and Hess, J. L. (2005) MLL associates specifically with a subset of transcriptionally active target genes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14765-14770

127. Hu, D., Garruss, A. S., Gao, X., Morgan, M. A., Cook, M., Smith, E. R., and Shilatifard, A. (2013) The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nature structural & molecular biology* **20**, 1093-1097

128. van Nuland, R., Smits, A. H., Pallaki, P., Jansen, P. W., Vermeulen, M., and Timmers, H. T. (2013) Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. *Mol Cell Biol* **33**, 2067-2077

129. Lee, J. H., Tate, C. M., You, J. S., and Skalnik, D. G. (2007) Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *J Biol Chem* **282**, 13419-13428

130. Guenther, M. G., Jenner, R. G., Chevalier, B., Nakamura, T., Croce, C. M., Canaani, E., and Young, R. A. (2005) Global and Hox-specific roles for the MLL1 methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8603-8608

131. Wysocka, J., Swigut, T., Milne, T. A., Dou, Y., Zhang, X., Burlingame, A. L., Roeder, R. G., Brivanlou, A. H., and Allis, C. D. (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859-872

132. Lin, J. J., Lehmann, L. W., Bonora, G., Sridharan, R., Vashisht, A. A., Tran, N., Plath, K., Wohlschlegel, J. A., and Carey, M. (2011) Mediator coordinates PIC assembly with recruitment of CHD1. *Genes & development* **25**, 2198-2209

133. Chi, P., Allis, C. D., and Wang, G. G. (2010) Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nature reviews. Cancer* **10**, 457-469

134. Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., van Schaik, F. M., Varier, R. A., Baltissen, M. P., Stunnenberg, H. G., Mann, M., and Timmers, H. T. (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58-69

135. Lauberth, S. M., Nakayama, T., Wu, X., Ferris, A. L., Tang, Z., Hughes, S. H., and Roeder, R. G. (2013) H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021-1036

136. Sanchez, R., and Zhou, M. M. (2011) The PHD finger: a versatile epigenome reader. *Trends in biochemical sciences* **36**, 364-372

137. Bienz, M. (2006) The PHD finger, a nuclear protein-interaction domain. *Trends in biochemical sciences* **31**, 35-40

138. Kusch, T. (2012) Histone H3 lysine 4 methylation revisited. *Transcription* **3**, 310-314

139. Wang, G. G., Song, J., Wang, Z., Dormann, H. L., Casadio, F., Li, H., Luo, J. L., Patel, D. J., and Allis, C. D. (2009) Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459**, 847-851

140. Voo, K. S., Carlone, D. L., Jacobsen, B. M., Flodin, A., and Skalnik, D. G. (2000) Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol Cell Biol* **20**, 2108-2121

141. Claverie, J. M. (2005) Fewer genes, more noncoding RNA. *Science* **309**, 1529-1530

142. Chodavarapu, R. K., Feng, S., Bernatavichute, Y. V., Chen, P. Y., Stroud, H., Yu, Y., Hetzel, J. A., Kuo, F., Kim, J., Cokus, S. J., Casero, D., Bernal, M., Huijser, P., Clark, A. T., Kramer, U., Merchant, S. S., Zhang, X., Jacobsen, S. E., and Pellegrini, M. (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* **466**, 388-392

143. Flaus, A., and Richmond, T. J. (1998) Positioning and stability of nucleosomes on MMTV 3'LTR sequences. *Journal of molecular biology* **275**, 427-441

144. Hughes, A. L., and Rando, O. J. (2014) Mechanisms underlying nucleosome positioning in vivo. *Annual review of biophysics* **43**, 41-63

145. Struhl, K., and Segal, E. (2013) Determinants of nucleosome positioning. *Nature structural & molecular biology* **20**, 267-273

146. Lemon, B., and Tjian, R. (2000) Orchestrated response: a symphony of transcription factors for gene control. *Genes & development* **14**, 2551-2569

147. Saxonov, S., Berg, P., and Brutlag, D. L. (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1412-1417

148. Bird, A. P. (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic acids research* **8**, 1499-1504

149. Talhaoui, I., Couve, S., Gros, L., Ishchenko, A. A., Matkarimov, B., and Saparbaev, M. K. (2014) Aberrant repair initiated by mismatch-specific thymine-DNA glycosylases

provides a mechanism for the mutational bias observed in CpG islands. *Nucleic acids research* **42**, 6300-6313

150. Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257

151. Varley, K. E., Gertz, J., Bowling, K. M., Parker, S. L., Reddy, T. E., Pauli-Behn, F., Cross, M. K., Williams, B. A., Stamatoyannopoulos, J. A., Crawford, G. E., Absher, D. M., Wold, B. J., and Myers, R. M. (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome research* **23**, 555-567

152. Edwards, C. A., and Ferguson-Smith, A. C. (2007) Mechanisms regulating imprinted genes in clusters. *Current opinion in cell biology* **19**, 281-289

153. Goll, M. G., and Bestor, T. H. (2005) Eukaryotic cytosine methyltransferases. *Annual review of biochemistry* **74**, 481-514

154. Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., and Eden, A. (2008) Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* **27**, 404-408

155. Bestor, T. H. (1992) Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J* **11**, 2611-2617

156. Kohli, R. M., and Zhang, Y. (2013) TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472-479

157. Ramirez-Carrozzi, V. R., Braas, D., Bhatt, D. M., Cheng, C. S., Hong, C., Doty, K. R., Black, J. C., Hoffmann, A., Carey, M., and Smale, S. T. (2009) A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114-128

158. Singh, H. (2009) Teeing up transcription on CpG islands. *Cell* **138**, 14-16

159. Bird, A. (2011) The dinucleotide CG as a genomic signalling module. *Journal of molecular biology* **409**, 47-53

160. Jones, P. A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews. Genetics* **13**, 484-492

161. Gille, H., Kortenjann, M., Thomae, O., Moomaw, C., Slaughter, C., Cobb, M. H., and Shaw, P. E. (1995) ERK phosphorylation potentiates Elk-1-mediated ternary complex formation and transactivation. *EMBO J* **14**, 951-962

162. Gille, H., Sharrocks, A. D., and Shaw, P. E. (1992) Phosphorylation of transcription factor p62TCF by MAP kinase stimulates ternary complex formation at c-fos promoter. *Nature* **358**, 414-417

163. Hill, C. S., Marais, R., John, S., Wynne, J., Dalton, S., and Treisman, R. (1993) Functional analysis of a growth factor-responsive transcription factor complex. *Cell* **73**, 395-406

164. Madak-Erdogan, Z., Lupien, M., Stossi, F., Brown, M., and Katzenellenbogen, B. S. (2011) Genomic collaboration of estrogen receptor alpha and extracellular signal-regulated kinase 2 in regulating gene and proliferation programs. *Mol Cell Biol* **31**, 226-236

165. Carlson, S. M., Chouinard, C. R., Labadorf, A., Lam, C. J., Schmelzle, K., Fraenkel, E., and White, F. M. (2011) Large-scale discovery of ERK2 substrates identifies ERK-mediated transcriptional regulation by ETV3. *Science signaling* **4**, rs11

166. Bardwell, L., Cook, J. G., Zhu-Shimoni, J. X., Voora, D., and Thorner, J. (1998) Differential regulation of transcription: repression by unactivated mitogen-activated protein kinase Kss1 requires the Dig1 and Dig2 proteins. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 15400-15405

167. Simone, C., Forcales, S. V., Hill, D. A., Imbalzano, A. N., Latella, L., and Puri, P. L. (2004) p38 pathway targets SWI-SNF chromatin-remodeling complex to muscle-specific loci. *Nat Genet* **36**, 738-743

168. Proft, M., and Struhl, K. (2004) MAP kinase-mediated stress relief that precedes and regulates the timing of transcriptional induction. *Cell* **118**, 351-361

169. Pokholok, D. K., Zeitlinger, J., Hannett, N. M., Reynolds, D. B., and Young, R. A. (2006) Activated signal transduction kinases frequently occupy target genes. *Science* **313**, 533-536

170. Lawrence, M. C., McGlynn, K., Shao, C., Duan, L., Naziruddin, B., Levy, M. F., and Cobb, M. H. (2008) Chromatin-bound mitogen-activated protein kinases transmit dynamic signals in transcription complexes in beta-cells. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13315-13320

171. Li, B. H., Zhang, L. L., Zhang, B. B., Yin, Y. W., Dai, L. M., Pi, Y., Guo, L., Gao, C. Y., Fang, C. Q., Wang, J. Z., and Li, J. C. (2013) Association between NADPH oxidase p22(phox) C242T polymorphism and ischemic cerebrovascular disease: a meta-analysis. *PLoS One* **8**, e56478

172. Flanagin, S., Nelson, J. D., Castner, D. G., Denisenko, O., and Bomsztyk, K. (2008) Microplate-based chromatin immunoprecipitation method, Matrix ChIP: a platform to study signaling of complex genomic events. *Nucleic acids research* **36**, e17

173. Nelson, J. D., LeBoeuf, R. C., and Bomsztyk, K. (2011) Direct recruitment of insulin receptor and ERK signaling cascade to insulin-inducible gene loci. *Diabetes* **60**, 127-137

174. Mikula, M., and Bomsztyk, K. (2011) Direct recruitment of ERK cascade components to inducible genes is regulated by heterogeneous nuclear ribonucleoprotein (hnRNP) K. *J Biol Chem* **286**, 9763-9775

175. Lawrence, M. C., Shao, C., McGlynn, K., Naziruddin, B., Levy, M. F., and Cobb, M. H. (2009) Multiple chromatin-bound protein kinases assemble factors that regulate insulin gene transcription. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 22181-22186

176. Wang, Y. N., Yamaguchi, H., Hsu, J. M., and Hung, M. C. (2010) Nuclear trafficking of the epidermal growth factor receptor family membrane proteins. *Oncogene* **29**, 3997-4006

177. Proft, M., Mas, G., de Nadal, E., Vendrell, A., Noriega, N., Struhl, K., and Posas, F. (2006) The stress-activated Hog1 kinase is a selective transcriptional elongation factor for genes responding to osmotic stress. *Mol Cell* **23**, 241-250

178. Kim, K. Y., and Levin, D. E. (2011) Mpk1 MAPK association with the Paf1 complex blocks Sen1-mediated premature transcription termination. *Cell* **144**, 745-756

179.	Lesch, B. J., and Page, D. C. (2014) Poised chromatin in the mammalian germ line. *Development* **141**, 3619-3626

180.	Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., and Lander, E. S. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326

181.	Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E. S., and Bernstein, B. E. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560

182.	Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., and Jaenisch, R. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353

183.	Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., Koseki, H., Fuchikami, T., Abe, K., Murray, H. L., Zucker, J. P., Yuan, B., Bell, G. W., Herbolsheimer, E., Hannett, N. M., Sun, K., Odom, D. T., Otte, A. P., Volkert, T. L., Bartel, D. P., Melton, D. A., Gifford, D. K., Jaenisch, R., and Young, R. A. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-313

184.	Hu, S., Xie, Z., Onishi, A., Yu, X., Jiang, L., Lin, J., Rho, H. S., Woodard, C., Wang, H., Jeong, J. S., Long, S., He, X., Wade, H., Blackshaw, S., Qian, J., and Zhu, H. (2009) Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* **139**, 610-622

185.	Rep, M., Reiser, V., Gartner, U., Thevelein, J. M., Hohmann, S., Ammerer, G., and Ruis, H. (1999) Osmotic stress-induced gene expression in Saccharomyces cerevisiae requires Msn1p and the novel nuclear factor Hot1p. *Mol Cell Biol* **19**, 5474-5485

186.	Rep, M., Krantz, M., Thevelein, J. M., and Hohmann, S. (2000) The transcriptional response of Saccharomyces cerevisiae to osmotic shock. Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes. *J Biol Chem* **275**, 8290-8300

187.	De Nadal, E., Zapater, M., Alepuz, P. M., Sumoy, L., Mas, G., and Posas, F. (2004) The MAPK Hog1 recruits Rpd3 histone deacetylase to activate osmoresponsive genes. *Nature* **427**, 370-374

188.	Mas, G., de Nadal, E., Dechant, R., Rodriguez de la Concepcion, M. L., Logie, C., Jimeno-Gonzalez, S., Chavez, S., Ammerer, G., and Posas, F. (2009) Recruitment of a chromatin remodelling complex by the Hog1 MAP kinase to stress genes. *EMBO J* **28**, 326-336

189.	Alepuz, P. M., Jovanovic, A., Reiser, V., and Ammerer, G. (2001) Stress-induced map kinase Hog1 is part of transcription activation complexes. *Mol Cell* **7**, 767-777

190.	Alepuz, P. M., de Nadal, E., Zapater, M., Ammerer, G., and Posas, F. (2003) Osmostress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II. *EMBO J* **22**, 2433-2442

191. Mouchel-Vielh, E., Rougeot, J., Decoville, M., and Peronnet, F. (2011) The MAP kinase ERK and its scaffold protein MP1 interact with the chromatin regulator Corto during Drosophila wing tissue development. *BMC developmental biology* **11**, 17

192. Gehani, S. S., Agrawal-Singh, S., Dietrich, N., Christophersen, N. S., Helin, K., and Hansen, K. (2010) Polycomb group protein displacement and gene activation through MSK-dependent H3K27me3S28 phosphorylation. *Mol Cell* **39**, 886-900

193. Rampalli, S., Li, L., Mak, E., Ge, K., Brand, M., Tapscott, S. J., and Dilworth, F. J. (2007) p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. *Nature structural & molecular biology* **14**, 1150-1156

194. Oya, H., Yokoyama, A., Yamaoka, I., Fujiki, R., Yonezawa, M., Youn, M. Y., Takada, I., Kato, S., and Kitagawa, H. (2009) Phosphorylation of Williams syndrome transcription factor by MAPK induces a switching between two distinct chromatin remodeling complexes. *J Biol Chem* **284**, 32472-32482

195. O'Donnell, A., Yang, S. H., and Sharrocks, A. D. (2008) MAP kinase-mediated c-fos regulation relies on a histone acetylation relay switch. *Mol Cell* **29**, 780-785

196. Balamotis, M. A., Pennella, M. A., Stevens, J. L., Wasylyk, B., Belmont, A. S., and Berk, A. J. (2009) Complexity in transcription control at the activation domain-mediator interface. *Science signaling* **2**, ra20

197. Martin, C., Chen, S., Heilos, D., Sauer, G., Hunt, J., Shaw, A. G., Sims, P. F., Jackson, D. A., and Lovric, J. (2010) Changed genome heterochromatinization upon prolonged activation of the Raf/ERK signaling pathway. *PLoS One* **5**, e13322

198. Peli, J., Schroter, M., Rudaz, C., Hahne, M., Meyer, C., Reichmann, E., and Tschopp, J. (1999) Oncogenic Ras inhibits Fas ligand-mediated apoptosis by downregulating the expression of Fas. *EMBO J* **18**, 1824-1831

199. Ma, Q., Alder, H., Nelson, K. K., Chatterjee, D., Gu, Y., Nakamura, T., Canaani, E., Croce, C. M., Siracusa, L. D., and Buchberg, A. M. (1993) Analysis of the murine All-1 gene reveals conserved domains with human ALL-1 and identifies a motif shared with DNA methyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 6350-6354

200. Xu, C., Bian, C., Lam, R., Dong, A., and Min, J. (2011) The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. *Nature communications* **2**, 227

201. Lee, J. H., Voo, K. S., and Skalnik, D. G. (2001) Identification and characterization of the DNA binding domain of CpG-binding protein. *J Biol Chem* **276**, 44669-44676

202. Carlone, D. L., and Skalnik, D. G. (2001) CpG binding protein is crucial for early embryonic development. *Mol Cell Biol* **21**, 7601-7606

203. Huntriss, J., Hinkins, M., Oliver, B., Harris, S. E., Beazley, J. C., Rutherford, A. J., Gosden, R. G., Lanzendorf, S. E., and Picton, H. M. (2004) Expression of mRNAs for DNA methyltransferases and methyl-CpG-binding proteins in the human female germ line, preimplantation embryos, and embryonic stem cells. *Molecular reproduction and development* **67**, 323-336

204. Monk, M., Boubelik, M., and Lehnert, S. (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371-382

205. Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & development* **12**, 2048-2060

206. Butler, J. S., Palam, L. R., Tate, C. M., Sanford, J. R., Wek, R. C., and Skalnik, D. G. (2009) DNA Methyltransferase protein synthesis is reduced in CXXC finger protein 1-deficient embryonic stem cells. *DNA Cell Biol* **28**, 223-231

207. Milhem, M., Mahmud, N., Lavelle, D., Araki, H., DeSimone, J., Saunthararajah, Y., and Hoffman, R. (2004) Modification of hematopoietic stem cell fate by 5aza 2'deoxycytidine and trichostatin A. *Blood* **103**, 4102-4110

208. Young, S. R., and Skalnik, D. G. (2007) CXXC finger protein 1 is required for normal proliferation and differentiation of the PLB-985 myeloid cell line. *DNA Cell Biol* **26**, 80-90

209. Tate, C. M., Lee, J. H., and Skalnik, D. G. (2009) CXXC finger protein 1 contains redundant functional domains that support embryonic stem cell cytosine methylation, histone methylation, and differentiation. *Mol Cell Biol* **29**, 3817-3831

210. Lee, J. H., and Skalnik, D. G. (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J Biol Chem* **280**, 41725-41731

211. Song, J., Rechkoblit, O., Bestor, T. H., and Patel, D. J. (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036-1040

212. Risner, L. E., Kuntimaddi, A., Lokken, A. A., Achille, N. J., Birch, N. W., Schoenfelt, K., Bushweller, J. H., and Zeleznik-Le, N. J. (2013) Functional specificity of CpG DNA-binding CXXC domains in mixed lineage leukemia. *J Biol Chem* **288**, 29901-29910

213. Capuano, F., Mulleder, M., Kok, R., Blom, H. J., and Ralser, M. (2014) Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. *Analytical chemistry* **86**, 3697-3702

214. Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y., Winston, F., and Allis, C. D. (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes & development* **15**, 3286-3295

215. Nislow, C., Ray, E., and Pillus, L. (1997) SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. *Mol Biol Cell* **8**, 2421-2436

216. Corda, Y., Schramke, V., Longhese, M. P., Smokvina, T., Paciotti, V., Brevet, V., Gilson, E., and Geli, V. (1999) Interaction between Set1p and checkpoint protein Mec3p in DNA repair and telomere functions. *Nat Genet* **21**, 204-208

217. Takahashi, Y. H., Westfield, G. H., Oleskie, A. N., Trievel, R. C., Shilatifard, A., and Skiniotis, G. (2011) Structural analysis of the core COMPASS family of histone H3K4

methylases from yeast to human. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 20526-20531

218. Roguev, A., Schaft, D., Shevchenko, A., Aasland, R., Shevchenko, A., and Stewart, A. F. (2003) High conservation of the Set1/Rad6 axis of histone 3 lysine 4 methylation in budding and fission yeasts. *J Biol Chem* **278**, 8487-8493

219. Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., Zhao, R., and Kutateladze, T. G. (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**, 100-103

220. Tate, C. M., Lee, J. H., and Skalnik, D. G. (2010) CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin. *The FEBS journal* **277**, 210-223

221. Garcia-Martinez, J., Aranda, A., and Perez-Ortin, J. E. (2004) Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell* **15**, 303-313

222. Blackledge, N. P., Thomson, J. P., and Skene, P. J. (2013) CpG island chromatin is shaped by recruitment of ZF-CxxC proteins. *Cold Spring Harbor perspectives in biology* **5**, a018648

223. Smith, E., Lin, C., and Shilatifard, A. (2011) The super elongation complex (SEC) and MLL in development and disease. *Genes & development* **25**, 661-672

224. Thingholm, T. E., Jorgensen, T. J., Jensen, O. N., and Larsen, M. R. (2006) Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nature protocols* **1**, 1929-1935

225. Cantin, G. T., Yi, W., Lu, B., Park, S. K., Xu, T., Lee, J. D., and Yates, J. R., 3rd. (2008) Combining protein-based IMAC, peptide-based IMAC, and MudPIT for efficient phosphoproteomic analysis. *Journal of proteome research* **7**, 1346-1351

226. Ibarrola, N., Kalume, D. E., Gronborg, M., Iwahori, A., and Pandey, A. (2003) A proteomic approach for quantitation of phosphorylation using stable isotope labeling in cell culture. *Analytical chemistry* **75**, 6043-6049

227. Junger, M. A., and Aebersold, R. (2014) Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. *Wiley interdisciplinary reviews. Developmental biology* **3**, 83-112

228. McAllister, F. E., and Gygi, S. P. (2013) Correlation profiling for determining kinase-substrate relationships. *Methods* **61**, 227-235

229. Wu, R., Haas, W., Dephoure, N., Huttlin, E. L., Zhai, B., Sowa, M. E., and Gygi, S. P. (2011) A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nature methods* **8**, 677-683

230. Hornbeck, P. V., Kornhauser, J. M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic acids research* **40**, D261-270

231. Mansour, S. J., Matten, W. T., Hermann, A. S., Candia, J. M., Rong, S., Fukasawa, K., Vande Woude, G. F., and Ahn, N. G. (1994) Transformation of mammalian cells by constitutively active MAP kinase kinase. *Science* **265**, 966-970

232. Vidali, G., Gershey, E. L., and Allfrey, V. G. (1968) Chemical studies of histone acetylation. The distribution of epsilon-N-acetyllysine in calf thymus histones. *J Biol Chem* **243**, 6361-6366

233. Gu, W., and Roeder, R. G. (1997) Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90**, 595-606

234. Kouzarides, T. (2000) Acetylation: a regulatory modification to rival phosphorylation? *EMBO J* **19**, 1176-1179

235. Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L., Grishin, N. V., White, M., Yang, X. J., and Zhao, Y. (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* **23**, 607-618

236. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E., and Mann, M. (2014) The growing landscape of lysine acetylation links metabolism and cell signalling. *Nature reviews. Molecular cell biology* **15**, 536-550

237. Yang, X. J. (2004) The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic acids research* **32**, 959-976

238. Khokhlatchev, A., Xu, S., English, J., Wu, P., Schaefer, E., and Cobb, M. H. (1997) Reconstitution of mitogen-activated protein kinase phosphorylation cascades in bacteria. Efficient synthesis of active protein kinases. *J Biol Chem* **272**, 11057-11062

239. Hunter, T., and Sefton, B. M. (1980) Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 1311-1315

240. Boyle, W. J., van der Geer, P., and Hunter, T. (1991) Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. *Methods in enzymology* **201**, 110-149

241. Robinson, F. L., Whitehurst, A. W., Raman, M., and Cobb, M. H. (2002) Identification of novel point mutations in ERK2 that selectively disrupt binding to MEK1. *J Biol Chem* **277**, 14844-14852

242. Schagger, H. (2006) Tricine-SDS-PAGE. *Nature protocols* **1**, 16-22

243. Boulton, T. G., and Cobb, M. H. (1991) Identification of multiple extracellular signal-regulated kinases (ERKs) with antipeptide antibodies. *Cell regulation* **2**, 357-371

244. Jivan, A., Earnest, S., Juang, Y. C., and Cobb, M. H. (2009) Radial spoke protein 3 is a mammalian protein kinase A-anchoring protein that binds ERK1/2. *J Biol Chem* **284**, 29437-29445

245. Lee, J. H., and Skalnik, D. G. (2002) CpG-binding protein is a nuclear matrix- and euchromatin-associated protein localized to nuclear speckles containing human trithorax. Identification of nuclear matrix targeting signals. *J Biol Chem* **277**, 42259-42267

246. Obenauer, J. C., Cantley, L. C., and Yaffe, M. B. (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic acids research* **31**, 3635-3641

247. Trudgian, D. C. S., R.; Cockman, M. E.; Ratcliffe, P. J.; Kessler, B. M. (2012) Modls: Post-translational modification localization scoring with automatic specificity expansion. *Journal of Proteomics and Bioinformatics* **5**, 283-289

248. Kinoshita, E., Kinoshita-Kikuta, E., and Koike, T. (2009) Separation and detection of large phosphoproteins using Phos-tag SDS-PAGE. *Nature protocols* **4**, 1513-1521

249. Wymann, M. P., Bulgarelli-Leva, G., Zvelebil, M. J., Pirola, L., Vanhaesebroeck, B., Waterfield, M. D., and Panayotou, G. (1996) Wortmannin inactivates phosphoinositide 3-kinase by covalent modification of Lys-802, a residue involved in the phosphate transfer reaction. *Mol Cell Biol* **16**, 1722-1733

250. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840

251. Lundby, A., Lage, K., Weinert, B. T., Bekker-Jensen, D. B., Secher, A., Skovgaard, T., Kelstrup, C. D., Dmytriyev, A., Choudhary, C., Lundby, C., and Olsen, J. V. (2012) Proteomic analysis of lysine acetylation sites in rat tissues reveals organ specificity and subcellular patterns. *Cell reports* **2**, 419-431

252. Guo, A., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K. A., Yang, V., Aguiar, M., Kornhauser, J., Jia, X., Ren, J., Beausoleil, S. A., Silva, J. C., Vemulapalli, V., Bedford, M. T., and Comb, M. J. (2014) Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Molecular & cellular proteomics : MCP* **13**, 372-387

253. Tang, Z., Chen, W. Y., Shimada, M., Nguyen, U. T., Kim, J., Sun, X. J., Sengoku, T., McGinty, R. K., Fernandez, J. P., Muir, T. W., and Roeder, R. G. (2013) SET1 and p300 act synergistically, through coupled histone modifications, in transcriptional activation by p53. *Cell* **154**, 297-310

254. Clouaire, T., Webb, S., and Bird, A. (2014) Cfp1 is required for gene expression dependent H3K4me3 and H3K9 acetylation in embryonic stem cells. *Genome biology* **15**, 451

255. Hazzalin, C. A., and Mahadevan, L. C. (2002) MAPK-regulated transcription: a continuously variable gene switch? *Nature reviews. Molecular cell biology* **3**, 30-40

256. Zhang, H. M., Li, L., Papadopoulou, N., Hodgson, G., Evans, E., Galbraith, M., Dear, M., Vougier, S., Saxton, J., and Shaw, P. E. (2008) Mitogen-induced recruitment of ERK and MSK to SRE promoter complexes by ternary complex factor Elk-1. *Nucleic acids research* **36**, 2594-2607

257. Bledau, A. S., Schmidt, K., Neumann, K., Hill, U., Ciotta, G., Gupta, A., Torres, D. C., Fu, J., Kranz, A., Stewart, A. F., and Anastassiadis, K. (2014) The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development* **141**, 1022-1035

258. Voigt, P., Tee, W. W., and Reinberg, D. (2013) A double take on bivalent promoters. *Genes & development* **27**, 1318-1338

259. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013) RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826

260. Porter, I. M., McClelland, S. E., Khoudoli, G. A., Hunter, C. J., Andersen, J. S., McAinsh, A. D., Blow, J. J., and Swedlow, J. R. (2007) Bod1, a novel kinetochore protein required for chromosome biorientation. *The Journal of cell biology* **179**, 187-197