Diverse Roles for the Mitogen-Activated Protein Kinase ERK2 Revealed by High-throughput Target Identification

By

Scott M. Carlson

B.S. Chemistry
Stanford University, 2005

M.Phil. Computational Biology
Cambridge University, 2006

SUBMITTED TO THE DEPARTMENT OF BIOLOGICAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGICAL ENGINEERING

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2012

Signature of Author:___

Department of Biological Engineering
December 16, 2011

Certified by:_____

Forest M. White
Associate Professor of Biological Engineering
Thesis Supervisor

Accepted by:_____

K. Dane Wittrup
Mares Professor of Chemical Engineering and Biological Engineering
Co-Chair, Graduate Program Committee

Members of the Thesis Committee voting in favor of the defense:

Chair, K. Dane Wittrup
Mares Professor of Chemical Engineering and Biological Engineering
Chair, Graduate Program Committee

Thesis Supervisor, Forest M. White
Associate Professor of Biological Engineering

Ernest Fraenkel
Associate Professor of Biological Engineering

Roger J. Davis
Professor of Molecular Medicine, University of Massachusetts
Howard Hughes Medical Institute

Biographical Note and Acknowledgement

The author attended Leland Stanford Junior University from 2001 to 2005 and received a Bachelor's of Science in Chemistry with a minor in Biology, and was awarded Distinction for academic performance, Honors for advanced coursework and research, a Firestone Medal for outstanding thesis in the natural sciences and the Merck Index Award for outstanding undergraduate in Chemistry.

Following the B.S. the author was awarded a Gates Cambridge Scholarship and attended Cambridge University from 2005-06 where he received a Master of Philosophy in Computational Biology.

The author began as a Ph.D. student in Biological Engineering at MIT in fall of 2006 and has been supervised by Associate Professor Forest M. White. He has been supported during this time by a Graduate Research Fellowship from the National Science Foundation, a Momenta Presidential Fellowship from MIT, a Graduate Fellowship from the David H. Koch Institute for Integrative Cancer Research, a Whitaker Health Science Fellowship, and a Siebel Scholarship.

Other financial support for the work described here has come from AstraZeneca, Pfizer Inc., and NIH grants R01DK42816, R01CA118705 and U54CA112967.

The research described here has been in collaboration with a large number of people and research groups. Individual contributions are noted in the text. Major collaborators include:

Professor Roger Davis, University of Massachusetts Medical School
Assistant Professor Norman Kennedy, University of Massachusetts Medical School

Institute Professor Phillip Sharp, MIT
Dr. Paul Boutz, Sharp Lab, MIT
Mohini Jangi, Sharp Lab, MIT

Associate Professor Ernest Fraenkel, MIT
Candace Coudinard, Fraenkel Lab, MIT
Adam Labadorf, Fraenkel Lab, MIT
Chris Ng, Fraenkel Lab, MIT

Additional technical assistance has been provided by Kevan Shokat, UCSF, and his graduate student Nicholas Hertz, as well as the Swanson Biotechnology Core Facilities at the Koch Institute.

Note on prior publication

Elements of this thesis were published in ACS Chemical Biology in January 2011 (*ACS Chem. Biol.*, 2011, 6 (1), pp 75–85), and have been published in Science Signaling (*Sci. Signaling*, 2011 Oct 25; 4(196): rs11). They are reused here under copyright agreements that allow reuse the material in a thesis or dissertation.

Diverse Roles for the Mitogen-Activated Protein Kinase ERK2 Revealed by High-throughput Target Identification

By

Scott M. Carlson

Submitted to the Department of Biological Engineering on December 16th 2011, in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biological Engineering

ABSTRACT

Many major human oncogenes contribute to cancer in large part by activating the mitogen-activated protein kinases (MAPK) ERK1 and ERK2 (ERK). These kinases are critical in normal physiological processes from development to memory, and their activation in cancer drives growth and metastatic invasion. Understanding the effects of ERK signaling is especially important since vemurafenib, the first pharmaceutical directly targeting the MAPK pathway, has recently been approved to treat advanced melanoma. ERK controls cellular phenotypes by phosphorylating over two hundred known substrate proteins, however new ERK targets are reported frequently. We have used a chemical genetics approach to identify over one hundred novel substrates of ERK2. This approach utilizes an ERK2 kinase with "gatekeeper" mutation that allows it to bind bulky ATP analogs (AS-ERK2). AS-ERK2 can be used to label its direct substrates with thiophosphate in an *in vitro* kinase reaction. To achieve sufficient sensitivity we have improved on existing protocols for identification of thiophosphorylated peptides. Our improved protocol identified over one hundred novel ERK2 substrates in 3T3-L1 fibroblasts and in colon carcinoma cell lines. We investigated one novel ERK2 substrate, the transcriptional repressor ETV3, in detail and found that phosphorylation abrogates binding to DNA by ETV3, and that mutation of key phosphorylated residues to alanine blocks this effect. We also identified several thousand ETV3 targets across the genome. The wide range of genes targeted by ETV3 suggests that it may act as a key regulator of cell cycle and metabolism in some cell types. We have also identified ERK2 substrates in the DLD1 colon carcinoma cell line, including several mRNA splicing factors and members of the MLL family of histone 3 methyltransferases. We are using high-throughput sequencing and biochemical experiments to determine whether these phosphorylation sites control the function of MLL proteins. Taken together these investigations greatly expand our knowledge of the ERK signaling pathway and have revealed greater connectivity among biological processes than had been appreciated.

Supervisor: Forest M. White
Title: Associate Professor of Biological Engineering

Table of Contents

1. Introduction

Virtually all cellular responses to environmental perturbations are controlled in part by tightly regulated protein phosphorylation signaling networks. Aberrant kinase activation or loss of phosphatase activity underlies many major human diseases including cancer, autoimmunity, and diabetes[1-3]. The early discovery of the transforming capability of constitutively active tyrosine kinases has led to extensive research on how signaling is deregulated in cancer[4]. As a result, increased expression or activation of kinases has been found to play a causative role in many cancers, including the BCR-ABL fusion protein in chronic myeloid leukemia (CML), B-Raf in melanoma, the receptor tyrosine kinase (RTK) EGFR in lung cancer and glioblastoma, and the related RTK HER2 in breast cancer[5,6]. Because kinase-activated signaling networks are associated with proliferative, migratory, and invasive phenotypes, small molecule kinase inhibitors are a major area of pharmaceutical research and development[7].

Clinical results and experiments in cell culture have shown that many cancer cells experience "oncogene addiction" such that inhibition or removal of the transforming "driver" mutation blocks malignant phenotypes and leads to apoptosis[8]. Although this effect has been observed in a wide range of cancers our understanding of the process is very limited. This was recently highlighted by Dr. Harold Varmus, Director of the National Institutes of Health, in a request for applications to study twenty-four "Provocative Questions" in cancer biology and treatment ("*While this cell death is an encouraging outcome for therapeutic approaches, we have little knowledge of why these cells become so strongly dependent on the continued expression of an active mutated oncogene*"). These questions will serve throughout this dissertation as a framework linking experiments in molecular biology to issues of long-term clinical significance.

The phenomenon of oncogene addiction suggests that many cancers may have a significant therapeutic window where targeted inhibition of an oncogene will destroy tumor cells or inhibit growth while having a minimal effect on normal tissues. Unfortunately clinical trials with targeted kinase inhibitors as mono-therapies for cancer treatment have been largely disappointing due to confounding issues including lack of specificity, primary resistance because of compensating or redundant signals, rapid development of acquired resistance, and toxicity[7].

Understanding why kinase inhibitors fail in the clinic and how to combine them with other chemotherapies or targeted treatments will require a much deeper knowledge of the signaling networks that they affect. The aim of this research has been to examine one of those pathways in detail by identifying downstream targeted by the mitogen-activated protein kinase (MAPK) ERK2. ERK2, along with the closely related ERK1, is a key signaling molecule involved in cell cycle control, differentiation and metabolism, among other processes. ERK1 and ERK2 are activated by the same upstream signals and share many of their targets; I will refer to their combined activity as the ERK pathway. ERK is downstream of several major oncogenes including over-expression or mutation in EGFR, mutation in KRAS, NRAS, and BRAF. In this dissertation I will present an optimized approach for labeling and identifying substrates of ERK2, discovery of over one hundred novel substrates in a variety of biological systems, a quantitative analysis of how ERK2 substrates are affected by the presence of an oncogenic signal, evidence that ERK2 controls ETV3, a global regulator of gene transcription, and initial work using high-throughput sequencing to define novel roles for ERK signaling in histone methylation and epigenetics.

Targeting kinases in oncology

There is tremendous interest in using small molecule kinase inhibitors as therapeutics for cancer. Imatinib (marketed as Gleevec) was the first clinically effective kinase inhibitor to be designed on the basis of a known disruption in kinase signaling[9]. Approximately 95% of CML carries a reciprocal translocation of chromosomes 9 and 22, leading to fusion of the genes for ABL tyrosine kinase and BCR. The BCR-ABL protein is constitutively active and its activity remains necessary for survival and growth of CML through the entire course of the disease. Motivated by the critical role of BCR-ABL in CML, imatinib was developed by *in vitro* rational design for inhibitors of tyrosine kinases. Imatinib was approved for clinical use in 2001 and often sends the disease into nearly complete remission. Unfortunately the cancer is not eliminated completely and will recur if treatment is withdrawn[9]. In some cases CML can also recur by acquiring secondary mutations that block the activity of imatinib.

Following the success of imatinib there was excitement about an entire generation of rationally-designed kinase inhibitors for treating cancer. Over the past decade this vision has proven extremely difficult to realize[10]. Starting from the 1980s one of the most promising kinase targets was over-expression and activating mutations in Epidermal Growth Factor Receptor (EGFR), which occurs frequently in non-small cell lung cancers, pancreatic cancers, and glioblastoma, as well as other cancer types[11,12]. Efforts to target EGFR led to development of gefitinib and then erlotinib, both of which are potent inhibitors of EGFR. Treatment of advanced cancers over-expressing EGFR with either inhibitor has been shown to extend life by an average of 3 to 6 months[13]. Balanced with potentially serious side-effects and a cost of tens of thousands of dollars it is still debated whether these treatments are cost-effective or even provide a net clinical benefit.

More recently rationally-targeted kinase inhibitors have been developed against BRAF, and specifically the constitutively active form carrying a V600E mutation[14]. BRAF is a kinase upstream in the MAPK pathway and it drives cell cycle and proliferation in large part by activating ERK. Activated BRAF is a critical mutation in approximately 60% of melanomas, and occurs less frequently in colon carcinoma and other cancers. Although BRAF mutation is an early event in development of melanoma (it is present in a large fraction of precancerous dysplastic nevi[15]), clinical trials have focused on treatment of late-stage disease for ethical reasons. The first clinical BRAF inhibitor PLX4032 (Vemurafenib) was very recently approved (August 17, 2011) for treatment of late-stage melanoma with BRAF V600E. Although this is one of the most promising targeted therapies in recent years the progression-free survival is only about 6 months. Known mechanisms of acquired resistance include activation of MEK1 (the immediate downstream target of BRAF), over-expression of the tyrosine receptor kinase PDGFRβ, and activating mutations in NRAS, although these mechanisms cover only about half of cases of acquired resistance[16,17]. Ongoing efforts are examining the effect of treatment with direct inhibitors of MEK, and co-treatment with inhibitors of the AKT pathway (especially targeting PI3K, mTOR, and mTORC).

Elucidating signaling networks

Understanding how to use targeted inhibitors more effectively will require a detailed knowledge of the signaling networks that they affect, and how normal and diseased tissues respond at the molecular and cellular levels. Even very basic questions, like why particular oncogenes occur frequently in specific types of cancer, are essentially unanswered. Another of the NIH Provocative Questions recently highlighted this uncertainty ("*Cancer-causing mutations*

arise under different selection pressures during tumor development. It has been recognized for some time that the frequency or timing of various cancer mutations differs widely among tissues, but we have little mechanistic understanding about why this occurs", PQ-22).

A range of chemical biology techniques have been developed for systemic analysis of signaling networks, including multiplexed or high-throughput assays for activity of key signaling nodes[18-20] and mass spectrometry-based approaches that probe signaling networks on a global scale[21,22]. These approaches are providing new insight into signaling networks, including their function, topology, and response to chemical perturbations. With these tools systemic effects of kinase inhibitors can be quantified with site-specific resolution across the signaling network (reviewed in 23).

Global phosphoproteomic analysis uses affinity enrichment of phosphorylated peptides to provide a broad and unbiased view of the signaling network, including phosphorylation of serine, threonine, and tyrosine residues[24-27]. In these experiments cell lysates are digested to peptides and phosphorylated peptides are enriched using approaches such as immobilized metal affinity chromatography or binding to a titanium dioxide surface. From there high-resolution mass spectrometry and analysis of peptide fragmentation patterns allow specific phosphorylated peptides to be identified by statistical comparison to peptides from *in silico* digest of the proteome[28]. These experiments routinely identify hundreds to thousands of phosphorylated peptides.

Unfortunately, this breadth tends to come at the cost of limited depth. Abundant proteins are more likely to be detected by global analysis, while many important signaling proteins are missed because they are present at much lower levels. Even so, this approach has been used to interrogate the effects of kinase inhibitors, as demonstrated by Pan *et al.*[29] In this study, global

phosphoproteomics was performed to analyze the effects of U0126, a MEK1/2 inhibitor, and SB202190, a p38α inhibitor, on EGFR signaling in HeLa cells, and to quantify the effects of dasatinib, an inhibitor of ABL and SRC family kinases, on the BCR-ABL signaling in a leukemia cell line. Of the several thousand sites identified in this study, approximately 500 increased after stimulation with EGF and 200-300 decreased with inhibitor co-treatment.

Relative to global analysis, targeted phosphoproteomics techniques enable more in-depth characterization of a subset of the phosphoproteome, allowing effects of small molecule inhibitors to be measured on low-abundance phosphorylation sites. This approach has been critical for the analysis of tyrosine kinase inhibitors, as tyrosine phosphorylation represents a small part (<1%) of total phosphorylation in mammalian cells[30]. The most common method for targeted phosphoproteomics uses pan-specific antibodies targeting a subset of the phosphoproteome, such as proteins or peptides containing phosphotyrosine or phosphorylation in the context of specific amino acid motifs[31,32].

In a recent example of targeted chemical phosphoproteomics, Moritz *et al.* developed antibodies recognizing phosphorylation in the context of the RXRXXS/T motif common to basophilic kinases including Akt, RSK and p70 S6K[33]. They used peptide immunoprecipitation (IP) followed by mass spectrometry analysis optimized for basic peptides to identify hundreds of basophilic phosphorylation sites affected by treatment with gefitinib, SU11274 and imatinib (inhibitors of activated EGFR, c-Met, and PDGFR respectively), as well as U0126, wortmannin, and rapamycin (inhibitors of MEK1/2, PI3K and mTOR). This analysis identified phosphorylation sites downstream of activated RTKs in several cancer cell lines, and signaling nodes where tyrosine and basophilic phosphorylation are likely to interact. Importantly, the authors note that their basophilic phosphorylation sites are not necessarily direct substrates of

any particular basophilic kinase, even when a site responds to a particular targeted inhibitor. Demonstration of a direct kinase substrate relationship requires both a direct physical interaction, often shown biochemically, as well as an *in vivo* functional relationship.

Although phosphoproteomic analysis can reveal unexpected downstream effects it provides very little information about the underlying interactions and molecular mechanisms. For example, up- or down-regulation of a particular phosphorylation site does not indicate which upstream kinases, phosphatases or signaling pathways are responsible. Without more specific mechanistic insight it is often difficult to generate and prioritize testable hypotheses. In some cases the upstream pathways can be inferred from literature, predicted by computational analysis looking at interaction network and influence diagrams, or guessed based on kinase sequence specificities. However, since our knowledge of kinase and phosphatase targets is so limited, it remains difficult to infer enzyme-substrate interactions among signaling molecules. The problem is made more difficult because many substrates are extremely low abundance, interactions are often transient, phosphorylation stoichiometry may be low, and particular residues are often phosphorylated by several kinases. Even using perturbation approaches like genetic knockdown or chemical inhibition may lead to pleiotropic effects due to complex feedback and cross-talk within signaling networks.

Chemical genetics

An alternative to global or targeted profiling of phosphorylation sites is to selectively tag and enrich direct kinase substrates, which allows exquisitely sensitive detection of novel substrates by almost eliminating background from other phosphorylated peptides. In particular, chemical genetics is an approach in which a kinase or substrate of interest is mutated to introduce

interactions with a bio-orthogonal small molecule. For instance, Blair *et al.* engineered EGFR and c-Src to be irreversibly inhibited by 6-acrylamido-4-anilinoquinazoline and used a fluorescent form of the inhibitor to study the fraction of active signaling molecules required to initiate downstream signals[34], while Maly *et al.* replaced a particular serine/threonine phosphorylation site with cysteine, then treated with an ATP analog to covalently crosslink the substrate to its kinase[35]. Perhaps the most powerful of these chemical genetics approaches was developed by the Shokat lab in 1997[36]. In this approach, structural analysis was used to design a mutation in the ATP-binding site of v-Src, thereby altering the structure to accommodate a bio-orthogonal ATP analog (for example, N6-benzyl adenosine triphosphosate). To differentiate substrates of the mutated, analog sensitive (AS) kinase from substrates of other kinases, a radiolabeled phosphate was included in the terminal position of the ATP analog, thereby resulting in the transfer of radiolabeled phosphate to the kinase substrate. In addition to radio-labeling of direct substrates, the altered ATP-binding pocket of the AS-kinase rendered it sensitive to inhibition by a variety of bio-orthogonal ATP analog inhibitors which have minimal affinity for endogenous ATP-binding pockets[37,38].

The AS-kinase approach has several advantages over other chemical methods: it allows direct kinase substrates to be definitively tagged, it bypasses the need to design and validate a specific inhibitor for every kinase, and it ensures a limited and consistent set of off-target effects since the same inhibitor can be used for different AS-kinases. Engineering inhibitor sensitivity into specific kinases can be particularly useful to differentiate roles of related kinases, such as the various Src-family kinases. AS alleles have been used to parse out the transcriptional and apoptotic roles of JNK1 and JNK2 following stimulation by tumor necrosis factor alpha and in p53-induced senescence[39,40]. The AS-kinase approach was also used to identify a role for polo-

like kinase 1 (PLK1) in positioning RhoA during cytokinesis, by replacing endogenous PLK1 with an analog-sensitive allele for targeted inhibition[41].

Discovering network topology by chemical genetics

In addition to studying the phenotypic effects of targeted inhibition of a selected kinase, the AS-kinase approach has been also been used to identify substrates for multiple different kinases, including v-Src[42], JNK2[43], PKA[44], CDK1[45], and ERK2[46]. All of these studies were relatively low-throughput, as they relied on radiolabeled ATP-analogs to label substrates, which were then identified through MS analysis of bands cut from 1D or 2D electrophoresis gels. Without the ability to selectively enrich radiolabeled substrates in a complex background of phosphoproteins and non-phosphoproteins, analysis of many of the bands resulted in identification of multiple different proteins, from which the true substrate had to be determined. Although over 40 AS-kinases have been developed, successful substrate identification using these kinases was limited by inability to selectively enrich substrates, and by challenges associated with designing and validating AS-kinases that retain enough activity to function in their biological context[47].

Radioactive tagging on γ-phosphate of the ATP analog was originally used for selective labeling and detection of AS-kinase substrates, but purification and identification remained problematic. Recently, two procedures have been reported that use γ-thiol-phosphate ATP analogs to thiophosphorylate substrates of cyclin/CDK complexes in whole cell lysate[48,49]. In both cases, thiophosphate serves as a nucleophilic "handle" to purify peptides from tryptic digests so that substrates can be identified by tandem mass spectrometry (Figure 1). These experiments rely on the unique ability of the AS-kinase to bind and utilize the ATP analog; any

non-specific utilization of the ATP analog by other kinases in the cell will increase the background and may lead to false-positive substrate identifications.

In each case dozens of substrates of cyclin-dependent kinases were identified in yeast and human cells. A remarkable result from these AS-kinase based substrate identification experiments is that only a small fraction (~10-20%) of the substrates have been previously reported. Given that dozens of substrates have been already been identified for many of these kinases by other methods, including classical biochemistry experiments, the AS-kinase results can be extrapolated to suggest that each of these kinases may have hundreds of substrates. In fact, compared to systems-level phosphoproteomic profiling experiments, the AS-kinase experiments interrogate a much smaller part of the signaling network but reveal a similar level of

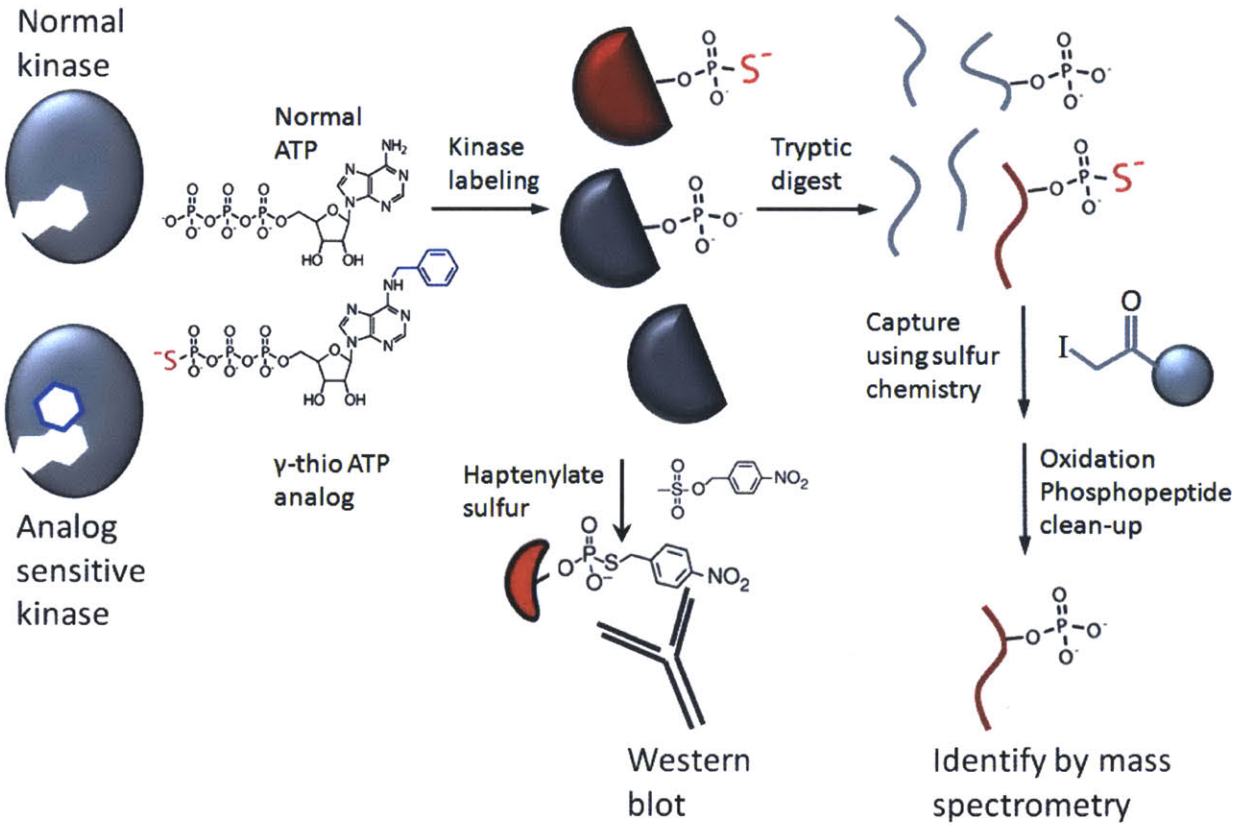


Figure 1: an overview of how AS-kinases are used to label and identify substrates by Western blotting and by covalent capture of thiophosphorylated peptides.

complexity. These experiments suggest that signaling networks are much more highly connected than had been previously recognized and that a tremendous amount of mechanistic detail remains to be discovered at the level of direct interactions among signaling molecules, regulators and effector proteins.

Systems biology of the ERK1/2 mitogen-activated protein kinases

Having introduced the methods used in this dissertation, I would like to introduce the biological system in more detail. The extracellular-signal regulated kinases 1 and 2 (ERK1/2) are ubiquitously expressed in mammalian cells and phosphorylate proteins involved in development[50], glucose homeostasis[51], immune function[52] and memory[53]. Deregulation of ERK1/2 activity is common in cancer and leads to proliferation, migration, resistance to apoptosis, and loss of differentiated phenotypes[54].

ERK1/2 regulates these various biological processes by phosphorylating hundreds of substrate proteins[55]. Almost all of these substrates were identified by traditional biochemical approaches such as in vitro kinase assays, or by tracking phosphorylation in cells treated with activators or inhibitors of the ERK1/2 pathway. Substrate recognition is mediated by protein-protein docking sites, while the substrate binding cleft allows for serine or threonine phosphorylation within an amino acid motif with proline strongly preferred at +1 and common at -2[56]. ERK activity and substrate specificity is further regulated by scaffolds and adaptors that assemble the Raf/MEK/ERK activation cascade and direct sub-cellular localization[57]. Even though many ERK substrates have been identified, incomplete knowledge of ERK targets remains a major hurdle to understanding the myriad biological consequences of ERK activity.

Despite the limitations of global phosphoproteomic analysis, several groups have used proteomic approaches to characterize ERK1/2 signaling by identifying phosphorylation sites that respond to MEK inhibition[29,58-60]. Most of these studies only examined phosphorylation sites that are up- or down-regulated in the presence of the inhibitor. Among these groups, Kosako *et al.* used a combination of affinity chromatography and 2D gel electrophoresis to identify MEK-dependent phosphorylation sites, including 24 potential novel ERK targets, many of them members of the nuclear pore, and validated several of them as direct ERK1/2 substrates. In follow-up experiments the same group showed that phosphorylation of the nuclear pore by ERK1/2 regulates nuclear transport by reducing affinity of the pore for members of the importin-β proteins.

All these studies of ERK1/2 signaling suffer from the weaknesses of global phosphoproteomic analysis that were discussed earlier. To address these concerns we set out to identify additional downstream targets and pathways by using the chemical genetics approach developed by the Shokat lab.

ERK2 with Q103G substitution (AS-ERK2) uses ATP and ATP analogs efficiently and interacts normally with known substrates[46]. AS-ERK2 was originally used to identify ubiquitin ligase EDD and nucleoporin TPR as direct substrates but discovery of additional targets was limited by their low abundance. To address this issue, we have optimized the solid-phase capture of thiophosphorylated peptides and added an additional phosphopeptide enrichment to reduce background, thereby enabling low-level substrate identification. Additionally, we have incorporated quantitative mass spectrometry using stable isotope labeling in cell culture (SILAC)[61] to establish statistical thresholds for identification of substrates over background phosphorylation, thereby decreasing false-positive substrate identification. Taken together these

constitute a major technical advance in the application of analog-sensitive kinases. Our approach is generally applicable to all AS-kinases and we hope that it will allow a much wider range of kinases and biological systems to be examined.

The second chapter of this dissertation describes protocol optimization and identification of novel ERK2 substrates in NIH 3T3-L1 fibroblasts. One of those substrates, a poorly studied the transcriptional repressor ETV3, is especially strongly phosphorylated by ERK1/2. The third chapter discusses the role of ERK activity in regulating ETV3, and presents a genome-wide analysis of ETV3 transcriptional targets. The fourth chapter extends substrate capture and identification to a colon carcinoma system and analyzes the effect of an oncogenic mutation in KRAS on downstream targets of the ERK pathway. This analysis also reveals new targets involved in mRNA processing and histone methylation that may be important in the global rewiring that occurs during carcinogenesis.

2. Protocol optimization and discovery of ERK2 substrates

To identify direct substrates of AS-kinases I began with the method originally developed by Allen *et al.*[62] and Blethrow *et al.*[48]. Proteins in cell lysate are labeled with thiophosphate in a reaction with AS-ERK2 and an N6-substituted γ-thio ATP analog. The lysate is digested to peptides and thiophosphorylated peptides are captured on a solid-phase agarose support functionalized with iodo-acetyl groups (SulfoLink beads), then released by oxidative hydrolysis to yield phosphorylated peptides. Phosphorylated peptides can then be identified by LC-MS/MS tandem mass spectrometry.

The first step in this project was to optimize a protocol to capture and identify thiophosphate-labeled peptides from a complex mixture. I approached this by starting with a simple set of recombinant proteins and progressively increasing the amount of non-labeled protein. My initial studies used recombinant c-Jun protein thiophosphorylated by JNK1 in a reaction with ATPγS. First 10 picomole of labeled c-Jun was digested directly and reacted with SulfoLink beads. Mild washing followed by 5 minute treatment with dilute potassium peroxomonosulfate recovered the labeled peptide with very little non-specific background. The same experiment with a 100-fold excess of bovine serum albumin added to the digest gave a similar result. Next I added 10 picomole of labeled c-Jun to 1 mg of cell lysate in 8M urea, a standard cell lysis buffer for mass spectrometry (the ratio of labeled to non-labeled peptide is approximately 1:100,000 by molar ratio). When I digested the mixture with trypsin and reacted with SulfoLink beads there was no detectable recovery of labeled peptide. I hypothesized that high salt concentration from the digest buffer may interfere with capture of the thiophosphorylated peptide. To address this issue I prepared the digest by buffer exchange on a

hydrophobic cartridge (a standard procedure for White Lab phosphoproteomics experiments). With the additional clean-up step I was able to detect a significant amount of labeled c-Jun peptide, although the apparent yield was greatly reduced by approximately 90% relative to the less complex mixtures.

Since the eventual goal was detection of ERK2 substrates labeled in an *in vitro* kinase reaction I next added thiophosphorylated c-Jun into 1 mg cell lysate collected in a mild buffer suitable for kinase activity. Under this condition there was no detectable recovery of labeled c-Jun. I suspected that detergent from the buffer might interfere with peptide binding to the hydrophobic cartridge. Protein precipitation with either ammonium sulfate or methanol/chloroform before digesting with trypsin resulted in a barely detectable recovery of labeled peptide, along with a dominant signal from non-phosphorylated background peptides. The signal for labeled c-Jun peptide was approximately 1/10,000 of the signal relative to analyzing the same amount of c-Jun by itself. This was encouraging but would not be sufficient to discover novel ERK2 substrates.

Next I considered how to reduce level of background from non-specific binding, and whether sample might be lost by adsorption on surfaces throughout the procedure. Dilute phosphorylated peptides can be rapidly lost by absorption on glass, metal and plastic surfaces. To ensure that this would not interfere with my experiment I assembled a system using microcapillary tubes so that labeled peptides were eluted directly onto a hydrophobic HPLC column. This also ensures that there are no losses from pipetting or transferring the sample between tubes.

Non-phosphorylated background can interfere with detection of phosphorylated peptides in two ways. First is that it takes up to a second for the mass spectrometer to isolate, fragment

and analyze a peptide, so a large number of peptides can overwhelm the ability of the instrument to collect fragmentation spectra. Second is that non-phosphorylated peptides take positive charge more readily than phosphorylated peptides, and can therefore suppress ionization of phosphorylated peptides.

To address the large non-phosphorylated background I first tried to adjust the wash stringency and elution conditions. Washing the SulfoLink beads in 50% acetonitrile, 5M sodium chloride, and 5% formic acid had only a minimal effect on the level of background peptides. This suggested that background peptides were not released from the agarose beads until treatment with oxidizing agent. I varied the concentration of the oxidizing buffer from 1 mg/mL to 10 mg/mL of potassium monopersulfate, and elution times from a few seconds to ten minutes. More intense oxidizing treatments resulted in higher background, but reducing the treatment also limited recovery of labeled peptides. To ensure complete recovery I settled on 10 minute treatment with 2 mg/mL oxidizing agent, even though it results in background signal that would be about 10 times too high to identify labeled ERK2 substrates.

To reduce the level of the non-phosphorylated background I included an additional phosphopeptide enrichment using immobilized metal affinity chromatography (IMAC) , which is a well-developed protocol in the White Lab[24]. This reduced the level of background by approximately 90%, an acceptable level for identifying novel phosphorylated peptides. At this stage the apparent recovery of labeled c-Jun peptide was in the range of 1-3%, and I attempted to identify ERK2 substrates in cell lysate labeled by an *in vitro* kinase reaction. This experiment identified two known ERK substrates, stathmin and TPR, both well-known ERK targets and extraordinarily abundant structural proteins, but the experiment did not identify promising novel targets.

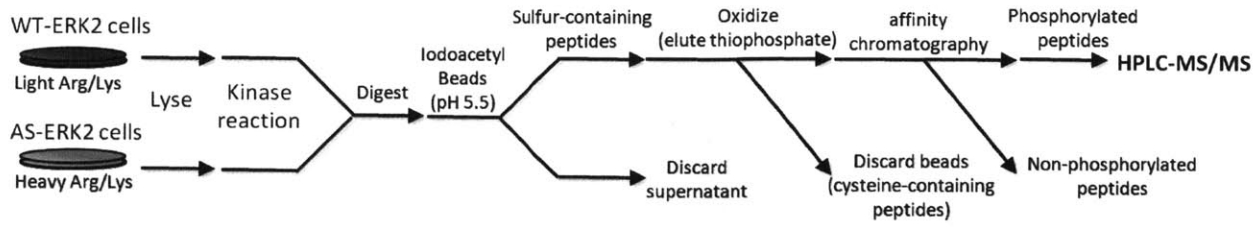


Figure 2: SILAC-labeled cells expressing AS- and wild-type ERK2 are lysed and substrates thiophosphorylated *in vitro*. After protein precipitation and tryptic digestion the labeled peptides are captured on iodoacetyl-agarose. The peptides are recovered by treatment with an oxidization agent and phosphorylated peptides are further enriched by IMAC prior to analysis by LC-MS/MS.

To further enhance recovery of labeled peptides I considered the binding chemistry of thiophosphate to iodoacetyl. Thiophosphate competes with cysteine to react with immobilized iodoacetyl. Because the reaction is greatly accelerated when the thiol group is negatively charged I decided use the difference in pKa between thiophosphoric acid and cysteine to enhance the selectivity of the reaction. Blethrow *et al.* originally used a binding reaction at pH 7.0, corresponding to about 9% deprotonation of cysteine thiol (pKa depends on the local environment but is in the range of 8.0 to 8.3). Since cysteine is at enormous excess relative to thiophosphate, this level of deprotonation could cause thiophosphate to lose out in competition to bind to the surface. Adjusting the pH of the binding buffer to 5.5 caused a several-fold increase in recovery of thiophosphate, with the final apparent yield in the range of 10-20%. The final protocol is represented in Fig. 2 (including SILAC labels as discussed below).

There are several areas where this protocol could be optimized further, especially the final phosphopeptide enrichment step. The microcapillary-based IMAC system is sensitive to handling and requires considerable attention and manipulation. Other phosphopeptide enrichment formats are available, notably batch-mode IMAC that requires much less handling, and titanium dioxide surfaces. Either of these approaches may improve the yield and specificity of the final enrichment and I hope that they will be tested in the future.

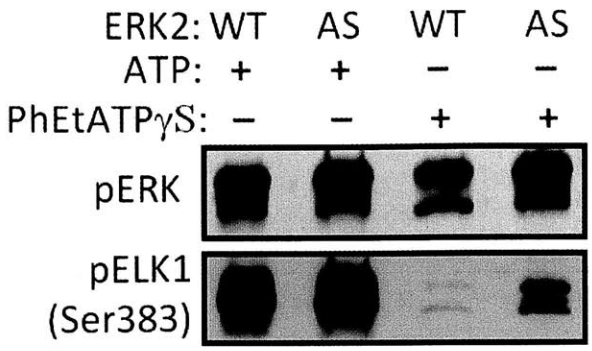


Figure 3: Activated WT and AS-ERK2 were immunoprecipitated from 3T3-L1 cells and incubated with ELK1 and either ATP or ATP analog. Both kinases use normal ATP but only AS-ERK2 thiophosphorylates ELK1 with the ATP analog. Thiophosphate is not recognized by the phospho-ELK1 antibody as well as normal phosphate.

Selective labeling of AS-ERK2 substrates

Dr. Norman Kennedy at University of Massachusetts Medical School provided the AS-ERK2 gene (mouse ERK2 with Q103G mutation). We produced a wild-type ERK2 (WT-ERK2) gene by site-directed mutagenesis and cloned both genes in pBabe retroviral vectors containing puromycin resistance. For an initial system we stably expressed HA-tagged WT and AS-ERK2 in 3T3-L1 fibroblast pre-adipocytes and verified by Western blot and ELISA that that both strains had similar expression of total ERK1/2. These cells were used in all the substrate tagging and identification experiments throughout this chapter. The 3T3-L1 cells were chosen as part of a collaboration involving Pfizer and several members of the White lab to study signaling down-stream of the insulin receptor.

Since the AS-kinase approach relies on selective affinity of the AS-kinase for bio-orthogonal ATP analogs, we confirmed selective utilization of PhEt-ATPγS by AS-ERK2 relative to wild-type ERK2 (WT-ERK2) by an *in vitro* kinase reactions using normal ATP or PhEt-ATPγS (Fig. 3). In this experiment the cells were serum-starved for 3 hours, treated for five minutes with EGF at 100 ng/mL, and lysed in cold kinase reaction buffer. We immunoprecipitated WT or AS-ERK2 using anti-HA antibody immobilized on agarose beads, and then washed the beads and combined them directly with the recombinant ELK protein and ATP or ATP analog as indicated. The reactions were treated with *p*-nitrobenzylmesylate

(PNBM) at 2.5 mM for two hours and then analyzed by Western blot for alkylated thiophosphate.

The next step was to determine the specificity of substrate labeling by AS-ERK2 in the context of a cell lysate. We serum-starved the AS-ERK2 and WT-ERK2 fibroblasts as before and treated them with PMA for two minutes, with or without 30 minute pretreatment with the MEK inhibitor U0126 at 10 μM. The cells were lysed as before, ATP analog was spiked directly into the lysates, and they were incubated at 30° C for *in vitro* labeling reactions. As expected, labeling was strongest in the presence of AS-ERK2, and was dramatically decreased by pre-treatment with U0126. However, there was considerable labeling in the control cells expressing WT-ERK2, indicating utilization of PhEt-ATPγS by wild-type kinases (Fig. 4). Interestingly, AS-ERK2 expressing cells treated with U0126 + PMA had a lower level of labeling than WT-ERK2 cells treated with PMA alone. This suggests that much of the non-specific labeling is coming from normal ERK (especially since it is expressed at a high level), or else from kinases activated downstream of the ERK signal.

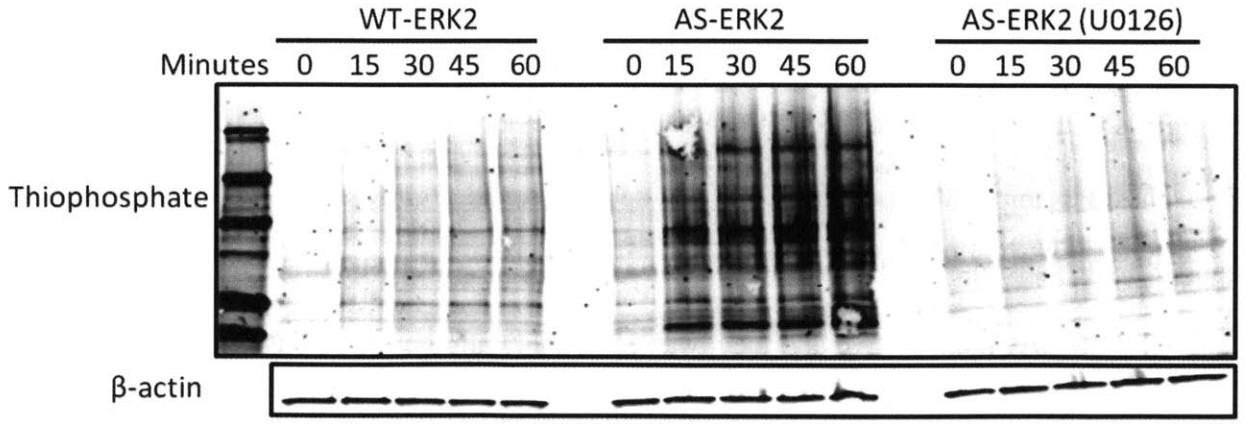


Figure 4: Substrate labeling by WT or AS-ERK2 in cell lysate. Labeling in the WT condition represents non-specific labeling by endogenous kinases. Much stronger labeling is observed with AS-ERK2, and labeling is almost completely abrogated by pretreatment with the MEK inhibitor U0126.

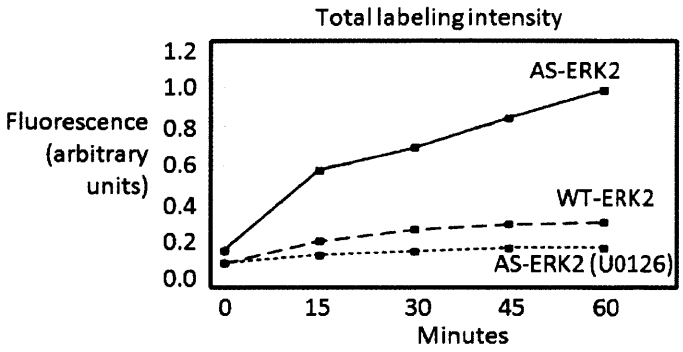


Figure 5: Total intensity for each of the conditions in Figure 4.

The cumulative non-specific labeling by endogenous kinases indicated that a quantitative approach to measuring non-specific labeling was necessary to resolve low abundance substrates from background signals.

Identification of AS-ERK2 substrates

To distinguish ERK2 substrates from non-specific background we used SILAC for quantitative comparison between AS-ERK2 and WT-ERK2 cell lysates. Cells expressing AS-ERK2 were grown in "heavy" L-arginine and L-lysine labeled with ^{13}C or ^{13}C and ^{15}N, while WT-ERK2 cells were grown in L-arginine and L-lysine containing entirely ^{12}C and ^{14}N. Samples were combined after *in vitro* kinase reactions and prior to tryptic digestion. The resulting peptides have distinctive mass signatures: background peptides have light and heavy ion signals with similar intensity, while ERK2 substrate peptides are enriched for heavy ions. Fig. 6 shows example MS spectra for one non-specific signal (FGFR substrate 2), one previously reported ERK substrate (ERF), and two novel substrates (CDC42EP1 and IRS2).

This SILAC quantitative approach was combined with the solid-phase capture protocol to identify thiophosphorylated peptides from six independent labeling reactions with lysates of EGF-stimulated 3T3-L1 cells expressing WT or AS-ERK2. We used MASCOT [63] to match MS/MS spectra to phosphorylated peptides and examined the results for peptides matching the

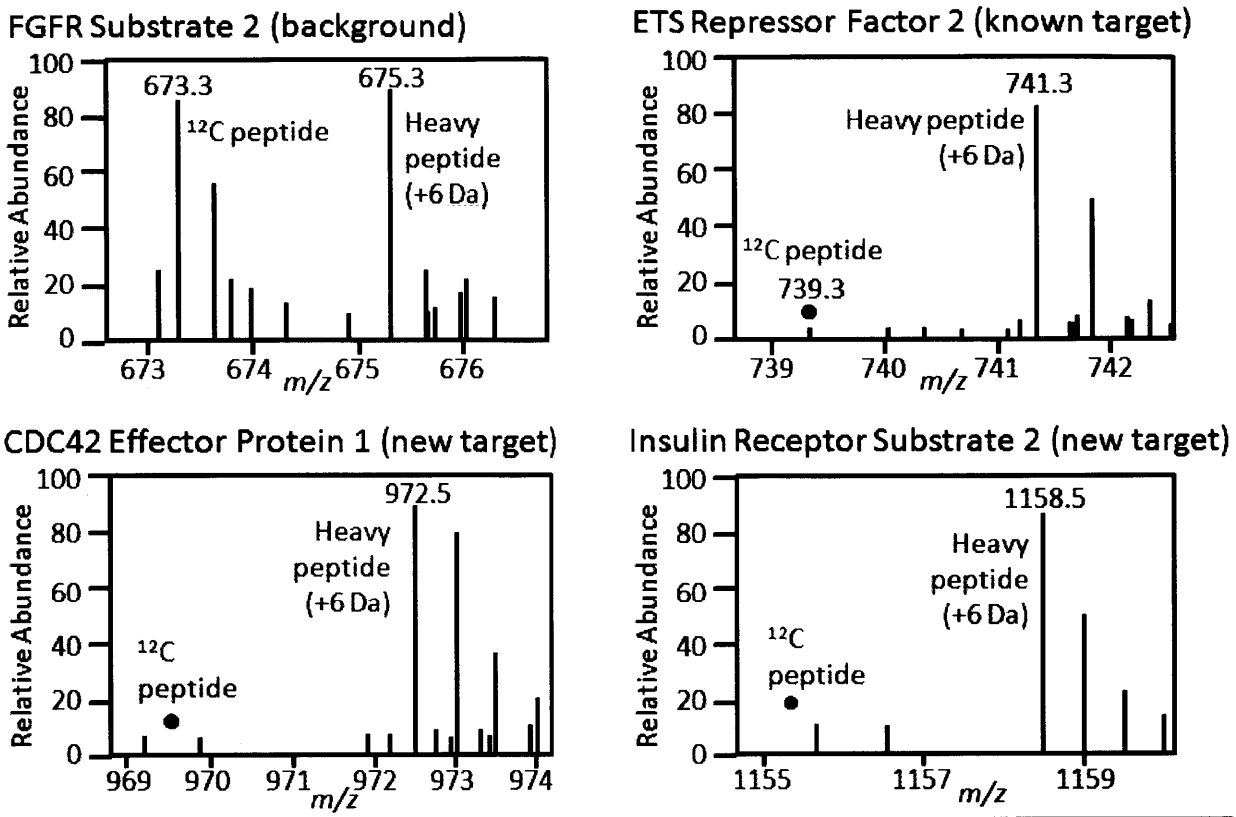


Figure 6: Representative MS spectra showing the effect of SILAC labeling. Light labeled peptides represent the negative control and heavy peptides the AS-ERK2 labeled lysate. The non-specific peptide has equal signals (top left), while the ERK2 substrates are dominated by signal from the heavy channel.

minimal ERK2 substrate motif: pSer or pThr followed by proline. The MS/MS spectra for all potential ERK2 substrates were manually inspected to ensure proper placement of the phosphorylated residue and validation with synthetic peptides was used to confirm any low-confidence assignment. SILAC values for phosphorylated peptides were normalized to values from non-phosphorylated peptides from within the same run.

Phosphorylated peptides from thirteen previously reported substrates were detected: Cx43, DYNC1I2, DYNC1LI1, EGFR, ERF, KSR, LMNA, MAPKAP2, NUP153, TPR, STMN1, RPS3, and SORBS3[55,60]. To differentiate AS-ERK2 substrates from background phosphorylation we used these thirteen reported substrates to set a threshold for the quantitative difference in phosphorylation between the AS-ERK2 and WT-ERK2 expressing cells.

Previously reported substrates had background signals ranging from 6% to 33% of the signal from AS-ERK2 (Appendix 1), while non-specific peptides had ratios around 100% with a standard deviation (S.D.) of 19%. Since heavy-label incorporation into AS-ERK2 cells was greater than 98% (*i.e.* the SILAC ratio of protein from AS-ERK2 cells alone was ratio of less than 2%), we concluded that ERK2 substrate sites were subject to varying degrees of non-specific labeling. This result highlights the importance of quantitatively analyzing the negative control to distinguish background from bona fide substrates. To evaluate approximately 200 potential ERK2 substrates we used a threshold of 3.5 standard deviations below 100% in log-space, corresponding to a light:heavy ratio of 47% and a Bonferroni-corrected familywise error rate of 0.05 assuming that ratios are log-normal distributed under the null-hypothesis (*e.g.* less than 5% chance of one false positive).

A total of 98 peptides on 80 proteins met the criteria for *in vitro* ERK2 substrates, including 67 previously uncharacterized ERK2 substrate proteins. Three phosphopeptides matched the full motif but had light to heavy ratios close to 100% (ATF7, HIVEP1, G3BP-1). In agreement with our data, ATF7 phosphorylation does not respond to MEK inhibition in HeLa cells despite being homologous to ATF2 Thr53 (71 in human), a well-known ERK2 substrate site[64]. Without quantitative data these peptides could have been mistaken as ERK2 substrates.

As expected based on the complex circuitry of the ERK2 network, substrates identified here span a wide range of both upstream and downstream-signaling pathways across many biological processes (Fig. 7). Consistent with dynamic ERK1/2 localization, substrates were associated with multiple cellular compartments, including cytoplasmic, nuclear, and membrane, among other compartments and organelles, and many substrates were further localized to structural components (cytoskeleton, microtubules, nuclear pore, plasma membrane) (Fig. 8).

Figure 7: The 80 proteins identified as *in vitro* substrates of ERK2 in 3T3-L1 fibroblasts. Proteins in red are previously reported ERK2 targets. Proteins in bold have ERK-dependent phosphorylation in intact cells, either in experiments described here (FOX2 and ETV3) or in phosphoproteomic experiments by other groups. Functional categories were derived by manually examining Uniprot and GO annotations for each protein.

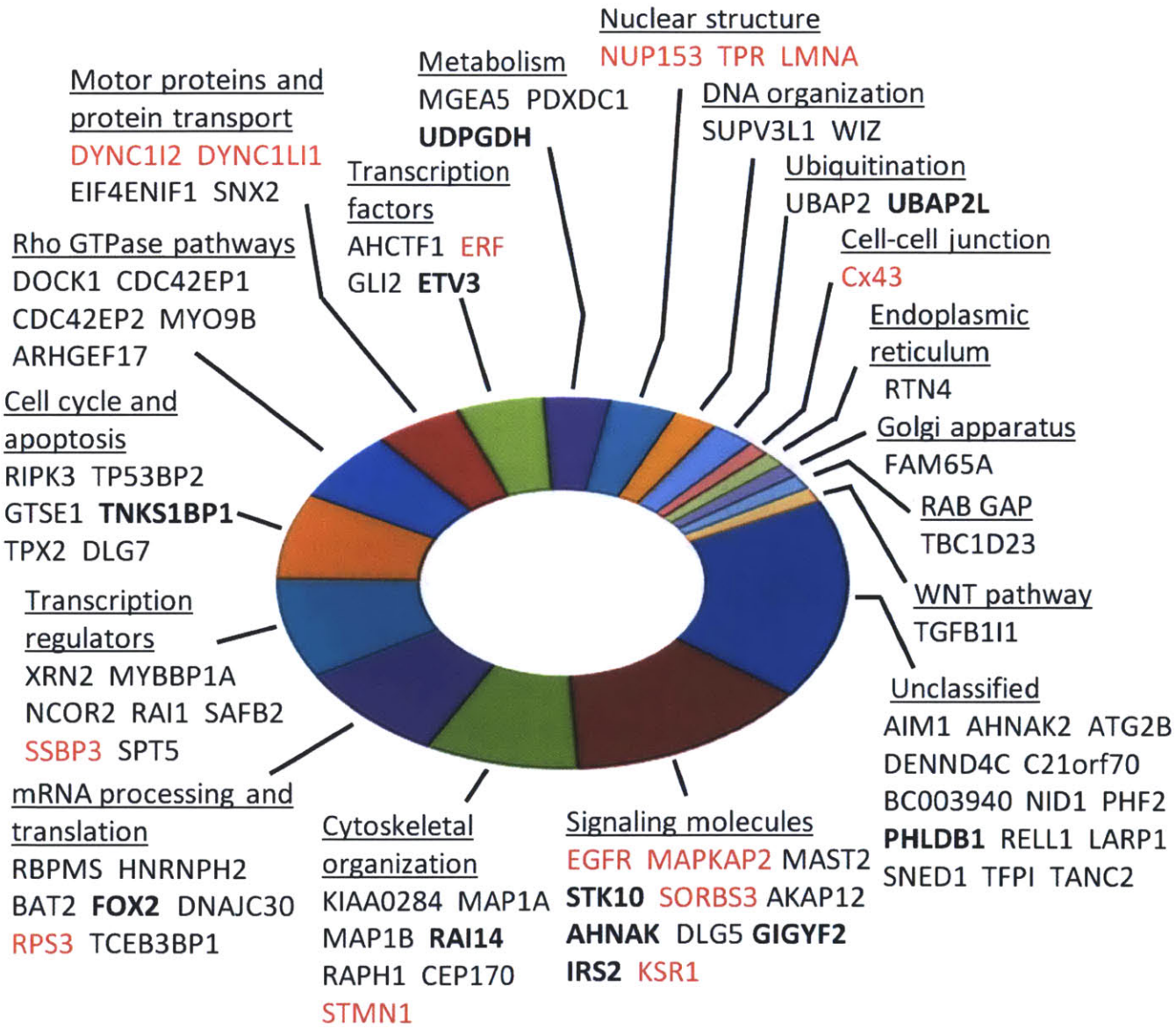
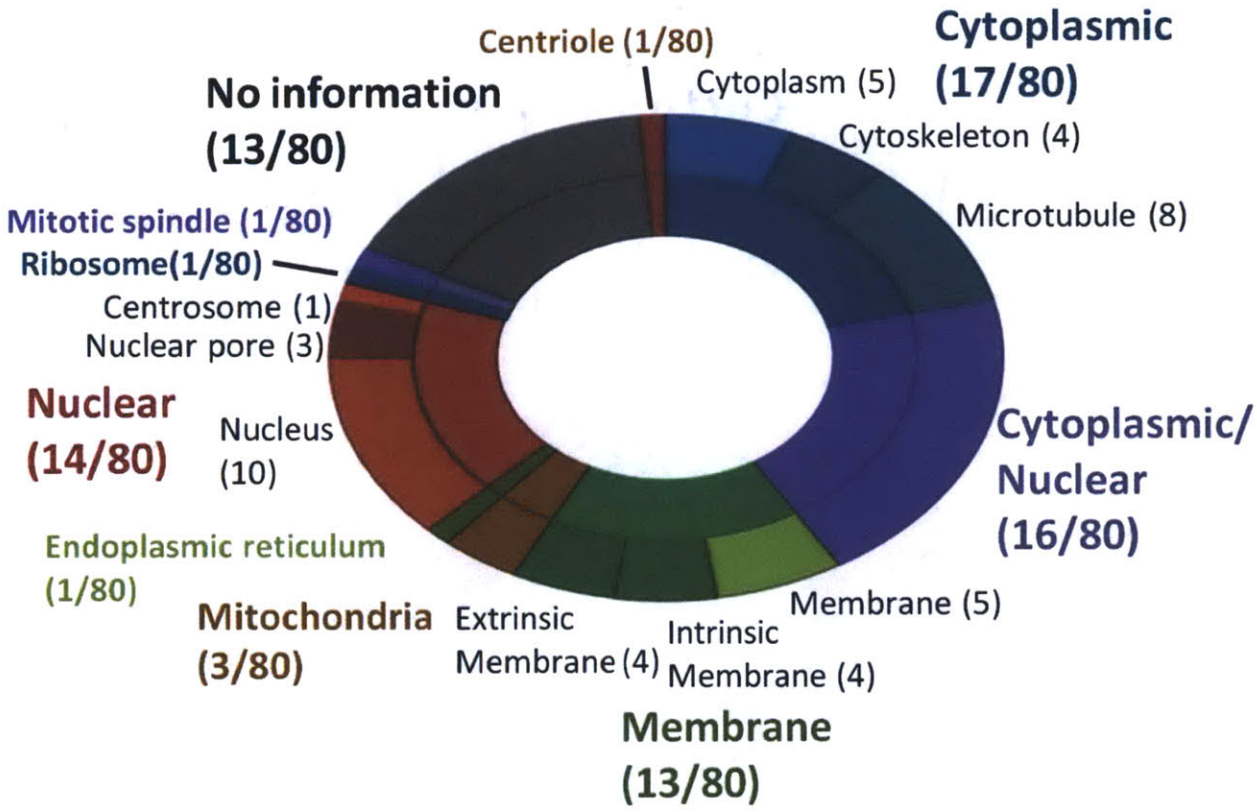


Figure 8: Subcellular location from GO annotations for each of the ERK2 substrates identified in 3T3-L1 cells.



Validating AS-ERK2 substrate specificity

Although the ERK2 Q103G mutation has previously been shown to have no effect on substrate specificity, we tested several of the substrates identified in this study for their ability to be phosphorylated by WT-ERK2. We generated seven recombinant substrate proteins and all seven were phosphorylated *in vitro* by ERK2 to varying degrees (Fig. 9). We also mapped phosphorylation of recombinant CDC42EP1, IRS2 and ETV3 by mass spectrometry following *in vitro* kinase reactions using wild-type ERK2 and identified additional ERK2-dependent phosphorylation sites on each protein (Appendix 2).

To determine whether the substrates identified in this study were consistent with known ERK1/2 binding motifs we used ScanSite to search substrates for ERK binding sequences: D-domains (positively charged residues three to five residues ahead of a hydrophobic sequence) and DEF-domains (FXFP, where X is any amino acid and Y may substitute for F)[65]. Both motifs were strongly enriched in the AS-ERK2 substrates compared to the SwissProt mouse proteome (Fig. 10). We also compared phosphorylation site motifs for AS-ERK2 substrates with known ERK2 substrates according to PhosphoSite (www.phosphosite.org) and found that major features of the two motifs were identical (Fig. 11).

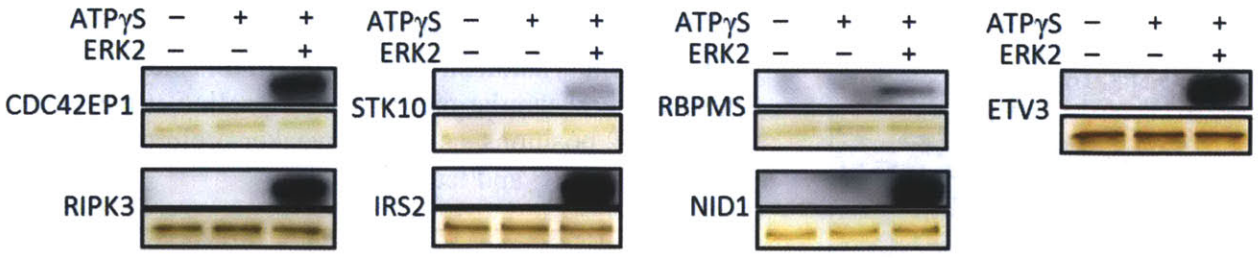


Figure 9: Recombinant ERK2 substrates are thiophosphorylated by wild-type ERK2 in *in vitro* kinase reactions. This indicates that AS-ERK2 substrates are not artifacts of the Q103G mutation.

Although most detected substrates match the minimal ERK1/2 consensus motif SP/TP, and many match the full PXSP/PXTP motif, these motifs are neither necessary nor sufficient to identify ERK2 substrates. In fact, one phosphorylation site on the transcriptional repressor ETV3 was found that lacks the adjacent proline in the +1 position (discussed in the next chapter).

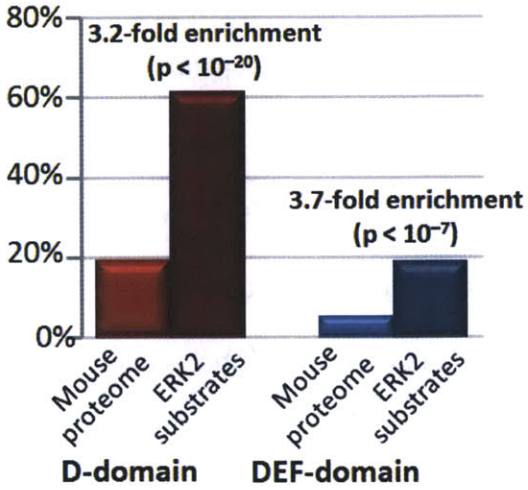


Figure 10: The frequency of ERK-recognition motifs identified by Scansite at high stringency. Both D-domains (consensus FXFP) and DEF-domains (consensus +++XXXLXL, where + is K/R) occur with more than three times their background frequency in the mouse proteome. Both of these enrichments are highly statistically significant.

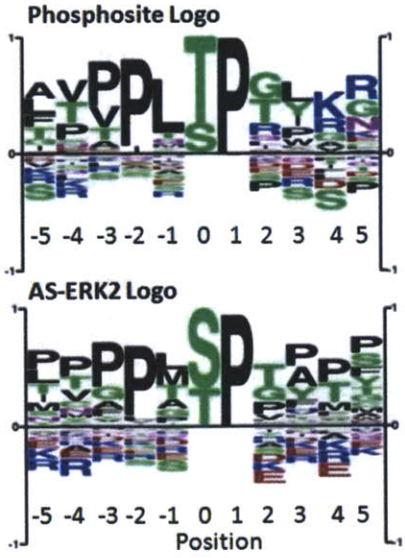


Figure 11: Amino acid frequencies normalized by information content around phosphorylation sites targeted by ERK2. The top plot shows previously described ERK2 substrates from the Phosphosite database. The bottom plot shows the 98 phosphorylation sites identified in this study. Both sets show similar preference for proline at the +1 and -2 positions, as well as weaker preferences for proline at -3 and leucine at -1.

In vivo MEK-dependence of AS-ERK2 substrates

Several of the AS-ERK2 substrates identified here have previously been show to have altered phosphorylation following MEK inhibition. In particular, Pan *et al.* found reduced phosphorylation of AHNAK, ETV3, GIGYF2, PHLDB1, RAI14, TNKS1BP1 and UBAP2L following treatment of EGF-stimulated HeLa cells with U0126[29]. Kosako *et al.* found MEK-dependent phosphorylation of UDPGDH in 3T3-L1 fibroblasts[60], Old *et al.* found MEK-dependent phosphorylation of STK10 in the WM115 melanoma cell line[59], and Fritsche *et al.* found ERK-dependent phosphorylation of IRS2[66]. The combination of *in vitro* interaction and cell-based assays supports these proteins as bona fide ERK2 substrates.

To identify additional phosphorylation sites affected by MEK inhibition in a global manner, we treated SILAC-labeled 3T3-L1 fibroblasts with EGF or PMA with or without U0126 pre-treatment. Lysates were digested with trypsin and antibodies recognizing the phosphorylated motifs PXpSP and PXpTP were used to immunoprecipitate peptides containing potential ERK1/2 substrate sites. These peptide IPs were processed by IMAC phosphopeptide enrichment and analyzed by LC-MS/MS. Only a couple dozen phosphopeptides were identified, probably because of restricted antibody specificity, but a number of known ERK2 substrates were identified, including ERF and the recently reported substrate NUP153, as well as FOX2, a splicing factor and the novel substrate identified by AS-ERK2[67]. Phosphorylation of both NUP153 and FOX2 was strongly dependent on activity of MEK, indicating that these substrates are direct ERK2 targets in their biological context (Fig. 12).

In collaboration with Dr. Paul Boutz in the laboratory of Professor Phillip Sharp we have also looked at phosphorylation of FOX2 in MCF10A cells. We have observed that FOX2 migrates more slowly by SDS-PAGE when ERK is activated. The effect is eliminated by

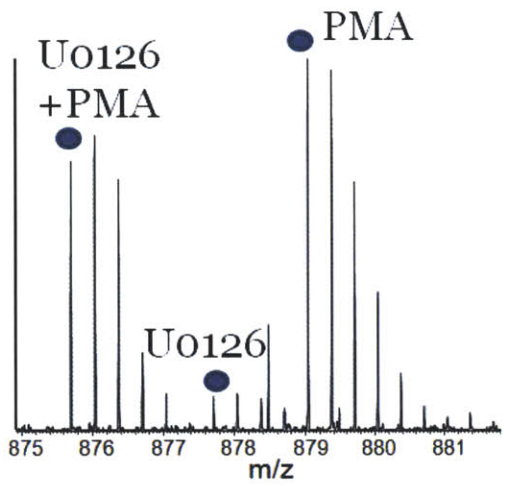


Figure 12: SILAC quantitation for a phosphorylated FOX2 peptide from 3T3-L1 cells treated with U0126 and/or PMA. This peptide shows a strong dependence on ERK activity. The same result has been confirmed in MCF10A breast epithelial cells.

treatment with phosphatase, or by expressing a form of FOX2 that does not include its second exon, which codes for the phosphorylation site. We have also performed detailed mapping for phosphorylation, methylation and acetylation across the FOX2 protein. We have observed only one phosphorylation site that appears to be targeted by ERK, as well as methylation of arginines 361 and 445 and acetylation of lysine 257. Interestingly, there is only one sequence in FOX2 that matches a known ERK-binding motif, and that sequence is itself phosphorylated by an unknown kinase. The ERK-binding motif and ERK-target site both fall within an alternatively spliced cassette that is predominantly expressed in well-differentiated breast epithelial and carcinoma cell lines (personal communication, Paul Boutz). This leads us to believe that FOX2 contains an ERK-responsive cassette that can be regulated by alternative splicing, and possibly by phosphorylation of the ERK-binding motif. We used SILAC-labeled cells for quantitative co-IP to look for changes in protein binding when FOX2 is phosphorylated by ERK. Although we identified a number of heterogeneous nuclear ribonucleoproteins that associate with FOX2, none of them seem to do so in a phosphorylation-dependent manner. Additional experiments with phosphorylation site mutants are ongoing.

A complete list of the peptides identified in the motif IP experiment is contained in Appendix 3, along with a similar experiment in 3T3-L1 adipocytes which also found MEK-

dependent phosphorylation of FOX2 following stimulation with insulin (adipocyte data was contributed by Dr. Katrin Schmelzle, a former postdoctoral associate in the White Lab).

Alternative methods and cellular systems

Comprehensive identification of kinase targets will require analysis of a wide range of cellular systems. A simple strategy would be to prepare activated AS-ERK2 from a strongly-expressing cell line, immunoprecipitate the kinase, and combine it with lysates from a range of cell types. We tested this by using thiophosphate Western blot to compare lysate labeled using immunoprecipitated AS-ERK2 with lysate from cells that had expressed AS-ERK2 directly. There was a stronger signal and more bands visible when AS-ERK2 was expressed in the target cells. We suspect that substrate recognition is enhanced by formation of protein complexes, possibly including scaffold and substrate proteins, prior to the lysis.

Based on this result we decided to explore expressing AS-ERK2 in additional cell culture systems. First we differentiated the 3T3-L1 fibroblasts into adipocytes using an established protocol[68]. It proved difficult to get strong substrate labeling in adipocyte lysates, perhaps because of the abundant lipids. We also expressed AS-ERK2 in MCF10A breast epithelial cells and in HepG2 hepatocellular carcinoma cells. MCF10A cells normally have tightly-packed sheet-like morphology but can be induced to undergo epithelial-mesenchymal transition (EMT) by treatment with TGFβ. When we conducted *in vitro* labeling reactions with these cells they showed an amount of labeling similar to the 3T3-L1 cells. Unfortunately when AS-ERK2 was expressed in MCF10A they changed their morphology in a similar manner to TGFβ treatment. The same did not occur when we expressed WT-ERK2 in the MCF10A cells, perhaps suggesting that a difference in basal activity resulting from the Q103G mutation. We determined by

Western blotting that exogenous ERK2 expression was similar in each case, but the phenotypic difference may have resulted from a small difference in expression. We abandoned this system since it is not clear how to interpret the biological relevance of ERK substrates in MCF10A cells following the apparent EMT.

Substrate labeling would be more biologically relevant if the reaction could occur in intact cells. We experimented with permeabilizing the 3T3-L1 fibroblasts and HepG2 cells. Treatment with the cholesterol-binding detergent digitonin did permeabilize both 3T3 and HepG2 cells to small molecules, but it also resulted in rapid detachment of 3T3 cells from the surface. HepG2 cells were more robust but we were unable to achieve a degree of substrate labeling even remotely close to that observed in *in vitro* reactions. There are several major challenges with this approach, especially competition with normal ATP (present at millimolar concentrations in most cells) and maintaining biological functions long enough to complete the reaction. Detecting substrates from *in vitro* labeling already stretches the sensitivity of our mass spectrometers, so we decided that identifying substrates from weaker labeling *in vivo* was unlikely to reveal new ERK2 targets. Our experience is consistent with work from the Shokat group. The only substrate identified when they conducted in-cell labeling reactions with ERK2 was TPR[62]. TPR represents the strongest signal that we have seen in our own experiments and seems to be two to three orders of magnitude more abundant than most other substrates.

An alternative approach to identify labeled substrates could be using the thiophosphate antibody for protein IP. The Shokat group has identified a small number of kinase substrates by this approach and I have conducted a number of similar experiments. In my experiments the antibody demonstrated very poor specificity, always pulling down a large excess of non-thiophosphorylated proteins. A major disadvantage of protein IP is that proteins can often be

detected without discovering the thiophosphorylated residue. Given this weakness we did not choose to spend time optimizing the IP.

Discussion

Through an improved solid-phase capture chemical genetics strategy we identified a large number of known and novel substrates of the AS-ERK2 kinase. These novel substrates reveal a wide range of new connections between ERK2 and other signaling pathways, including adaptors and signaling molecules, regulation and effector proteins of the Rho GTPases, transcription factors, splicing regulators, and structural proteins. Validation using wild-type ERK2 shows that these substrates are not artifacts of the Q103G mutation, and cell-based assays indicate that a significant fraction display MEK-dependence, and are therefore likely to be relevant *in vivo*. It is interesting that only 13 of the 80 substrates have been previously reported, and that these represent a small fraction of the approximately 200 reported substrates. The limited overlap suggests that ERK2 may phosphorylate a much larger number of proteins than previously recognized. We may not have identified a larger fraction of all the previously reported substrates because this cell line only expresses a restricted set of substrates. It may also be that some targets are not efficiently phosphorylated in the context of an in vitro reaction, either because they depend on particular localization or because they involve bridging interactions through scaffolds. Comprehensive identification of these targets will require a variety of complementary technologies and detailed examination of a wide range of cell types.

Novel substrates expand on the already considerable cross-talk between ERK and other pathways. For example, regulation of cell shape and migration is mediated in part by Rho GTPases, but very few ERK substrates have been reported in Rho signaling pathways[69].

Identification of ERK2 substrate sites on RhoGEFs DOCK1 and ARHGEF17, the RhoGAP MYO9B and the effector proteins CDC42EP1 and CDC42EP2 suggest that ERK regulates these pathways both upstream and downstream from the Rho GTPases. The ERK phosphorylation sites on these proteins may provide novel connections between MAPK signaling and regulation of cell morphology and migration. Additional investigation is needed to determine whether these phosphorylation sites function in normal physiological regulation of cytoskeletal structure and cell shape, or in pathological processes such as cancer metastasis.

Phosphorylation of adaptor proteins may represent feedback loops by which ERK regulates PI3K/Akt activity. Homology between the IRS2 target residues and a well-characterized sites on IRS1 suggests that phosphorylation by ERK may reduce the effect of insulin signaling by competing with an overlapping PI3K p85 binding site[70]. Serine 907 on mouse IRS2 was recently identified as ERK-responsive in hepatoma cells and liver tissue[66]. Identification of additional ERK phosphorylation sites raises the possibility that the kinase regulates IRS2 through several modes of negative feedback. This mechanism is likely to be part of normal feedback regulation associated with insulin and other growth factor signaling, and could connect pathological ERK activation to insulin resistance caused by inflammation. Phosphorylation of GIGYF2A, GRB10-associated adaptor protein, is another potential interaction with insulin signaling[71], although this protein is not as well characterized as IRS2.

Another interesting point of cross-talk between ERK and other pathways is represented by phosphorylation of the GLI2 transcription factor. Although the ERK phosphorylation site has not previously been reported, it is proximal to residues near the C-terminus that are phosphorylated by PKA and GSK3 to regulate GLI2 stability, and may therefore be critical for MEK-dependent cross-talk between Hedgehog and EGFR signaling[72].

We have also shown that some novel substrates including FOX2 are phosphorylated by ERK2 in a biological context. *In vivo* validation is important because kinases are often promiscuous when removed from their biological environment, possibly because of loss of regulation from sub-cellular localization and protein complexes. Some of the phosphorylation sites identified *in vitro* have also been observed to have MEK-dependent phosphorylation in previous phosphoproteomics studies. The combination of *in vitro* and cell-based assays will continue to be important in elucidating the structure of kinase signaling networks.

Further development of methods for analysis of thiophosphorylated peptides will continue to expand the range of biological systems amenable to investigation by AS-kinases, and quantitative characterization of background labeling makes it possible to identify substrates even in the presence of WT-kinase utilization of PhEt-ATPγS. We expect that comparisons between cell types, growth factor treatments, or genetic perturbations will reveal how MAP kinase signaling is rewired in different contexts.

Methods

Antibodies and reagents

Antibodies for phospho-ELK1, phospho-ERK1/2, phospho-motifs, anti-HA Alexa488 conjugate, goat anti-rabbit Alexa488 conjugate, PMA and recombinant ERK2 kinase were from Cell Signaling Technology. Normal rabbit IgG for immunofluorescence was from Santa Cruz Biotechnology. Monoclonal rabbit antibody against alkylated thiophosphate and *p*-nitrobenzylmesylate (PNBM) were from Epitomics. N6-(2-Phenylethyl)adenosine-5'-O-(3-thiotriphosphate) (PhEt-ATPγS) was from Biolog (Bremen, Germany). Adenosine-5'-O-(3-thiotriphosphate) (ATPγS), U0126, anti-HA agarose beads and anti-ETV3 antibody were from Sigma-Aldrich. Anti-HA antibody for ChIP was from Abcam. Media and amino acids for SILAC and iodoacetyl-agarose beads were from Pierce/Thermo Scientific. Epidermal growth factor (EGF) was from Peprotech. Q103G ERK2 (AS-ERK2) with N-terminal HA tag was a gift from Dr. Roger Davis (University of Massachusetts Medical School). HA-tagged human ETV3 was synthesized by GeneWiz Inc. Genes for mammalian expression were cloned into the pBabe-puro-IRES-EGFP vector (AddGene plasmid 14430).[73]

In vitro substrate labeling

Matched WT-ERK2 and AS-ERK2 expressing 3T3-L1 cells were generated by retroviral transduction followed by puromycin selection, and 3T3-L1 cells were grown in DMEM containing stable isotope labeled L-Lysine and L-Arginine, with 10% dialyzed fetal bovine serum and 230 mg/L L-Proline. Cells were treated as indicated, washed with cold PBS and lysed in 500 µL kinase buffer. Protein levels were determined by BCA assay and lysates were diluted to uniform concentration. Lysates were immediately supplemented with 1 mM GTP and 50 µM

N6-PhEt-ATPγS and incubated at 30°C for 30 to 90 minutes. Time-course reactions were terminated by adding EDTA to 50 mM and placing the reaction on ice. For Western blots, 13.5 μL of lysate was treated with 1.5 μL 25 mM PNBM in 50% DMSO for 2 hours.

Identification of thiophosphorylated peptides

Proteins from kinase reactions were precipitated in methanol/chloroform, suspended in digest buffer by sonication in 100 mM NH_4Ac pH 8.9, 1 mM $CaCl_2$, and digested with trypsin (1:100 w/w) overnight with rotation. Digests were acidified with acetic acid (1:6 v/v) and desalted on a C18 Sep-Pak (Waters), lyophilized and re-suspended at 20 mg/mL in binding buffer (25 mM HEPES pH 7.0 in 50% ACN) with 1 mM TCEP and 250 μg/mL bovine serum albumin (final pH 5.5). SulfoLink beads (25 μL per mg peptide) were equilibrated in binding buffer. Peptides were added to beads and rotated overnight in darkness at room temperature. Beads were washed 2x in binding buffer, quenched with DTT, and washed in binding buffer, 5% formic acid, and again in binding buffer. Beads were loaded in a fused silica capillary (530 μm I.D.), washed briefly with 0.1% acetic acid and peptides were eluted using 2 mg/mL potassium peroxomonosulfate (Oxone) directly onto a POROS R2 trapping column. The POROS column was washed with 0.2M acetic acid and peptides were eluted with 70% ACN 0.2M acetic acid to a immobilized metal affinity chromatography column (IMAC) for IMAC-LC-MS/MS analysis[24] on an LTQ-Orbitrap XL mass spectrometer (ThermoFisher Scientific). A detailed protocol is available in the Supplementary Methods. Data files were converted to MASCOT generic format (.mgf) using DTASuperCharge (version 1.19) and searched using MASCOT[63] (version 2.1) against the NCBI mouse proteome using peptide tolerance of 12 ppm and MS/MS tolerance of 0.7 Da. Peptide identification and phosphorylation site assignments were manually validated to

ensure correct identification and quantitation. Any peptide with a low-confidence assignment was confirmed by comparison to a synthetic peptide.

Phospho-motif peptide immunoprecipitation

SILAC-labeled fibroblasts were pretreated with either 5 or 10 μM U0126 for 10 minutes and then stimulated with EGF (100 ng/mL) for 20 minutes. Differentiated 3T3-L1 adipocytes were produced as described[68]. Adipocytes were pretreated with U0126 at 10 μM for 10 minutes and then with insulin at 150 nM for 15 minutes. Cells were lysed with a solution containing 8 M urea with 1 mM Na_3VO_4 and processed as described for peptide labeling with 4-plex iTRAQ reagents (AB Sciex). Samples were dissolved (Tris-HCl pH 7.4, 0.03% NP40), immunoprecipitated with antibodies against PXpSP and PXpTP phospho-motifs, and analyzed by IMAC-LC-MS/MS using an LTQ-Orbitrap XL or QStar XL (AB Sciex) mass spectrometer (for fibroblasts or adipocytes respectively) as described[68].

Recombinant ERK2 substrates and *in vitro* kinase reactions

Clones were from OpenBiosystems (clone IDs ETV3: 4918332, RIPK3: 3590770, STK10: 40111177, NID1: 40110859, RBMPS: 3586451) except CDC42EP1 (OriGene catalog #MC205344) and IRS2 (Addgene plasmid 11373). CDC42EP1, ETV3, NID1, RBPMS and RIPK3 were cloned into pET-16b (Stratagene), while IRS2 and STK10 were inserted into pET-100 (Invitrogen) using directional TOPO cloning. Vectors were transformed into BL21 Star (DE3) *E.coli* (Invitrogen) and induced using 1 mM IPTG for three hours during log-phase growth of a 2 mL culture. Bacterial pellets were lysed using B-Per bacterial lysis kit with DNAase, lysozyme, and Halt protease inhibitors (Thermo Scientific). Recombinant proteins

were recovered from inclusion bodies by solubilization in 8M urea, except RIPK3, which was solubilized with 0.2% sodium dodecyl sulfate in 40 mM Tris-HCl pH 7.5, 100 mM NaCl. Recombinant proteins were stored at 10 ng/uL in 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 50% glycerol.

Kinase reactions used 10-30 ng of substrate, 12.5-25 ng recombinant ERK2 and ATPγS at 50-100 μM in kinase buffer (25 mM Tris-HCl pH 7.5, 5 mM β–glycerolphosphate, 2 mM dithiothreitol, 0.1 mM Na_3VO_4, 10 mM $MgCl_2$), and incubated 30 minutes at 30°C. Reactions with RIPK3 were supplemented with 100 mM thiourea to increase solubility. Half of each reaction was silver stained as a loading control, while half was treated with PNBM at 2.5 mM for two hours prior to Western blotting for thiophosphate.

Phosphorylation mapping of recombinant proteins by HPLC-MS/MS

Kinase reactions were performed for 1 hour with approximately 100 ng of each substrate and ATP at 100 μM. Matched negative control reactions omitted ATP. Each reaction was run by SDS-PAGE and proteins visualized by Coomassie stain or mass-spectrometry compatible silver stain (Thermo Scientific). Matched bands were cut and destained, and processed as described for overnight digestion in trypsin (Promega, 12.5 ng/uL in 50 mM ammonium bicarbonate pH 8.9) or chymotrypsin (Sigma-Aldrich, 125 ng/uL in 100 mM Tris 10 mM $CaCl_2$ pH 8.0 at 30° C).[74] Reactions were quenched with 5% formic acid in 50% acetonitrile and peptides were eluted by dehydrating the gel band twice with 100% acetonitrile. Peptide samples were dried to about 2 uL in a vacuum centrifuge and resuspended in 30 uL 0.1% acetic acid. IMAC was used to enrich phosphopeptides from digests of CDC42EP1 and IRS2, while ETV3 was loaded directly onto a C18 pre-column (10 cm, 100 μm I.D.). Pre-columns were placed in-line on an HPLC connected

to a C18 column with electrospray tip (10 cm, 50 μm I.D., 1 μm tip flowing at approximately 20-40 nL/min). Peptides were eluted using a piece-wise linear gradient from 0% to 70% acetonitrile in 0.2M acetic acid (4 min: 9.1%, 50 min: 29.4%, 57 min: 42%, 60 min: 70%) and analyzed on an Orbitrap XL hybrid mass spectrometer. The mass spectrometer was running in data-dependent mode where each cycle included an Orbitrap MS scan with 100,000 target resolution followed by isolation and Collision-induced Dissociation of up to six ions (charge state 2-5) for analysis in the LTQ ion trap. MS/MS spectra were extracted using DTASuperCharge (version 1.19) and identified by MASCOT (Matrix Science, version 2.1). Phosphorylated peptides were confirmed by manual inspection of each spectrum.

ERK binding site motif analysis

ScanSite 2.0 (http://scansite.mit.edu) was used to search SwissProt protein sequences for high-stringency ERK1/2 binding sequences (ScanSite motifs "Erk D-domain" and "Erk1 Binding" for D-domain and DEF-domain respectively).[75] Where an appropriate protein sequence was not available in SwissProt, protein sequences were uploaded directly from the National Center for Biotechnology Information Entrez protein database. The frequency of each motif in AS-ERK2 substrates was compared to SwissProt mouse proteome (Sept 22 2009 release) using the X^2-test.

3. Phosphorylation by ERK regulates the transcriptional repressor ETV3

Having identified novel substrates of ERK2 we set out to determine whether any of the novel substrates are regulated somehow by phosphorylation. The transcriptional repressor ETV3 (also called METS and PE-1) was an especially interesting target because it is closely related ETS-domain transcription factors that are already known to be regulated by ERK. It was also intriguing that serine 139 of ETV3 was the only site identified in Chapter 2 that is phosphorylated by ERK2 despite lacking proline at the +1 position. Previous work on ETV3 is very limited (relevant papers can literally be counted on your fingers), but initial work suggested that it is not regulated by ERK[76].

Before analyzing the biology of ETV3 in detail it was important to validate the identification of ETV3 phosphorylation by generating an MS/MS of the synthetic peptide, and conducting an *in vitro* kinase reaction that confirmed that ERK2 strongly phosphorylates ETV3. MEK-dependent ETV3 phosphorylation was also detected in cells, as shown by the strong MEK-dependent gel shift by Western blot in DLD1 colon carcinoma cells and HEK 293T cells treated with PMA compared with those treated with the MEK inhibitor U0126 (Fig. 13).

To determine whether ERK2 phosphorylates other sites on ETV3 we produced the

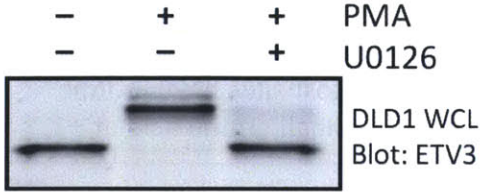


Figure 13: The migration of ETV3 by SDS-PAGE is greatly reduced when ERK is activated. The same result was obtained with ETV3 transfected into 293T cells and visualized by blotting for HA tag.

recombinant protein in *E.coli* and mapped phosphorylation by LC-MS/MS following an *in vitro* kinase reaction. Eight phosphorylated serine residues were detected, none of which were detected in a negative control (Fig. 14). Five of the eight sites are followed by proline, while two are PXpSAP motifs and one is PXpSG. It seems that proline at the -2 position and a small amino acid at +1 may be sufficient for recognition by ERK2 in the absence of proline at +1. As other groups have reported non-canonical sites phosphorylated by various MAP kinases, we suspect that these sites may be systematically overlooked because it is uncommon to comprehensively map phosphorylation on putative kinase substrates[77].

ETV3 is related to ERF, another ETS domain-containing transcriptional repressor. ERF translocates rapidly from nucleus to cytoplasm following phosphorylation by ERK but ETV3 has been reported to be unresponsive to ERK activity in reporter gene assays[76]. As a result, ETV3 is thought to permanently repress its transcriptional targets. Consistent with this previous observation, we detected predominantly nuclear localization of ETV3 in 293T cells transfected with HA-tagged ETV3 regardless of phosphorylation state. We observed the same result in DLD1 colon cancer cells treated with U0126 or stimulated with the PMA and stained for endogenous ETV3 (Fig. 15).

Figure 14: Phosphorylation sites identified on recombinant ETV3 following *in vitro* kinase reaction. The phosphorylation motif around each site is indicated. Most sites have proline at the +1 position, but there are several sites with proline at -2 and a small amino acid at +1.

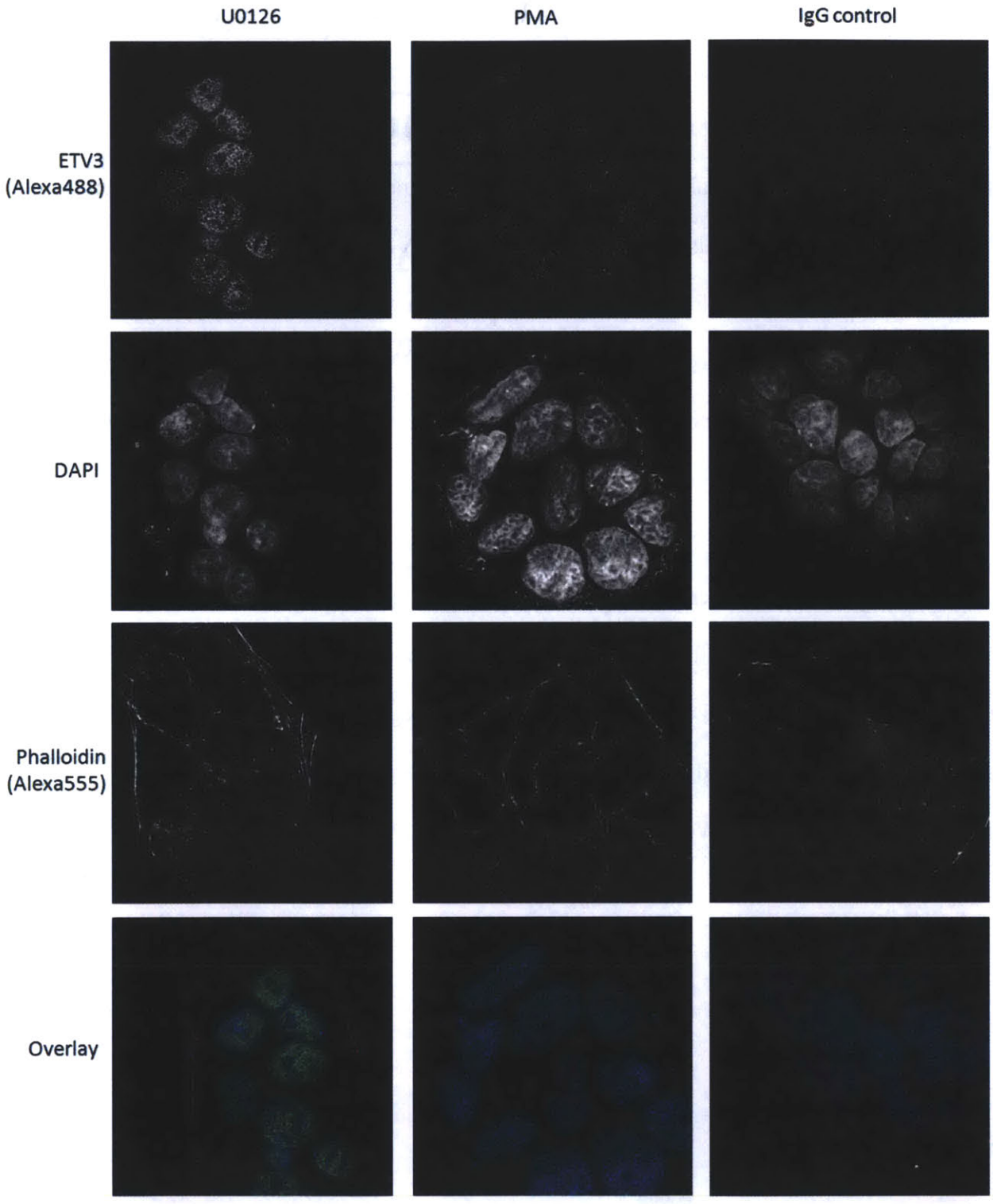


Figure 15: Immunofluorescence of ETV3 in DLD1 cells. Unlike the related ETS-domain transcriptional repressor ERF, ETV3 remains in the nucleus regardless of phosphorylation by ERK. We obtained the same result with ETV3 transfected in 293T cells.

Because transcriptional target sites of ETV3 have never been broadly characterized we began by conducting chromatin immunoprecipitation and high-throughput sequencing (ChIP-SEQ). Only a mouse clone of the ETV3 gene was available so we had the human ETV3 gene synthesized with a C-terminal HA tag (the DNA-binding domain is N-terminal). We expressed this protein in HEK293T cells, treated the cells with PMA or U0126, and prepared the cells for analysis by ChIP-SEQ. ChIP and sequencing preparation were conducted by Candace Chouinard in the laboratory of Professor Ernest Fraenkel (MIT Biological Engineering), high-throughput sequencing was conducted by the MIT BioMicro Center, and the data was processed by Adam Labadorf, also in the Fraenkel Lab.

This experiment revealed several thousand binding sites (see the supplemental data in Carlson *et al.* in press at Science Signaling), most of which were within 1 kilobase of known transcription start sites (Fig. 16), including previously reported binding near the beginning of the *myc* gene (Fig. 17). We also observed binding near the beginning of *etv3* itself, *etv6* (a.k.a. *tel*) and *ddx20*, a partner required for ETV3 activity, suggesting that ETV3 negatively regulates itself and related ETS factors[78]. We used THEME to identify DNA-binding protein motifs near ETV3

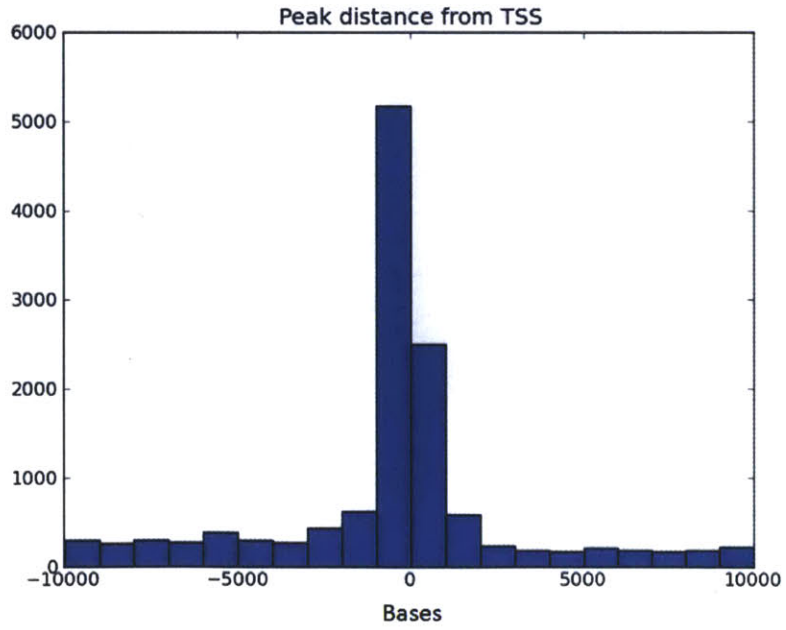
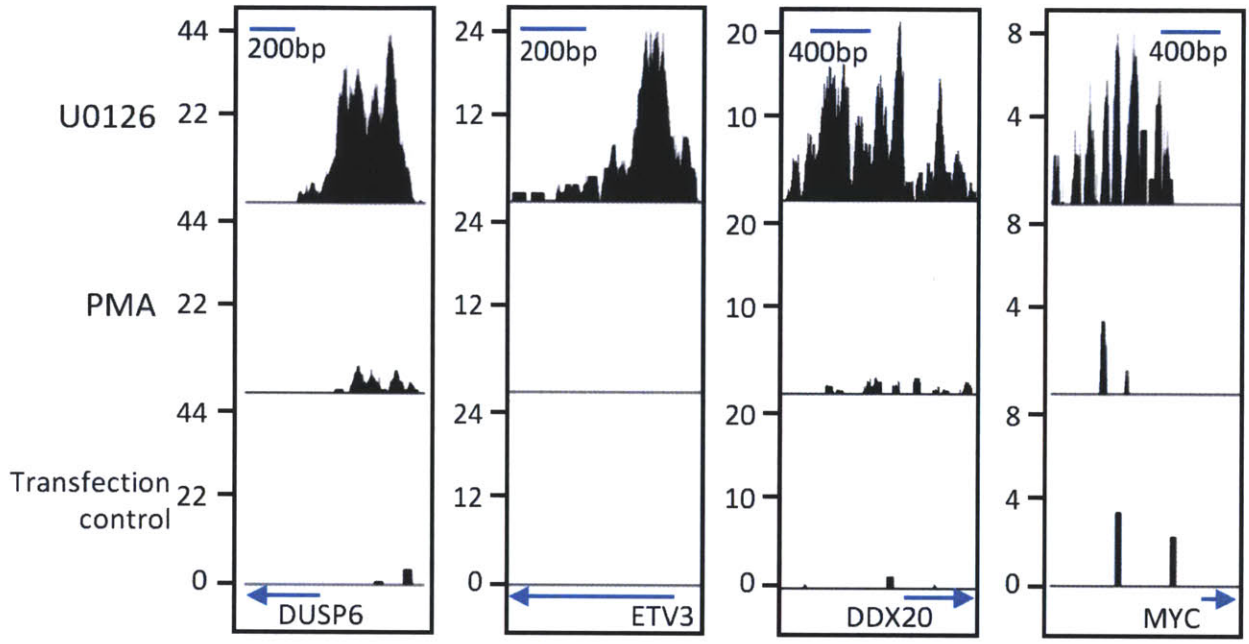


Figure 16: ETV3-binding sites identified by ChIP-seq are located almost entirely within 1 kilobase of transcriptional start sites.

Figure 17: Histograms of ChIP-seq peaks for HA-tagged ETV3 near several transcription start sites in 293T cells. MYC is a reported target for ETV3 but the MAPK phosphatase DUSP6, ETV3 itself, and its binding partner DDX20 show much stronger binding. In each case binding is completely abrogated by activation of ERK signaling.



binding sites and found that GGAA sequences characteristic of ETS-family proteins were very strongly enriched (Fig. 18)[79].

Because phosphorylation by ERK eliminates the repressive activity of ERF we hypothesized that it might have a similar effect on ETV3. Intriguingly, 20 minute stimulation with PMA led to an almost complete loss of DNA binding (Fig. 18). ChIP-PCR was used to confirm binding sites and showed strong phosphorylation-dependent binding near *ddx20* and the MAPK phosphatase *dusp6*, and weak binding near *myc* (Fig. 19). The weaker binding near *myc*

Figure 18: The consensus sequence for ETV3-binding determined by THEME. The core GGAA sequence is common to all ETS-domain proteins.

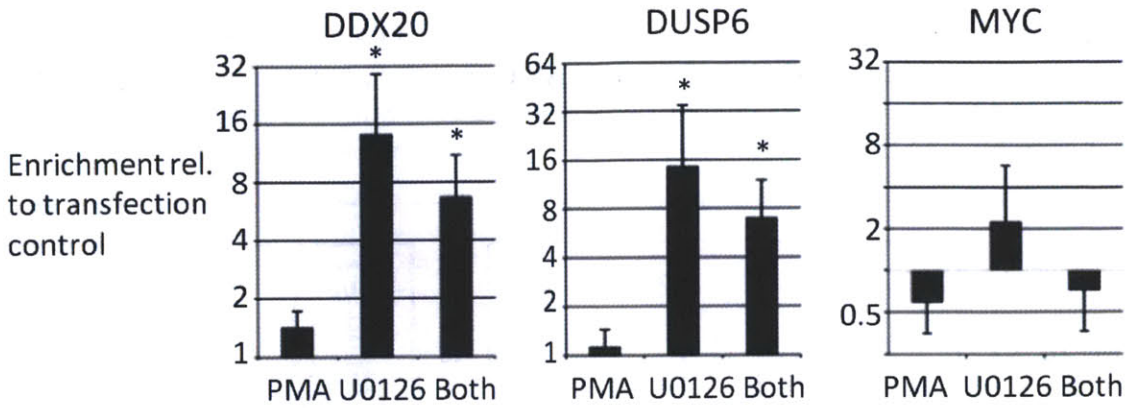


Figure 19: HA-tagged ETV3 was expressed in 293T cells and analyzed by ChIP-qPCR. Treatment with PMA abrogates binding of ETV3 to DNA, and treatment with U0126 partly blocks that effect. The same trend is observed for the MYC gene but with much smaller effect size. We obtained the same result with two different PCR primers for the MYC promoter. N=6, * $p < 0.05$ relative to DNA-bind in PMA-treated cells.

may reflect the relatively lower signal observed by ChIP-seq. We verified by Western blot that the HA antibody captures ETV3 in these ChIP experiments regardless of phosphorylation state.

To determine whether phosphorylation of ETV3 is directly responsible for eliminating DNA binding it was necessary to mutate key serine residues to alanine. Since the protein contains at least eight ERK phosphorylation sites we decided to prioritize by aligning ETV3 to ERF (there is considerable data available about specific phosphorylation sites on ERF), and by looking at sequence conservation between mouse and human forms of ETV3. We selected serines at positions 139, 159, 245 and 250 in human ETV3 because they are homologous to ERK target residues in ERF that affect DNA-binding and are conserved between mouse and human forms of the protein. We also mutated serine 29 because it is proximal to the DNA-binding domain. Converting serine 29 or 139 to alanine reduces the PMA-induced gel shift by Western blot, and the combination of 139, 159, 245 and 250 reduces the gel shift by slightly more (Fig. 20). This suggests that 29 and 139 are probably the major phosphorylated positions, with a smaller contribution from the other residues.

We used ChIP followed by quantitative PCR to test the effect of the quadruple S to A 139/159/245/250 mutation on ERK-responsive DNA binding. The ETV3 quadruple mutant binds DNA similarly to the wild-type form, it maintains DNA binding after PMA treatment at a level not statistically different from U0126-treated cells, and its DNA-binding remains higher than the wild-type form treated with PMA (Fig. 21).

As a control to verify that phosphorylation does not simply lead to a reduced protein level we compared protein stability in 293T cells treated with cycloheximide and PMA or cycloheximide and PMA+U0126. We found that ETV3 was constitutively unstable with a half-life of approximately 2 hours, and that phosphorylation has no effect on protein stability (Fig. 22).

STAT3 has been shown to transiently up-regulate ETV3 in macrophages co-stimulated with IL10 and lipid polysaccharide[80]. In the same work, over-expression of ETV3 was shown to repress NF-κB p65 (RelA) in COS cells. We identified ETV3 binding sites near many genes

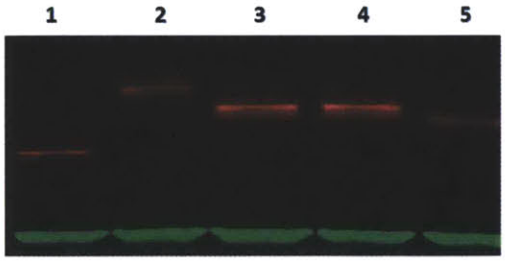


Figure 20: Retarded migration of ETV3 phosphosite mutants expressed in 293T cells (Red – ETV3, green – β-actin. Lane 1) ETV3 with no treatment, 2) ETV3 with PMA treatment, 3) ETV3 S29A with PMA, 4) ETV3 S139A with PMA, 5) ETV3 S139/159/245/250A with PMA.

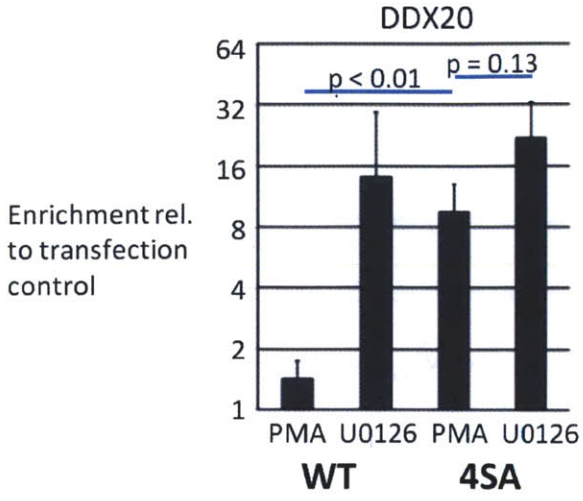


Figure 21: Mutation of the four serines at 139, 159, 245 and 250 prevents ERK from abrogating DNA binding by ETV3. N=4.

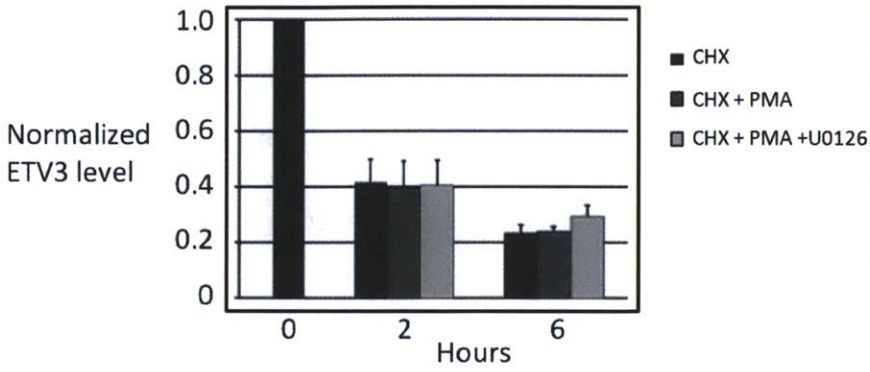


Figure 22: ETV3 is unstable in cycloheximide-treated cells 293T cells with half-life of about 2 hours, but protein stability is not affected by ERK activity.

involved in NF-κB signaling (*nfkb2*, *nkap*, *nkiras1*, *nkiras2*, *nfkbil1*, *trib3* and a weak signal near *rela*). This set includes both positive and negative regulators of NF-κB activity, suggesting that inactivation of ETV3 may be an important step in priming the pathway to respond to other signals. ETV3 targets span a wide range of other pathways and processes including ribosomal components and members of the spliceosome (Bonferroni-correct p-value $< 10^{-4}$ using DAVID)[81,82], and important targets in p53 signaling and cell cycle control (notably *tp53*, *chek1*, *cdkn2a*, *e2f4*, *ccng2* and *mdm4*).

Discussion

ETV3 was the only ERK target in our experiment that was identified with phosphorylation outside the canonical S/TP motif. We found that ETV3 is heavily phosphorylated by ERK2 on at least 8 serine residues, including three non-canonical positions. Although ETV3 has previously been reported to be unaffected by ERK activity[76], we found that its DNA-binding activity was rapidly abrogated following phosphorylation by ERK1/2. To characterize this effect we used ChIP-SEQ to identify several thousand binding sites for ETV3 across the genome, most of which contain an ETS consensus sequence, and all of which are lost following ERK1/2 activation. ETV3 target genes including *etv3*, *ddx20*, and *dusp6* provide negative feedback regulation of ETV3 production and activity. Negative feedback along with constitutive instability may serve to tightly regulate ETV3 expression. In this model

phosphorylation by ERK2 relieves repression by ETV3, allowing activation of cell cycle control genes including *myc*, components of the NF-κB pathway, and genes required for mRNA processing and translation. This concerted effect may prime cells to respond to additional stimuli, such as IL10 in bone-marrow derived macrophages[80]. Once ERK2 activity ceases, newly translated ETV3 can rapidly down-regulate its target genes (Fig. 23), providing a temporal burst of transcriptional activity following ERK activation. We believe that previous work did not observe an interaction between ERK and ETV3 because experiments focused on time-scales of several hours, enough time for repression by ETV3 to be reestablished. Unlike ERF, a related ETS repressor factor, loss of DNA-binding by ETV3 does not appear to result from sequestration in the cytoplasm. Although we have not observed phosphorylation of the ETS domain itself, phosphorylation regulates many ETS domain-containing factors through

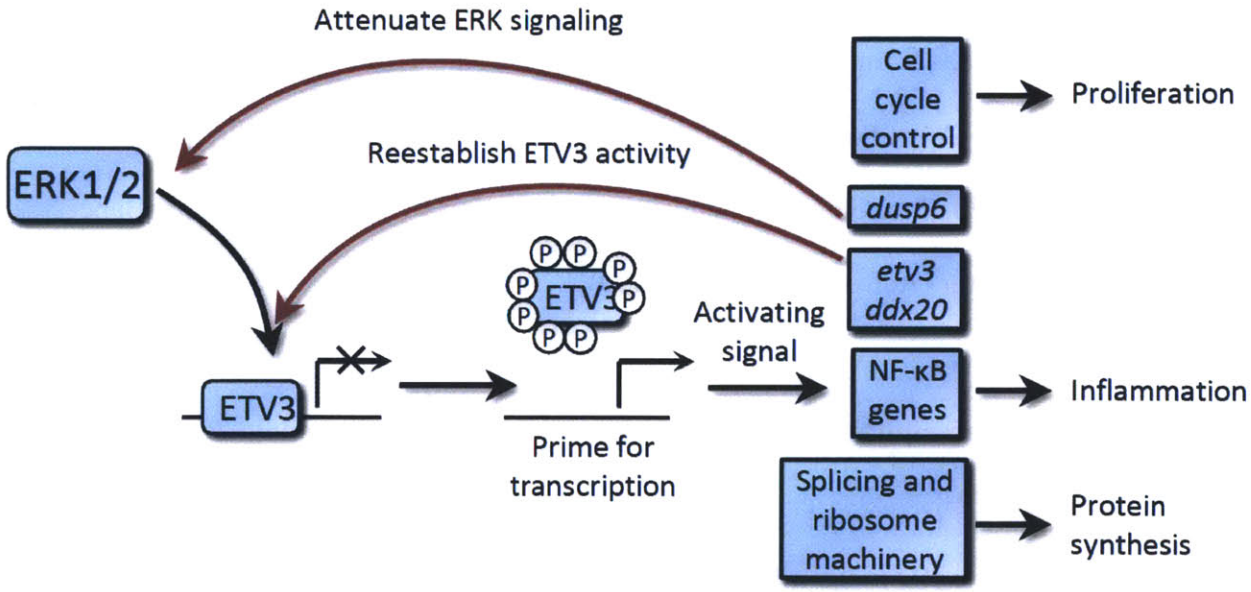


Figure 23: A model for the function of ETV3 biological activity. In its basal state ETV3 represses a wide range of genes including cell cycle proteins and protein metabolism. When ETV3 is phosphorylated it stops repressing target genes, making them accessible to activating signals. ETV3 targets include negative feedback through repression of ETV3 and DDX20.

regulation of protein-protein interactions or auto-inhibitory domains[83,84].

Interestingly, loss of heterozygosity has recently been observed associated with the *etv3* gene in B-cell chronic lymphocytic leukemia and follicular lymphoma, suggesting that ETV3 functions as a tumor suppressor in these cell types[85]. Given the thousands of ETV3-phosphorylation-dependent targets, ETV3 may thus function as a master regulator downstream of ERK activation in some cell types, capable of modulating transcription of genes spanning a broad range of cellular activities.

Methods

Antibodies and reagents

Antibodies for anti-HA Alexa488 conjugate, goat anti-rabbit Alexa488 conjugate, PMA and recombinant ERK2 kinase were from Cell Signaling Technology. Normal rabbit IgG for immunofluorescence was from Santa Cruz Biotechnology. Anti-ETV3 antibody was from Sigma-Aldrich. Anti-HA antibody for ChIP was from Abcam. HA-tagged human ETV3 was synthesized by GeneWiz Inc. Genes for mammalian expression were cloned into the pBabe-puro-IRES-EGFP vector (AddGene plasmid 14430)[73].

Chromatin immunoprecipitation (ChIP) and high-throughput sequencing

Three 15-cm plates of HEK 293T cells were transfected with HA-tagged human ETV3 (20 μg plasmid, 60 μL FuGene HD in 0.5 mL PBS) for two days to a density of 2.5 x 10^7 cells per plate. Cells were treated for 20 minutes with U0126 (10 μM) and/or PMA (100 nM). A control experiment was conducted with non-transfected cells. The ChIP protocol was modified from methods previously described[86,87]. Briefly, frozen pellets of formaldehyde-fixed cells were

resuspended and lysed. The cells were then sonicated using a Misonix Sonicator 3000 at 33W for twelve cycles of a 20 second pulse followed by a 60 second pause. Lysate was centrifuged at 14,000 rpm for 10 minutes at 4°C to remove impurities. The lysate was incubated with 50 µL of Dynal Protein G magnetic beads bound to 10 µg of antibody for 15 hours at 4°C. The purified DNA was prepared for sequencing on a Beckman Coulter SPRI-TE following manufacturer's instructions. The seq-prepped DNA was PCR amplified using Illumina primers for 18 cycles. Samples were sequenced using Illumina Solexa Genome Analyzers 2.0. Reads were mapped to the hg18 reference genome using the GERALD alignment software package. Uniquely mapping reads were run through the MACS v1.4.0beta peak calling software using default parameters. Reads from 293T cells transfected with ETV3-HA and treated with U0126 were used as the foreground and untransfected 293T cells as the background. Resulting peaks were filtered for *p*-value < 1e-10. Peaks were mapped to UCSC Known Genes if they fell within a 10 kb window around a gene's transcription start site. ChIP-Seq data has been uploaded to the NCBI Sequence Read Archive (accession SRA029365.1).

ChIP-qPCR

8 x 10^6 293T cells were prepared and processed as described for ChIP-Seq experiments, except that only 5 ug of antibody were used. Samples were analyzed by real-time PCR on a Roche Lightcycler 480 system using Roche SYBR Green I PCR master mix. Primers for *ddx20* were 5' GAGGCGGAGATACGAACT TG 3' and 5' TACCACATTGGCTGGTGTGT 3', for *dusp6* 5' GCTGGAACAGGTTGTGTTGA 3' and 5' AAGTGCCCTGGT TTATGTGC 3', and for *myc* 5' CCAACAAATGCAATGGGAGT 3' and 5' CCAGAGTCCCAGGGAGAGTG 3'. Mutations in the ETV3-pBabe-IRES-EGFP vector were introduced using Stratagene QuikChange.

Immunofluorescence

Approximately 10^5 per cm^2 of DLD1 or HEK 293T cells were thoroughly dispersed and plated overnight on coverslips. 293T cells were then transfected with HA-tagged mouse ETV3 in the pBabe-IRES-EGFP vector (AddGene vector 14430) using FuGene HD (Roche) according to the manufacturer's instructions and grown for one more day. Cells were treated as described with U0126 (10 μM) and/or PMA (100 nM), fixed in 4% formaldehyde for 15 minutes and washed 3 times with PBS. Cells were blocked for one hour with 5% goat serum (DLD1) or 5% mouse serum (293T) in PBS with 0.1% Triton-X. DLD1 cells were incubated overnight at 4°C with anti-ETV3 at 8 μg/mL in PBS with 1% BSA 0.1% Triton-X, then washed three times with PBS and incubated for 1 hour with anti-rabbit Alexa488 secondary antibody at 1:1000 dilution with goat anti-rabbit Alexa488 conjugate. Following antibody incubations slides were washed twice with PBS, treated with DAPI in PBS for 15 minutes, washed once with PBS, treated with phalloidin conjugated to Alexa647 (Invitrogen), washed twice with PBS, and mounted on slides using ProLong Gold (Invitrogen). Images were captured on a DeltaVision Spectris microscope from Applied Precision and processed by software deconvolution.

ETV3 binding motif analysis

To identify an optimal ETV3 binding motif, the most enriched unrefined ETS family motif (ELK-1) from TRANSFAC found under the U0126-treated ChIP-Seq peaks was used as the initial hypothesis. The expected motif strength score for ELK-1 was calculated by scanning a randomly sampled set of background sequences that match the length, GC content, and distance from TSS distributions of the U0126 peaks. In this dataset, peak strength as determined by p-

value positively correlates with motif strength, and the *p*-value cutoff of 1e-25 was identified as the threshold below which the U0126 peaks contain stronger motifs than the expected background. The ELK-1 motif was refined by THEME using the peaks with *p*-value < 1e-25 to arrive at an ETV3 motif. See THEME citation for more information on the motif refinement procedure.

4. Quantitative rewiring of the ERK1/2 pathway in colon cancer driven by activated K-Ras

Returning to the NIH Provocative Question about oncogene addiction, the methods developed in chapter 2 can also be used to examine pathway rewiring in the context of an oncogenic signal. If we can discover how pathways downstream of major oncogenes change during carcinogenesis then we may be able to identify new therapeutic targets and mechanisms of resistance to kinase inhibitors. Two major oncogenes are constitutively active KRAS and BRAF. Activating mutations in KRAS are especially common cancers of the lung, colon and pancreas, although they can occur in a very wide range of cancer types. KRAS signals through several pathways, most notably PI3K/AKT and ERK1/2 (Fig. 24). There are ongoing efforts to treat KRAS-mutant

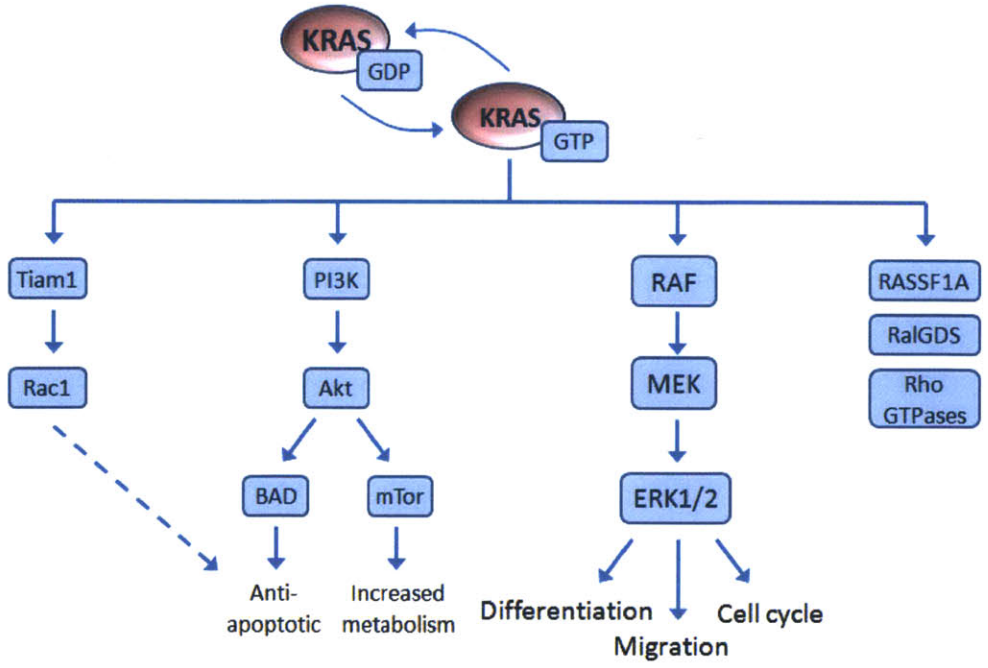


Figure 24: Pathways regulated by KRAS signaling.

cancer by simultaneously inhibiting PI3K and MEK, but there is not currently any targeted treatment in clinical use. Activation of BRAF occurs in almost two-thirds of melanomas and in a smaller fraction of colon cancers and other cancer types. As discussed in the introduction, the first targeted inhibitor for activated BRAF has recently entered clinical use for treating advanced melanoma.

ERK signaling controls a range of downstream transcriptional targets including feedback regulators of ERK activity (both phosphatases that deactive ERK and scaffolds that target its activity)[57,88]. We therefore hypothesized that constitutive signaling may lead to system-wide rewiring of the ERK signaling cascade. To investigate this possibility we chose to use AS-ERK2 and SILAC to both identify and quantify ERK2 substrates in cellular systems carrying activating mutations in the MAPK pathway.

A variety of cell culture models systems are available for studying oncogenic signals. For example, Bert Vogelstein generated a series of paired colon carcinoma cell lines in which one of each pair carries activated KRAS (DLD1 and HCT116 cell lines) or BRAF (VACO432 and RKO cell lines)[89]. We have also received Caco2 colon carcinoma cells with inducible BRAF V600E from Dr. Tilman Brummer at the Zentrum für Biosystemanalyse in Frieburg, Germany.

We have used the paired DLD1 cells lines to identify ERK2 substrates in the context of colon cancer, and to compare the levels of those substrates in the presence and absence of activated KRAS. These data indicate wide-spread changes in ERK signaling as a result of oncogene activation, especially in the areas of mRNA splicing and components of the nuclear pore. We have also identified four members of the mixed-lineage leukemia (MLL) family of histone 3 lysine 4 (H3K4) methyltransferases that are phosphorylated by ERK, and we are in the

process of exploring how those phosphorylation sites contribute to regulation of H3K4 methylation.

Quantitative identification of AS-ERK2

To label substrates relevant in colon cancer we began with the four cell lines provided by Bert Vogelstein. The DLD1 and HCT116 lines carry activating mutations in KRAS, and the RKO and VACO432 lines carry activated BRAF. Each of the four is paired with an isogenic line in which the oncogene has been removed by homologous recombination. We used a retroviral vector to express WT and AS-ERK2 in each of the eight lines, selected for stable expression, and analyzed the level of ERK expression by Western blot (Fig. 25) (VACO432 cells did not survive selection for ERK2 expression, perhaps because of a toxic interaction with the MAPK activating signal from BRAF V600E). Neither of the activated BRAF-expressing lines expressed a high level of exogenous ERK2, while HCT116 expressed a moderate level, and DLD1 expressed the largest amount. All subsequent identification experiments were performed in DLD1 cells.

To accomplish a quantitative comparison between ERK2 substrates in parental DLD1

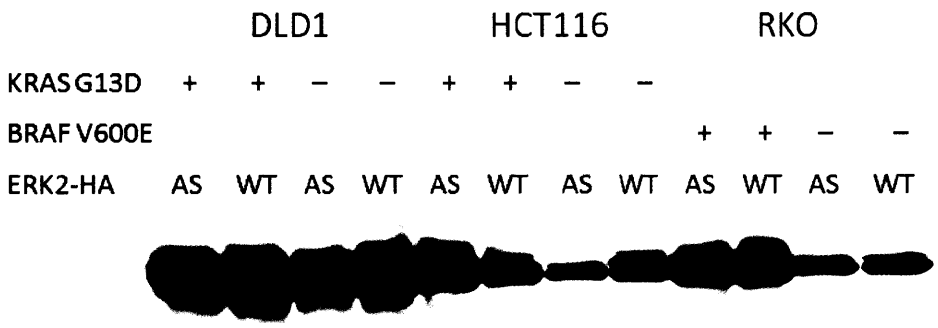


Figure 25: Blot for HA tag in colon cancer cell lines transfected with AS-ERK2 or WT-ERK2. Only the DLD1 cells expressed similar high levels of ERK2 with and without activated oncogene.

cells and DLD1 cells without activated KRAS (DKO) we labeled both cell lines using SILAC. Light label was used for DLD1 cells expressing AS-ERK2, the medium label was used for DLD1 cells expressing WT-ERK2, and the heavy label was used for DKO cells expressing AS-ERK2. Labeling reactions were conducted as described in the previous chapter, with the cells serum-starved for 12 hours and treated with PMA for 2-3 minutes prior to lysis. Western blot for alkylated thiophosphate showed strong labeling in the

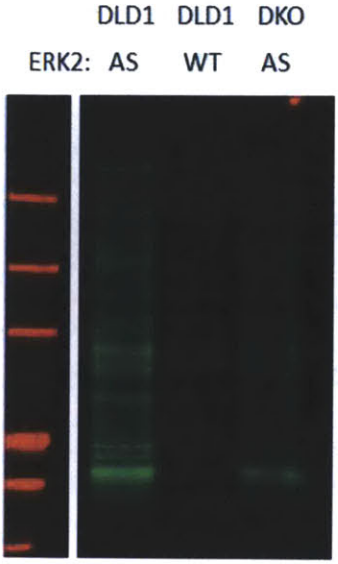


Figure 26: Thiophosphate labeling in lysates from DLD1 and DKO cells expressing AS or WT-ERK2 as indicated. There is more labeling visible in the DLD1 cells, consistent with the broadly higher in ERK2 substrate levels detected in thiophosphate capture experiments.

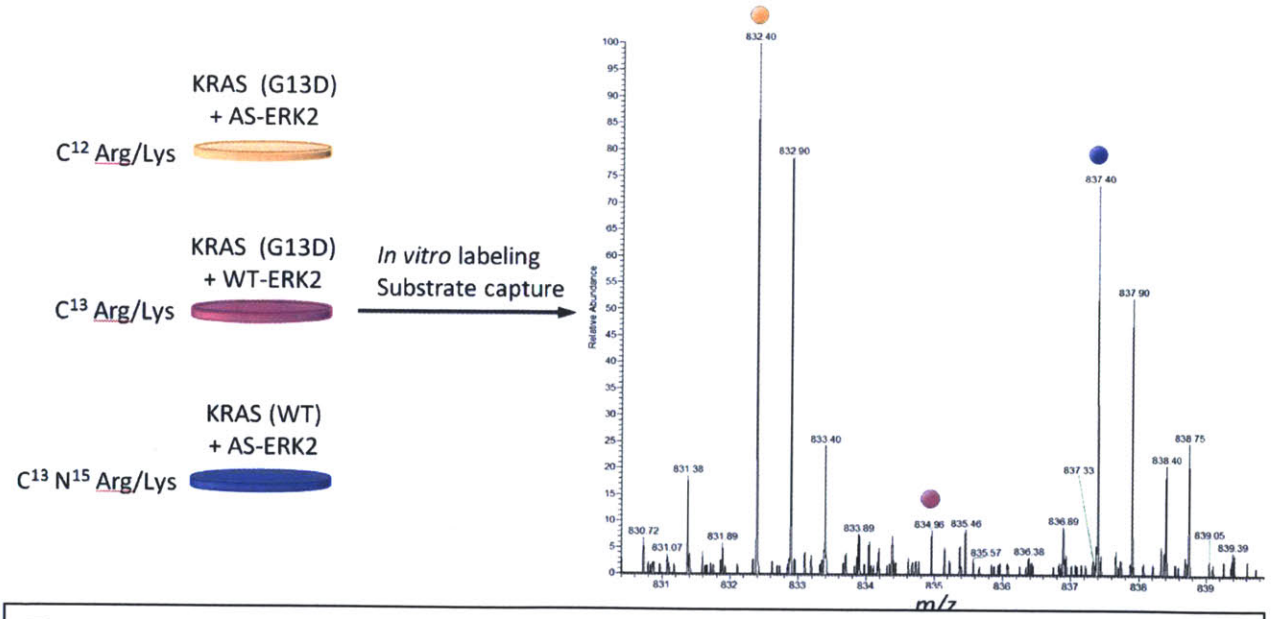


Figure 27: 3-way SILAC allows quantitative analysis of ERK2 substrates in different cell types. This example spectrum shows labeling of the methyltransferase MLL2 in DLD1 cells with or without activated KRAS. Once normalized to protein loading this spectrum indicates that MLL2 is about 40% higher in DLD1 cells with activated KRAS compared with DKO.

AS-ERK2 expressing lysates, with slightly more labeling in the cells with activated KRAS (Fig. 26).

The quantitative substrate capture experiment was performed as described in the previous chapter. The experiment was conducted twice, and analysis of additional replicates is underway with assistance from postdoctoral associate Dr. Aaron Gajadhar. Quantitative data for each phosphorylated peptide was normalized using non-phosphorylated peptides in the same LC-MS/MS experiment, and peptides strongly enriched in AS-ERK2 conditions were identified as ERK2 substrates (Appendix 4). Figure 27 on the previous page shows a representative SILAC MS1 mass spectrum for MLL2.

It is important to note that quantitative data here does not necessarily reflect the degree of phosphorylation *in vivo*. Because these substrates are identified following *in vitro* kinase reactions the data are likely to reflect total protein levels, although it is possible that they reflect

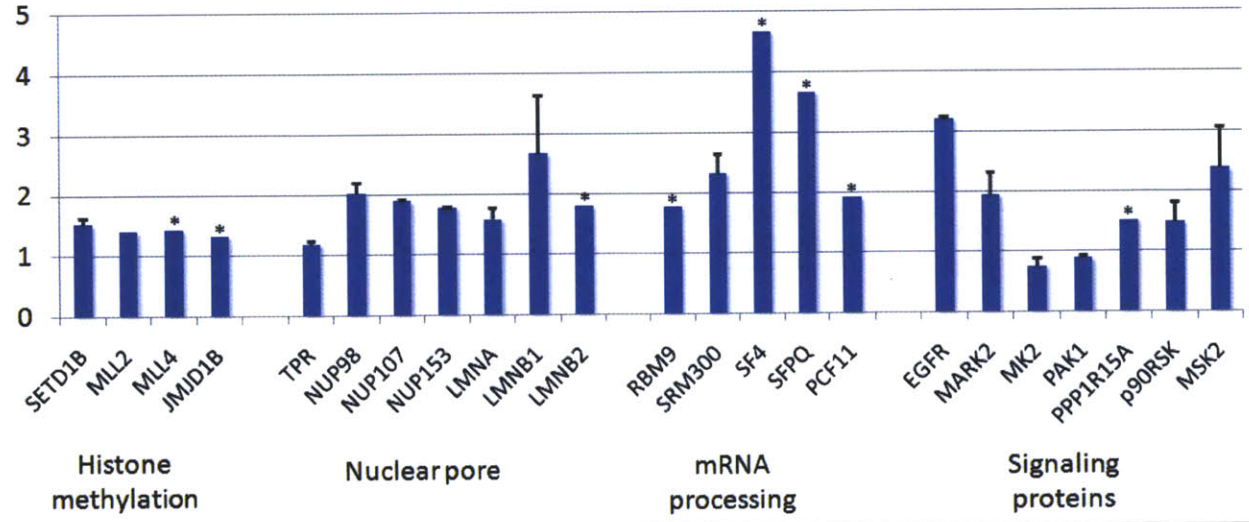


Figure 28: The ratio of selected ERK2 substrates detected in DLD1 with activated KRAS over DLD1 cells without activated KRAS. Functionally related proteins, such as members of the nuclear pore, tend to have similar values, while functionally unrelated proteins, such as kinases in different pathways, do not move together. Error bars – S.D. with N=2, * - detected in only one experiment.

formation of different protein complexes prior to cell lysis. We plan to determine expression for a subset of these proteins by quantitative Western blotting, and to examine mRNA levels using ChIP-Seq. The targets identified in these experiments had considerable overlap with targets in the previous chapter, although approximately 30% had not been reported in the literature or in the previous experiment.

Many of the ERK substrates detected in this experiment can be clustered into groups based on the molecular function. For example, splicing factors FOX2, SF4, SFPQ, as well as splicing coactivator SRm300 and mRNA processing protein PCF11 were all detected (Fig. 28). These are all novel ERK targets except for FOX2, which was identified in the previous chapter. Phosphorylation of all of these mRNA processing proteins by AS-ERK2 is greater in DLD1 relative to DKO cells. Similarly, phosphorylation of components of the nuclear pore, all well-known ERK targets, are uniformly increased by activated KRAS. In contrast, kinases and phosphatases to not move together, perhaps because they have distinct biological functions despite their related enzymatic activities.

Intriguingly, another of the NIH Provocative Questions is "How do changes in RNA processing contribute to tumor development?" (PQ-11) The discovery of several ERK targets involved in mRNA splicing, all of which are up-regulated by activated KRAS, suggests that a combination of ERK signaling and changes in the splicing machinery could drive changes in mRNA processing. To investigate this possibility we have prepared mRNA from DLD1, DLD1 treated with U0126, and DKO cells. In collaboration with the Fraenkel Lab we are conducting high-throughput sequencing of these samples to measure gene expression and global exon utilization. At the time of writing these samples are queued for analysis at MIT's BioMicro Center.

One especially interesting group of novel ERK2 targets is mixed-lineage leukemia 2 and 4 (MLL2 and MLL4), as well as SETD1B. These three proteins are all SET-domain containing histone lysine methyltransferases specific for the lysine at position 4 of histone 3 (H3K4). Phosphorylation of this family is uniformly up-regulated by about 40% by activated KRAS. To our knowledge this is the first direct connection between ERK activity and modification histones on lysine. Interestingly, the H3K9 demethylase JMJD1B was also identified as an ERK2 target. Methylation of H3K9 is associated with inactive chromatin regions. This may indicate a role for ERK in a coordinated program to control methylation at both K4 and K9 of histone H3.

H3K4 Methyltransferases

The MLL family of proteins consists of MLL1-4 and SETD1A and SETD1B (MLL5 contains a similar enzymatic domain but has a very different overall sequence). These proteins are large, ranging from 186 to 564 kD, and have many domains involved in interactions with chromatin, DNA, regulatory proteins, and other chromatin modifying enzymes. Structure/function relationships in MLL proteins are reviewed in [90].

The MLL family is subdivided into three groups with redundant function: MLL1/2, MLL3/4, and SETD1A/SETD1B. They are all specific for H3K4 and capable of mono-, di- or tri- methylating that residue[91,92]. All the MLL proteins are synthesized as a full-length protein and then cleaved by taspase to produce a functional dimer. The detailed enzymatic activities of these proteins are still being worked out but they are known to depend on the presence of a core complex consisting of RBBP5, ASH2L and WDR5.

We initially attempted to study MLL2 and MLL4 by Western blotting, IP and immunofluorescence (IF). All these experiments were limited by poor antibody specificity and

lack of available full-length cDNA clones. IF experiments in DLD1 cells suggested that MLL2 and MLL 4 are both primarily nuclear with one or two additional cytoplasmic foci for MLL4. These results are difficult to interpret because there are no antibodies for either protein validated for IF. IP of MLL2 followed by LC-MS/MS indicated that the antibody has stronger affinity for several other proteins than it has MLL2 itself. In any case, there was no visible change in IF following treatment with PMA or U0126 so we did not pursue this line of inquiry.

MLL1 is the best-studied member of the MLL family, and was originally identified as a fusion protein in mixed-lineage leukemia[93]. Many fusion partners of MLL1 have been identified in mixed-lineage leukemia, suggesting that it has serves a conserved function in cancer regardless of the fusion partner (reviewed in [94]). Because MLL1 is well-studied there are a range of high-quality reagents available, especially antibodies against both fragments of the active form, MLL1-null mouse embryonic fibroblasts, and cDNA clones.

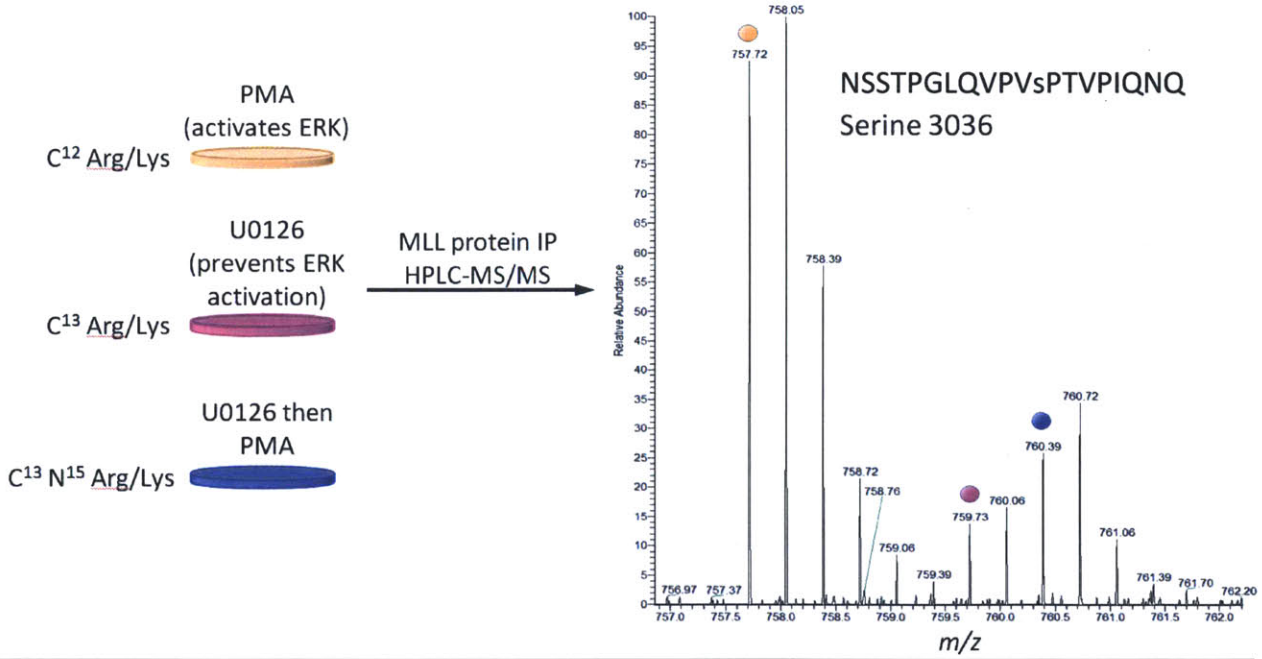


Figure 29: Quantitative analysis of MLL1 phosphorylation reveals MEK-dependent phosphorylation at residues 1837, 1845 and 3036. This is a very large protein and we expect that there are additional target sites that are not amenable to mass spectrometry.

To determine whether MLL1 is also an ERK target we prepared HEK293T cells with light, medium and heavy SILAC labels. We treated the cells with PMA, U0126 or U0126 then PMA for 30 minutes, immunoprecipitated MLL1, and either digested the IP directly or purified the protein by SDS-PAGE. In each case phosphorylated peptides were enriched by batch-mode IMAC and identified by LC-MS/MS. MEK-dependent phosphorylation was identified at residues 1837, 1845 and 3036 (Fig. 29). A similar experiment pulling down MLL2 from SILAC-labeled DLD1 cells showed MEK-dependent phosphorylation at the same residue as identified with AS-ERK2.

Taken together these data indicate that MLL1 and MLL2 are direct ERK2 substrates in a biological context, and MLL4 and SETD1B are phosphorylated *in vitro*. Based on these results we hypothesized that phosphorylation by ERK may be a general mechanism regulating some aspect of MLL activity.

Phosphorylation of MLL proteins

Although MLL proteins are heavily phosphorylated in global phosphoproteomic screens very little is known about their kinases and phosphatases, or about the functional importance of phosphorylation. Phosphorylation has been shown to affect subcellular localization of MLL5 but that result is unlikely to be relevant here because of broad differences in the protein structures. A recent study identified serine 516 of MLL1 as a target of the ATM kinase in response to DNA-damage[95]. Phosphorylation blocks the association of MLL1 with the E3 ubiquitin ligase Skp2, and stabilizes MLL1 following DNA damage during S-phase, which prevents progression to mitosis. Cells in culture accumulated chromosomal abnormalities when this phosphorylation site

was mutated to alanine. This mechanism only functions during S-phase, while other mechanisms are responsible for controlling MLL1 levels through G2, M, G0 and G1.

Based on the function of serine 516 we hypothesized that phosphorylation by ERK may have a role in regulating MLL1 levels. DLD1 cells and 293T cells were synchronized with double-thymidine block and analyzed over a 16 hour period, or else synchronized with mimosine (G1/S transition) or nicodazole (G2/M), and treated with U0126 and/or PMA. Although there seemed to be a small increase in MLL1 by Western blot when DLD1 cells or HEK293T cells were treated with PMA the effect is not consistent and does have a clear dependence on cell cycle. Unlike ATM, we believe that ERK does not act primarily through control of MLL1 stability.

Global H3K4 trimethylation profiling

Trimethylation of H3K4 is associated with chromatin in an accessible state and tends to accumulate near sites of active transcription[96], while dimethylation indicates genes primed for transcription. H3K4 methylation recruits a variety of proteins containing PHD-finger domains that activate or repress transcription. Disregulation of H3K4 methylation, especially at *hox* genes, is one mechanism by which MLL fusion proteins cause mixed lineage leukemia[97].

Phosphorylation of MLL1 may be involved in targeting to specific genomic loci or in regulating its enzymatic activity. In either case ERK signaling downstream of activated KRAS should contribute to changes in H3K4 methylation. To explore this possibility we are collaborating with the Fraenkel Lab to conduct and analyze ChIP-Seq experiments targeting trimethylated H3K4. DLD1 cells, DKO cells and DLD1 cells treated for 12 hours with U0126 were crosslinked and subject to ChIP-Seq for trimethylated H3K4 as described in Chapter 3,

with ChIP for total H3 serving as the background control. Analysis presented here was provided by graduate student Chris Ng and technician Adam Labadorf in the Fraenkel Lab.

To examine the overall structure of H3K4 trimethylation, profiles from near transcriptional start sites were filtered for minimum tag count and grouped using K-means clustering. Figure 30 shows results with five clusters for the DLD1 cells expressing mutant KRAS. It is clear from visual inspection that there is a basic bimodal structure around transcription start sites, and much of the variability among genes is described by the relative strength of the 5' and 3' peaks, and by how far H3K4 trimethylation spreads in the 3' direction. The first four clusters are not significantly enriched for any meaningful set of GO terms, but the fifth cluster is enriched for a wide range of GO terms related to RNA metabolism (this cluster is defined by a strong 3' peak and wide spread in the 3' direction). In a similar analysis of H3K4 trimethylation in neuronal cell culture the equivalent cluster is statistically enriched for neuron-specific genes (personal communication, Chris Ng). This may suggest that the fifth cluster is associated with active transcription. We are in the process of testing this hypothesis by correlating cluster assignments with mRNA expression determined by RNA-Seq. The bottom right panel of Figure 30 shows the center-of-mass for H3K4 trimethylation near each transcription start site across all three conditions (filtered as described in the methods).

When the center-of-mass is considered across all genes in the DLD1, DKO and DLD1 with U0126 conditions there is a small overall 3' shift DLD1 relative to the other two conditions. (Fig. 31). This data does not indicate whether the effect is direct (*i.e.* ERK changes activity of methyltransferases) or whether it is a by-product of changes in gene expression. Figure 32 shows trimethylation profiles for several genes with especially clear 3' shift when ERK is activated. The similar effect between loss of KRAS activity and direct inhibition of MAPK

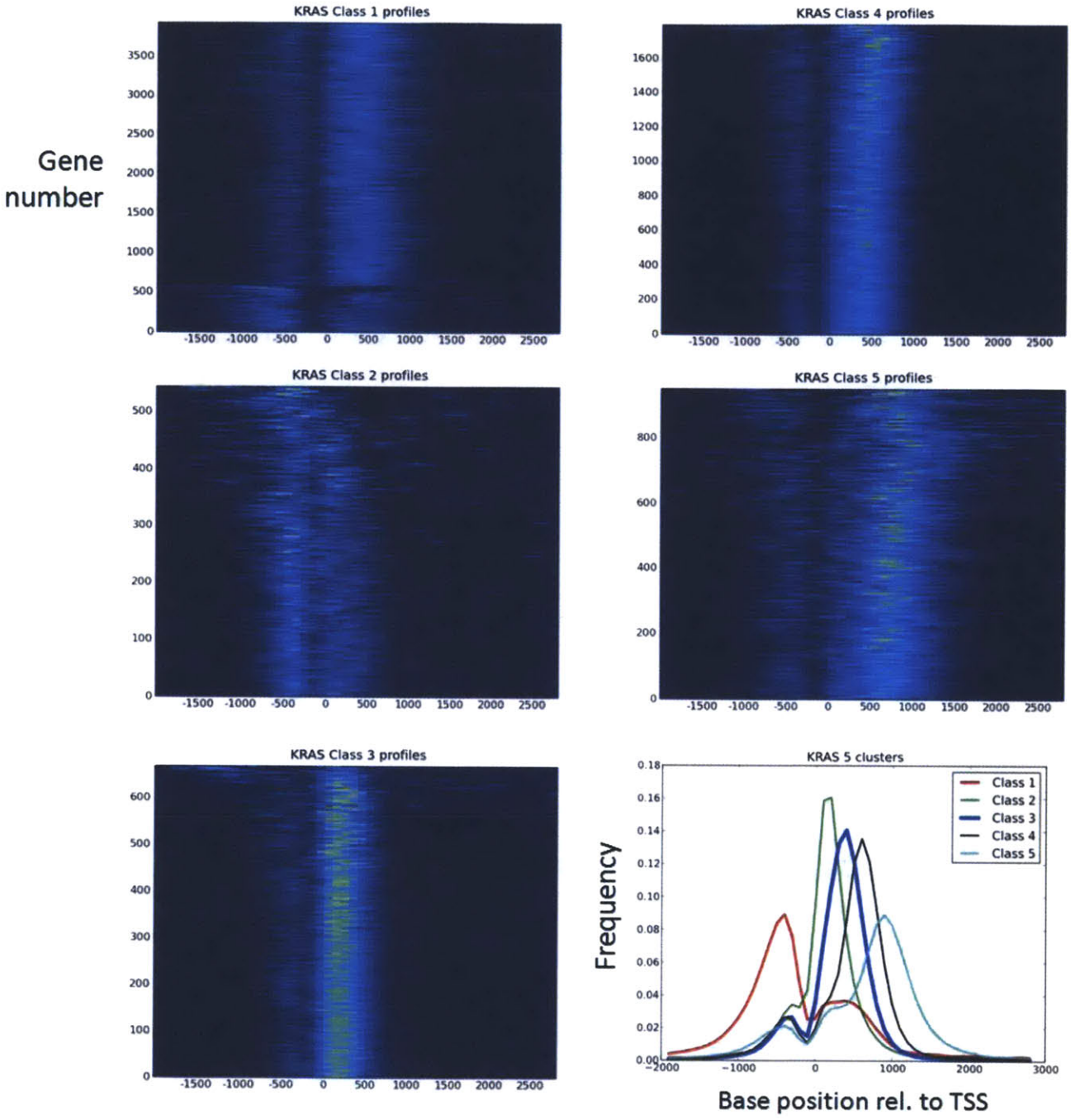


Figure 30: H3K4 trimethylation profiles from transcription start sites (TSS) in DLD1 cells grouped by K-means with five clusters. The x-axis indicates position around the annotated TSS (positioned at 0). Each cluster shows a distinct relationship between the 5' peak (left) and the 3' peak (right). The bottom right panel shows the frequency distribution of center-of-mass for each cluster, showing that each cluster is defined by the distance of 3' spreading.

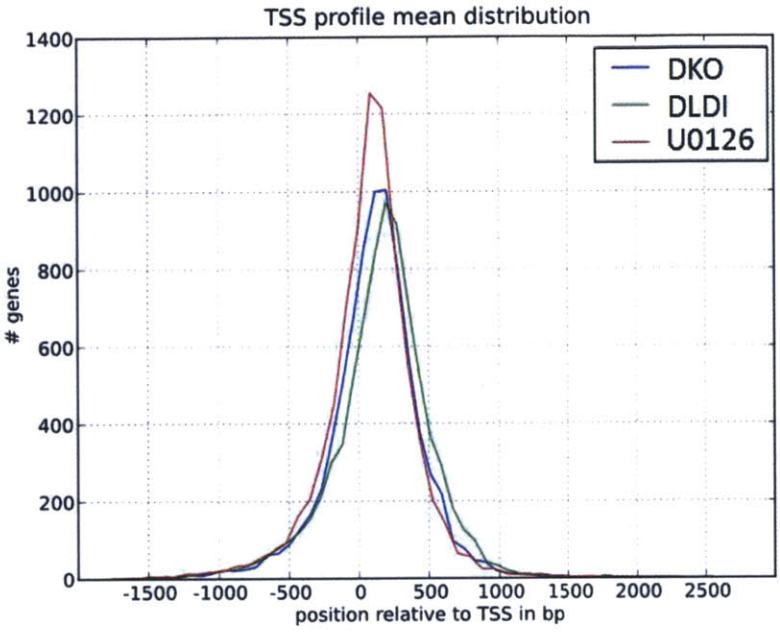


Figure 31: Center-of-mass for H3K4 trimethylation near all transcription start sites in the DLD1, DKO and DLD1 treated with U0126. TSS were filtered for minimum tag count. The DLD1 cells show a global 3' shift of about 50 bp relative to DKO and DLD1 + U0126, indicating that MAPK signaling may have a role in regulating the fine structure of H3K4 methylation.

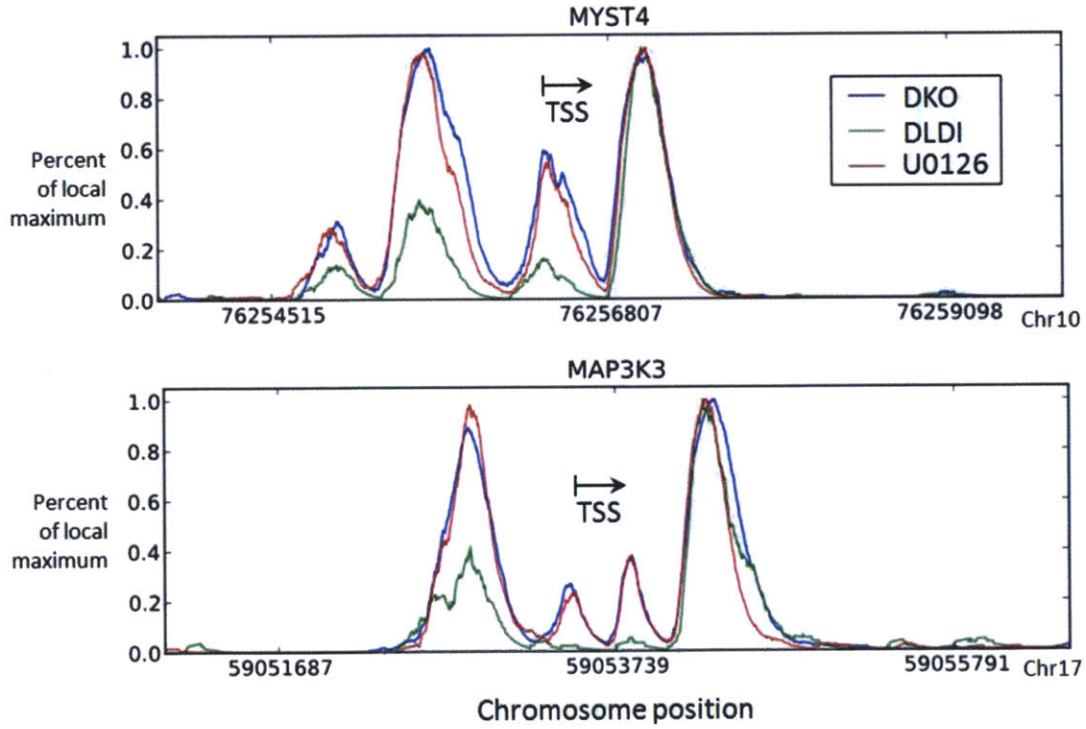


Figure 32: H3K4 trimethylation profiles for two genes selected for having a clear visual difference between DLD1 cells and the other two conditions. In both cases the DLD1 cells have more weight 3' to the transcription start site. The same shift is observed globally to a varying degree at different genes.

signaling suggests that ERK may be responsible for much of the difference between DLD1 and DKO cells. Based on these results we have additional experiments underway to determine if there is a direct mechanism by which ERK regulates H3K4 methylation or methyltransferase activity.

ChIP-Seq for ERK and MLL

It has recently been reported that ERK1/2 and other components of the activated MAPK signaling cascade associate with promoter regions for transcriptional targets of ERK signaling[98,99]. This led us to hypothesize that association of ERK with specific loci might be correlated with changes in H3K4 methylation. We have attempted test this hypothesis using ChIP-Seq from DLD1 cells under basal, U0126-treated, PMA-treated and EGF-treated conditions. Three different antibodies, including the antibody used in Nelson *et al.*[99] failed to capture any detectable DNA beyond the amount in a negative control. Despite the lack of detectable DNA we conducted sequencing on material captured by the antibody used in the original analysis. Results for all three conditions were identical to the IgG control. At this point we are unsure whether the original report is incorrect, if we failed to execute the protocol correctly, or if there may be a problem with the particular antibodies tested.

We have also prepared samples for ChIP-Seq analysis of MLL itself in DLD1 cells, DKO cells, DLD1 cells treated U0126, and DLD1 cells treated with PMA. A well-validated antibody for MLL1 is available and we are in the process of conducting ChIP-Seq to determine whether MLL1 targeting to genetic loci is affected by ERK activity. We will attempt to correlate any effect on MLL1 targeting with changes in H3K4 trimethylation and with mRNA levels. If it appears that ERK activity may be responsible for functional changes in MLL1 activity then phosphosite mutants will be generated and their targeting to DNA examined by ChIP-PCR.

Enzymatic function and protein interactions

In addition to high-throughput sequencing we are examining the biochemical function of MLL1 to determine whether it is regulated by phosphorylation. We have received a Flag-tagged MLL1 gene and MLL1-null mouse embryonic fibroblasts from the Hsieh Lab at Memorial Sloan Kettering Cancer Center. They generated MLL-null MEFs with a single FRT site recognized by the Flp recombinase. We are currently generating a stable MLL1-expressing cell line to investigate the effect of ERK activity on MLL1.

Phosphorylation may regulate protein-protein interactions with members of the core methyltransferase complex, with associated histone-modifying proteins like MYST1 acetyltransferase, or with DNA-binding proteins involved in targeting MLL1 activity. To address this question we will label MLL1-expressing MEFs with light and heavy SILAC labels, and label MLL1-null MEFs with a medium SILAC label. Light and heavy cells will be treated with PMA and U0126 respectively, and all three plates will be lysed in low-detergent hypotonic lysis buffer. Each lysate will be immunoprecipitated for FLAG-tag, washed, and eluted gently with Flag peptide to maintain protein complexes. We have verified that IP of MLL under these conditions also captures a significant amount of ASH2L, one of the core components of the methyltransferase complex. Purification of clean MLL complexes generally requires at least two enrichment steps so the eluate will be subjected to a secondary IP for MLL protein. Proteins captured in the second pull-down will be combined, run by SDS-PAGE, visualized by silver stain, and bands will be cut for analysis by LC-MS/MS. We aim to identify any proteins that interact selectively with MLL1 in the PMA or U0126-treated cells.

A second possibility is that phosphorylation regulates MLL1 methyltransferase activity. Methyltransferase activity assays based on colorimetric and fluorescent reporters are

commercially available. We will treat MLL1-expressing MEFs with various combinations of ERK-pathway activators and inhibitors, immunoprecipitate MLL1 complexes, elute gently with Flag peptide, and measure enzymatic activity toward recombinant histone H3.

Conclusions and ongoing work

In this chapter we identified additional ERK2 substrates in a colon cancer cell line, and generated preliminary data suggesting that many of these substrates are up-regulated following constitutive KRAS activation. Up-regulation seems to be coordinated among proteins with related function and additional replicates of the experiment are underway to establish statistics for determining which particular substrates are affected. In collaboration with Dr. Jamie Spangler, formerly of the Wittrup Lab at MIT, we found that EGFR surface expression is increased by about 3-fold in DLD1 cells expressing activated KRAS relative to no activated KRAS. This is exactly consistent with the mass spectrometry data. We also plan to follow-up by Western blotting to confirm quantitative results from the mass spectrometry experiments.

Some of the most intriguing novel substrates are involved in mRNA splicing and processing, and in methylation of H3K4. RNA-Seq experiments are underway to examine gene expression and exon utilization in DLD1, DKO, and DLD1 cells with U0126 treatment. We will collaborate with the Fraenkel Lab to determine if alternative splicing is significantly different among these conditions, and we will continue working with the Sharp lab to determine whether phosphorylation of splicing factors, and FOX2 in particular, may be responsible for some of the splicing changes seen in cancer.

We have shown that there is a broad 3' shift in H3K4 trimethylation near transcription start sites in DLD1 cells relative to cells without activated KRAS or cells in which the ERK

pathway is inhibited. The change itself is not surprising since these conditions probably have broad changes in gene expression. We hope however that ongoing experiments will to show that these changes correlate with MLL1 target genes or they depend on MLL1 phosphorylation by ERK. We are using ChIP-Seq to determine whether MLL1 associates with specific loci that have differential H3K4 trimethylation under each condition, and whether association with MLL1 is predictive of changes in the H3K4 trimethylation profile. This will allow us to generate hypotheses that can be tested using phosphorylation site mutants and ChIP-PCR for H3K4 methylation state at specific genomic loci.

Methods

Cell culture and SILAC labeling

Colon carcinoma cells were grown in DMEM supplemented by 10% FBS. AS-ERK2 and WT-ERK2 expressing lines were generated by retroviral transduction and selection in puromycin as described in Chapter 2. 3-way SILAC labeling and *in vitro* substrate labeling was conducted as described earlier.

Protein immunoprecipitation and phosphorylation analysis

SILAC-labeled DLD1 or 293T cells were grown in 10-cm plates, and each set of three SILAC-labeled cells was lysed in 1 mL of RIPA buffer supplemented with Halt protease and phosphatase inhibitors, PMSF and EDTA. Lysates were pre-cleared for 1 hour with 50 μL protein G-agarose. 10 μg anti-MLL1 antibody (ActiveMotif) was conjugated to 50 μL of protein G-agarose beads pre-conjugated to 10 μg mouse IgG bridging antibody (ActiveMotif) and washed three times with RIPA. Cleared lysate was incubated overnight with the sample, washed four times with RIPA, and eluted with either 8M urea for 30 minutes or by boiling in SDS-PAGE loading buffer. Immunoprecipitated protein was digested directly with trypsin or run by SDS-PAGE, visualized with silver stain (Pierce), and processed by in-gel digestion. Phosphorylated peptides were enriched by incubation with iron-NTA and analyzed by LC-MS/MS.

ChIP and high-throughput sequencing

Each ChIP experiment used 50 x 10^6 DLD1 or DKO cells grown in 15 cm plates. Cells were grown to approximately 80% confluence, treated with U0126 or PMA as described, cross-linked

processed for ChIP as described in Chapter 3. H3K4 trimethylation ChIP used 10 μg of Abcam #ab1012. Sequencing and analysis were performed as described in Chapter 3.

H3K4 trimethylation clustering

Short read sequences from DLD1, DKO, and U0126 treated DLD1 cells were aligned against the hg18 reference genome with BOWTIE. The genomic positions of each read were mapped by shifting the 5' ends of each read towards the 3' ends of their respective strands given a peak shift model built by MACS. To control for read complexity and PCR biases, a maximum of one mapped read was constrained for each genomic position. Genomic coordinates for annotated transcription start sites (TSS) of genes were obtained from RefSeq. This list of annotated of TSS's was filtered for genomic regions containing more than one TSS within 5000bp. For genes with more than one TSS, a primary TSS was empirically chosen based on the maximum number of reads +/- 1000bp from each TSS across all conditions. The genomic coordinates of the bound/unbound vector in each condition was mapped to -2kb/+3kb windows around the resulting list of 7884 primary isolated TSS's. Each TSS profile was smoothed by a moving average window of 200bp and binned in 100bp segments. Density profiles were calculated by dividing by the total number of bound coordinates at each TSS window. TSS profiles for each condition were concatenated into matrices and clustered by the k-means algorithm for k=5 using Euclidean distance as the metric.

5. Conclusions and future directions

Although chemical genetics is a powerful strategy for discovering signaling network topology without depending on existing knowledge or hypotheses there are a number of challenges and caveats to be considered. First, ATP-analogs are not readily taken up by cells, therefore labeling of kinase substrates typically occurs in cell lysates, either from cells expressing the AS-kinase or by addition of AS-kinase to the lysate. Since the reaction occurs outside the normal cellular environment loss of biological context, including subcellular localization, can lead to spurious interactions, while the dilution of scaffold and substrate molecules may cause some substrates to be missed. Second, bio-orthogonal ATP analogs are not perfectly specific for the AS-kinase, leading to background substrate labeling due to endogenous wild-type kinases in the cell lysate. Depending on the cellular context and the expression and activity level of the AS-kinase the level of non-specific labeling can make a substantial contribution to total labeling, significantly altering the results of the assay. It is important to include an appropriate negative control (*e.g.* replacement of the AS-kinase with equivalent amount of wild-type kinase). Third, many substrates are present at very low levels, such that identification is limited by the sensitivity of even the latest generation of mass spectrometers. Further improvements in the protocol for substrate enrichment combined with more sensitive instrumentation will greatly expand the range of applications for AS-kinases.

Extending this approach to other kinases will depend greatly on how amenable each one is to this approach. There is no guarantee that every kinase will have good activity with γ-thiol analogs, or that there will be an N6-substituted analog appropriate for every AS-kinase. It also requires the kinase to have high activity in an *in vitro* context, and for substrates to be

sufficiently abundant and suitable for identification by LC-MS/MS. My own experience analyzing substrates of AS-JNK1 in collaboration with Dr. Norman Kennedy is that substrates of JNK1 are more difficult to detect, probably because of low abundance or weaker *in vitro* activity. After extensive optimization we identified only TPR as a substrate of JNK1, although at a residue different from that targeted by ERK2. Dr. Sandra Morandell in the group of Professor Michael Yaffe at MIT has had a similar experience with kinase MK2. Her data suggest that the kinase has poor activity with γ-thiol ATP (personal communication).

The multitude of new ERK2 substrates identified in this work has revealed many downstream pathways that may be important in both normal biology and in disease. For example, the transcriptional repressor ETV3 is regulated by phosphorylation and targets thousands of genes. Since ETV3 has a role in regulating immune cell function this may be an important mechanism connecting growth factor signals to immune response. ETV3 is also widely expressed in cancer cells, and its inactivation may be an important process in development of some cancers.

Substrates involved in mRNA processing and histone modification are two of the most exciting areas for future work. The fact that signaling through the ERK pathway regulates many global regulators of other signaling processes underscores the interconnected nature of biological signaling processes. These types of highly connected networks are difficult to understand using reductionist approaches such as perturbation of individual proteins. Systems-level experiments and computational methods that span a range of signaling mechanisms will be necessary to elucidated the behavior of these networks.

Analysis of mRNA splicing in a variety of cellular systems is ongoing in collaboration with the Sharp and Fraenkel groups. We are also continuing to examine histone modification

through ChIP-SEQ and biochemical analysis of the MLL1 H3K4 methyltransferase. Despite these collaborations, most of the substrates identified here remain almost entirely unexplored. With such a wide range of biological processes represented in these datasets, I hope that the data presented here will be valuable for researchers in other areas of biology.

Ongoing work in with AS-kinases in the White Lab is now focused on analysis of AS-ERK1. Postdoctoral associate Aaron Gajadhar is preparing to compare the targets of ERK1 and ERK2 with the goal of understanding how such similar kinases mediate different biological effects.

The most striking result from studies with AS-kinases is that only a small fraction of kinase-substrate interactions appear already in the literature. It is humbling to realize that even such a basic biochemical process in cell signaling is so poorly understood, even ignoring the complexity of different cell types and biological context, and interactions between different types of regulation (signaling, transcription, splicing, epigenetics, etc.). Even with all the technological sophistication available today, developing a mechanistic system-level understanding of cell signaling will clearly require enormous effort in method development, high-throughput experiments, and biological characterization.

References

1. Hanahan, D. & Weinberg, R. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
2. Vang, T. et al. Protein tyrosine phosphatases in autoimmunity. *Annu Rev Immunol* **26**, 29-55 (2008).
3. Fröjdö, S., Vidal, H. & Pirola, L. Alterations of insulin signaling in type 2 diabetes: a review of the current evidence from humans. *Biochim Biophys Acta* **1792**, 83-92 (2009).
4. Karess, R., Hayward, W. & Hanafusa, H. Cellular information in the genome of recovered avian sarcoma virus directs the synthesis of transforming protein. *Proc Natl Acad Sci U S A* **76**, 3154-8 (1979).
5. Daley, G., Van Etten, R. & Baltimore, D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science* **247**, 824-30 (1990).
6. Aaronson, S. Growth factors and cancer. *Science* **254**, 1146-53 (1991).
7. Zhang, J., Yang, P. & Gray, N. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* **9**, 28-39 (2009).
8. Luo, J., Solimini, N.L. & Elledge, S.J. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* **136**, 823-37 (2009).
9. An, X. et al. BCR-ABL tyrosine kinase inhibitors in the treatment of Philadelphia chromosome positive chronic myeloid leukemia: a review. *Leuk Res* **34**, 1255-68 (2010).
10. Knight, Z.A., Lin, H. & Shokat, K.M. Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer* **10**, 130-7 (2010).
11. Paez, J.G. et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497-500 (2004).

12. Ciardiello, F. & Tortora, G. EGFR antagonists in cancer treatment. *N Engl J Med* **358**, 1160-74 (2008).

13. Herbst, R.S., Fukuoka, M. & Baselga, J. Gefitinib--a novel targeted approach to treating cancer. *Nat Rev Cancer* **4**, 956-65 (2004).

14. Davies, H. et al. Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-54 (2002).

15. Pollock, P.M. et al. High frequency of BRAF mutations in nevi. *Nat Genet* **33**, 19-20 (2003).

16. Nazarian, R. et al. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468**, 973-7 (2010).

17. Wagle, N. et al. Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J Clin Oncol* **29**, 3085-96 (2011).

18. Janes, K. et al. A high-throughput quantitative multiplex kinase assay for monitoring information flow in signaling networks: application to sepsis-apoptosis. *Mol Cell Proteomics* **2**, 463-73 (2003).

19. Krutzik, P., Crane, J., Clutter, M. & Nolan, G. High-content single-cell drug screening with phosphospecific flow cytometry. *Nat Chem Biol* **4**, 132-42 (2008).

20. Ciaccio, M., Wagner, J., Chuu, C., Lauffenburger, D. & Jones, R. Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nat Methods* **7**, 148-55 (2010).

21. Schmelzle, K. & White, F. Phosphoproteomic approaches to elucidate cellular signaling networks. *Curr Opin Biotechnol* **17**, 406-14 (2006).

22. Choudhary, C. & Mann, M. Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol* **11**, 427-39 (2010).

23. Grimsrud, P., Swaney, D., Wenger, C., Beauchene, N. & Coon, J. Phosphoproteomics for the masses. *ACS Chem Biol* **5**, 105-19 (2010).

24. Ficarro, S.B. et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* **20**, 301-5 (2002).

25. Moser, K. & White, F. Phosphoproteomic analysis of rat liver by high capacity IMAC and LC-MS/MS. *J Proteome Res* **5**, 98-104 (2006).

26. Olsen, J. et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-48 (2006).

27. Dephoure, N. et al. A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A* **105**, 10762-7 (2008).

28. Macek, B., Mann, M. & Olsen, J. Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu Rev Pharmacol Toxicol* **49**, 199-221 (2009).

29. Pan, C., Olsen, J., Daub, H. & Mann, M. Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics. *Mol Cell Proteomics* **8**, 2796-808 (2009).

30. Hunter, T. & Sefton, B. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci U S A* **77**, 1311-5 (1980).

31. Rush, J. et al. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* **23**, 94-101 (2005).

32. Zhang, Y. et al. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* **4**, 1240-50 (2005).

33. Moritz, A. et al. Akt-RSK-S6 kinase signaling networks activated by oncogenic receptor tyrosine kinases. *Sci Signal* **3**, ra64 (2010).

34. Blair, J. et al. Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat Chem Biol* **3**, 229-38 (2007).

35. Maly, D., Allen, J. & Shokat, K. A mechanism-based cross-linker for the identification of kinase-substrate pairs. *J Am Chem Soc* **126**, 9160-1 (2004).

36. Shah, K., Liu, Y., Deirmengian, C. & Shokat, K. Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc Natl Acad Sci U S A* **94**, 3565-70 (1997).

37. Bishop, A. et al. Design of allele-specific inhibitors to probe protein kinase signaling. *Curr Biol* **8**, 257-66 (1998).

38. Liu, Y., Shah, K., Yang, F., Witucki, L. & Shokat, K. Engineering Src family protein kinases with unnatural nucleotide specificity. *Chem Biol* **5**, 91-101 (1998).

39. Ventura, J. et al. Chemical genetic analysis of the time course of signal transduction by JNK. *Mol Cell* **21**, 701-10 (2006).

40. Das, M. et al. Suppression of p53-dependent senescence by the JNK signal transduction pathway. *Proc Natl Acad Sci U S A* **104**, 15759-64 (2007).

41. Burkard, M. et al. Chemical genetics reveals the requirement for Polo-like kinase 1 activity in positioning RhoA and triggering cytokinesis in human cells. *Proc Natl Acad Sci U S A* **104**, 4383-8 (2007).

42. Shah, K. & Shokat, K. A chemical genetic screen for direct v-Src substrates reveals ordered assembly of a retrograde signaling pathway. *Chem Biol* **9**, 35-47 (2002).

43. Habelhah, H. et al. Identification of new JNK substrate using ATP pocket mutant JNK and a corresponding ATP analogue. *J Biol Chem* **276**, 18090-5 (2001).

44. Niswender, C. et al. Protein engineering of protein kinase A catalytic subunits results in the acquisition of novel inhibitor sensitivity. *J Biol Chem* **277**, 28916-22 (2002).

45. Ubersax, J. et al. Targets of the cyclin-dependent kinase Cdk1. *Nature* **425**, 859-64 (2003).

46. Eblen, S. et al. Identification of novel ERK2 substrates through use of an engineered kinase and ATP analogs. *J Biol Chem* **278**, 14926-35 (2003).

47. Blethrow, J., Zhang, C., Shokat, K. & Weiss, E. Design and use of analog-sensitive protein kinases. *Curr Protoc Mol Biol* **Chapter 18**, Unit 18.11 (2004).

48. Blethrow, J., Glavy, J., Morgan, D. & Shokat, K. Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc Natl Acad Sci U S A* **105**, 1442-7 (2008).

49. Chi, Y. et al. Identification of CDK2 substrates in human cell lysates. *Genome Biol* **9**, R149 (2008).

50. Aouadi, M., Binetruy, B., Caron, L., Le Marchand-Brustel, Y. & Bost, F. Role of MAPKs in development and differentiation: lessons from knockout mice. *Biochimie* **88**, 1091-8 (2006).

51. Bost, F., Aouadi, M., Caron, L. & Binétruy, B. The role of MAPKs in adipocyte differentiation and obesity. *Biochimie* **87**, 51-6 (2005).

52. Dong, C., Davis, R.J. & Flavell, R.A. MAP Kinases in the Immune Response. *Ann. Rev. Immunol.* **20**, 55-72 (2002).

53. Govindarajan, a., Kelleher, R.J. & Tonegawa, S. A clustered plasticity model of long-term memory engrams : Article : Nature Reviews Neuroscience. *Nat. Rev. Neurosci.* **7**, 575-583 (2006).

54. Dhillon, A.S., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26**, 3279-3290 (2007).

55. Yoon, S. & Seger, R. The extracellular signal-regulated kinase: multiple substrates regulate diverse cellular functions. *Growth Factors* **24**, 21-44 (2006).

56. Lawrence, M.C. et al. The roles of MAPKs in disease. *Cell Res.* **18**, 436-442 (2008).

57. Kolch, W. Coordinating ERK/MAPK signalling through scaffolds and inhibitors. *Nat. Rev. Mol. Cell. Biol.* **6**, 827-837 (2005).

58. Lewis, T. et al. Identification of novel MAP kinase pathway signaling targets by functional proteomics and mass spectrometry. *Mol Cell* **6**, 1343-54 (2000).

59. Old, W. et al. Functional proteomics identifies targets of phosphorylation by B-Raf signaling in melanoma. *Mol Cell* **34**, 115-31 (2009).

60. Kosako, H. et al. Phosphoproteomics reveals new ERK MAP kinase targets and links ERK to nucleoporin-mediated nuclear transport. *Nat Struct Mol Biol* **16**, 1026-35 (2009).

61. Ong, S.E. & Mann, M. Stable isotope labeling by amino acids in cell culture for quantitative proteomics. *Methods Mol Biol* **359**, 37-52 (2007).

62. Allen, J. et al. A semisynthetic epitope for kinase substrates. *Nat Methods* **4**, 511-6 (2007).

63. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-67 (1999).

64. Camuzeaux, B. et al. p38beta2-mediated phosphorylation and sumoylation of ATF7 are mutually exclusive. *J Mol Biol* **384**, 980-91 (2008).

65. Roux, P.P. & Blenis, J. ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. *Microbiol Mol Biol Rev* **68**, 320-44 (2004).

66. Fritsche, L. et al. INSULIN-INDUCED SERINE PHOSPHORYLATION OF IRS-2 VIA ERK1/2 AND mTOR: STUDIES ON THE FUNCTION OF SER 675 AND SER 907. *Am J Physiol Endocrinol Metab* (2010).

67. Venables, J.P. et al. Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* (2009).

68. Schmelzle, K., Kane, S., Gridley, S., Lienhard, G.E. & White, F.M. Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* **55**, 2171-9 (2006).

69. Pullikuth, A. & Catling, A. Scaffold mediated regulation of MAPK signaling and cytoskeletal dynamics: a perspective. *Cell Signal* **19**, 1621-32 (2007).

70. De Fea, K. & Roth, R. Modulation of insulin receptor substrate-1 tyrosine phosphorylation and function by mitogen-activated protein kinase. *J Biol Chem* **272**, 31400-6 (1997).

71. Giovannone, B. et al. Two novel proteins that are linked to insulin-like growth factor (IGF-I) receptors by the Grb10 adapter and modulate IGF-I signaling. *J Biol Chem* **278**, 31564-73 (2003).

72. Schnidar, H. et al. Epidermal growth factor receptor signaling synergizes with Hedgehog/GLI in oncogenic transformation via activation of the MEK/ERK/JUN pathway. *Cancer Res* **69**, 1284-92 (2009).

73. Morgenstern, J.P. & Land, H. A series of mammalian expression vectors and characterisation of their expression of a reporter gene in stably and transiently transfected cells. *Nucleic Acids Res* **18**, 1068 (1990).

74. Brar, G.A. et al. Rec8 phosphorylation and recombination promote the step-wise loss of cohesins in meiosis. *Nature* **441**, 532-6 (2006).

75. Obenauer, J., Cantley, L. & Yaffe, M. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* **31**, 3635-41 (2003).

76. Hester, K.D. et al. Differential repression of c-myc and cdc2 gene expression by ERF and PE-1/METS. *Cell Cycle* **6**, 1594-604 (2007).

77. Sluss, H.K. & Davis, R.J. H2AX is a target of the JNK signaling pathway that is required for apoptotic DNA fragmentation. *Mol Cell* **23**, 152-3 (2006).

78. Klappacher, G.W. et al. An induced Ets repressor complex regulates growth arrest during terminal macrophage differentiation. *Cell* **109**, 169-80 (2002).

79. Macisaac, K.D. et al. A hypothesis-based approach for identifying the binding specificity of regulatory proteins from chromatin immunoprecipitation data. *Bioinformatics* **22**, 423-9 (2006).

80. El Kasmi, K.C. et al. Cutting edge: A transcriptional repressor and corepressor induced by the STAT3-regulated anti-inflammatory signaling pathway. *J Immunol* **179**, 7215-9 (2007).

81. Dennis, G. et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).

82. Huang, d.W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).

83. Verger, A. & Duterque-Coquillaud, M. When Ets transcription factors meet their partners. *Bioessays* **24**, 362-70 (2002).

84. Oikawa, T. & Yamada, T. Molecular biology of the Ets family of transcription factors. *Gene* **303**, 11-34 (2003).

85. Green, M.R. et al. A new method to detect loss of heterozygosity using cohort heterozygosity comparisons. *BMC Cancer* **10**, 195 (2010).

86. Odom, D.T. et al. Core transcriptional regulatory circuitry in human hepatocytes. *Mol Syst Biol* **2**, 2006.0017 (2006).

87. Harbison, C.T. et al. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-104 (2004).

88. Amit, I. et al. A module of negative feedback regulators defines growth factor signaling. *Nat Genet* **39**, 503-12 (2007).

89. Shirasawa, S., Furuse, M., Yokoyama, N. & Sasazuki, T. Altered growth of human colon cancer cell lines disrupted at activated Ki-ras. *Science* **260**, 85-8 (1993).

90. Cosgrove, M.S. & Patel, A. Mixed lineage leukemia: a structure-function perspective of the MLL1 protein. *FEBS J* **277**, 1832-42 (2010).

91. Patel, A., Dharmarajan, V., Vought, V.E. & Cosgrove, M.S. On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem* **284**, 24242-56 (2009).

92. Dou, Y. et al. Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121**, 873-85 (2005).

93. Rowley, J.D. et al. Mapping chromosome band 11q23 in human acute leukemia with biotinylated probes: identification of 11q23 translocation breakpoints with a yeast artificial chromosome. *Proc Natl Acad Sci U S A* **87**, 9358-62 (1990).

94. Slany, R.K. The molecular biology of mixed lineage leukemia. *Haematologica* **94**, 984-93 (2009).

95. Liu, H. et al. Phosphorylation of MLL by ATR is required for execution of mammalian S-phase checkpoint. *Nature* **467**, 343-6 (2010).

96. Karlić, R., Chung, H.R., Lasserre, J., Vlahovicek, K. & Vingron, M. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A* **107**, 2926-31 (2010).

97. Chi, P., Allis, C.D. & Wang, G.G. Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nat Rev Cancer* **10**, 457-69 (2010).

98. Mikula, M. & Bomsztyk, K. Direct recruitment of ERK cascade components to inducible genes is regulated by heterogeneous nuclear ribonucleoprotein (hnRNP) K. *J Biol Chem* **286**, 9763-75 (2011).

99. Nelson, J.D., LeBoeuf, R.C. & Bomsztyk, K. Direct recruitment of insulin receptor and ERK signaling cascade to insulin-inducible gene loci. *Diabetes* **60**, 127-37 (2011).

Appendix 1 - AS-ERK2 sites in 3T3-L1 cells

			WT:AS Ratio †							
Gene	**Site**	**Known substrate**	**1**	**2**	**3**	**4**	**5**	**6**	**Mean**	**# Obs**
XRN2	439		0.12		< 0.06		0.15		0.11	3
AKAP12	273			0.16	< 0.14	0.07			0.12	3
AIM1	655			< 0.36		0.08			0.22	2
AHNAK	5567		0.07	0.06	0.04	0.05	0.08	0.10	0.07	6
	4890		0.33	0.10	0.05	0.04	0.09		0.12	5
	695				< 0.08				0.08	1
	2985		0.27		0.04	0.07		0.10	0.12	4
	217		< 0.18	0.11	< 0.07	0.09	0.20	< 0.16	0.13	6
	4879			0.05	0.11	0.04	0.07		0.07	4
	4906			0.10	< 0.11	0.28			0.16	3
AHNAK2	510		< 0.65		< 0.14				0.14 *	2
RAI14	249				< 0.17				0.17	1
AHCTF1	1954				0.19				0.19	1
ATG2B	398			0.14					0.14	1
MGEA5	709		< 0.32		0.20				0.32	1
CDC42EP1	197			0.27					0.27	1
	113		< 0.15	0.27	0.12		< 0.08	0.05	0.14	5
CDC42EP2	101			0.09	0.07	0.05	0.11	0.06	0.08	5
CEP170	562**					< 0.22			0.22	1
DOCK1	1772		0.13	0.09	0.06	< 0.07			0.09	4
DENND4C	1270			< 0.24					0.24	1
DLG5	1279		< 0.24	< 0.17	< 0.17	< 0.10			0.17	4
DLG7	328		0.25		0.13	0.13			0.17	3
DNAJC30	137				< 0.12				0.12	1
DYNC1I2	81**	Kosako H, et al. (2009) Nat St	< 0.07		0.05				0.07	1
DYNC1LI1	516	Kosako et al.		< 0.27					0.27	1
EGFR	695	Gonzalez FA, et al. (1991) J B	0.26		0.17				0.21	2
ERF	529	Le Gallic L, et al. (1999) Mol (	0.08	< 0.02	0.03	0.04	0.09	0.09	0.06	6
ETV3	139			0.16	0.17				0.16	2
EIF4ENIF1	114			< 0.22					0.22	1
GTSE1	460				< 0.21				0.21	1
GJA1	255					< 0.26			0.26	1
GLI2	885		< 0.07	0.04	0.04	0.04			0.05	4
GIGYF2	30		0.08	0.05	0.05	0.05	0.05		0.06	5
HNRNPH1	104**			0.24	< 0.45	0.18			0.29	3
BAT2	609**			< 0.60					0.60	1
	1002			< 0.12					0.12	1
	1217				0.18				0.18	1
C21orf70	38			< 0.19	< 0.11	< 0.09			0.13	3
Fam195b	30		< 0.06	< 0.31	< 0.24	< 0.35	0.08	0.14	0.17 *	5
KIAA0284	968			< 0.33	< 0.28			< 0.09	0.24	3
	529			0.13				< 0.25	0.19	2
FAM65A	351			< 0.02	< 0.04	< 0.05	0.09	< 0.16	0.07	5

IRS2	653		< 0.22	< 0.19					0.21	2
KSR1	260	Cacace AM, et al. (1999) Mol Cell Biol		< 0.14			< 0.12		0.13	2
LARP1	278		0.05		0.04	< 0.10	0.09		0.07	4
	751		< 0.27			0.07			0.17	2
LMNA	22	Peter M, et al. (1992) Eur J Biochem			0.07		0.11	0.17	0.11	3
MAPKAPK2	320	Engel K, et al. (1995) J Biol Chem			< 0.20		0.32		0.26	2
MAP1A	2420		0.04	< 0.03	0.06				0.04	3
	2044			0.22		0.10			0.10	2
MAP1B	1784			0.07	0.07	0.09			0.08	3
MAST2	1337				< 0.31				0.31	1
MYBBP1A	1280				< 0.36				0.36	1
MYO9B	1317				< 0.15		0.25		0.20	2
NID1	331		0.29	0.12	< 0.03	0.04	0.12		0.12	4
TPR	2110	Eblen S, et al. (2003). J Biol C	< 0.45	0.05					0.05 *	2
	2149	Eblen, et al.		0.09	0.04	0.07	< 0.20	0.06	0.07 *	5
	2131	Eblen, et al.		0.07	0.05	0.05	0.07	0.07	0.06	5
NCOR2	1884						< 0.24		0.24	1
NUP153	610	Kosako et al.			< 0.11		0.15		0.13	2
PHF2	651**				< 0.27		0.21	< 0.22	0.23	3
PHLDB1	463**		0.05		0.05	0.06	0.12	0.04	0.07	5
	445				0.08				0.08	1
PDXDC1	691		< 0.26	0.05	< 0.09	0.13		< 0.15	0.13	5
RAPH1	746		< 0.46		< 0.10				0.28	2
	611		< 0.38		< 0.12				0.25	2
RIPK3	399				< 0.19				0.19	1
RELL1	262		< 0.44						0.44	1
RTN4	156				< 0.08				0.08	1
RAI1	1051					0.07	0.07		0.07	2
RPS3	221	Kim HD, et al. (2005) Biocher	0.32	0.29	0.36	0.67		0.52	0.41	5
ARHGEF17	695		< 0.34	0.23	0.15				0.24	3
FOX2	65			0.41					0.41	1
RBPMS	118		< 0.27	< 0.22	< 0.05		< 0.12	< 0.27	0.19	5
SAFB2	387						< 0.16	< 0.25	< 0.16	2
STK10	950			< 0.17	< 0.05		< 0.08	< 0.06	0.09	4
SSBP3	360			< 0.13	< 0.11				0.12	2
SORBS3	648	Mitsushima M, et al. (2004). J Biol Chei		< 0.23	0.15	0.09		0.14	0.12	3
SNX2	104		0.09	0.21	0.10	< 0.16	< 0.06		0.12	5
STMN1	25	Marklund U, et al. (1993) J Bi	0.08	0.05	0.07	0.06	0.10	0.14	0.09	6
	38			0.25	0.24		0.50		0.33	3
SUPT5H	664		0.24	0.27	0.16				0.22	3
SUPV3L1	725				< 0.06	0.15			0.11	2
SNED1	1239					< 0.06	< 0.11	< 0.21	0.13	3
TNKS1BP1	1025**			0.18	< 0.15				0.17	2
	131		0.10	0.03	0.04	0.06	0.07	0.07	0.06	6
TBC1D23	514			0.12	< 0.11				0.12	2
TFPI	210		< 0.31		< 0.08	0.13			0.18	3
TANC1	1561			< 0.10	< 0.23				0.17	2

TPX2	369			0.33				0.33	1
TCEB3BP1	529			< 0.10				0.10	1
TGFB1I1	124	< 0.33	< 0.28			< 0.15		0.21 *	3
TRP53BP2	591			< 0.23				0.23	1
UBAP2	964			< 0.09				0.09	1
UBAP2L	844	< 0.17		0.03	0.03	0.04	0.06	0.06	5
UDPGDH	474		0.08	0.09	0.08	0.19	< 0.11	0.11	5
WIZ	288			< 0.13				0.13	1

† Ratios are calculated by averaging SILAC peak intensities from three MS spectra around the maximum chromatographic elution. Missing values were rounded up to the noise floor (estimated manually for each spectrum). Ratios are normalized to a panel of non-specific signals shared across the experiments. Less-than symbols indicate that the light ion was below the detection limit so the reported value is rounded up the noise/signal ratio.

* Excludes quantitation close to the detection limit

** MS/MS spectrum does not uniquely determine the phosphorylated position. Phosphorylation is assigned to the residue consistent with the ERK recognition motif (only one residue is followed by proline)

Appendix 2: additional ERK2 phosphorylation sites identified on IRS2, ETV3, and CDC42EP1

a. Insulin receptor substrate 2 (GI: 124487073)

```
   1 masaplpgpp asaggdgpnl nnnnnnnnhs vrkcgylrkq khghkrffvl rgpgtggdea
  61 saaggsppqp prleyyesek kwrskagapk rvialdccln inkradakhk ylialytkde
 121 yfavaaeneq eqeqwyralt dlvsegrsge ggsgttggsc saslpgvlgg sagaagcddn
 181 yglvtpatav yrevwqvnlk pkglgqsknl tgvyrlclsa rtigfvklnc eqpsvtlqlm
 241 nirrcghsds fffievgrsa vtgpgelwmq addsvvaqni hetileamka lkelfefrpr
 301 sksqssqssa thpisvpgar rhhhlvnlpp sqtglvrrsr tdslaatppa akctscrvrt
 361 asegdggaag gagtaggrpm svagsplspg pvraplsrsh tlsagcggrp skvtlapagq
 421 alqhsrsmsm pvahsppaat spgslssssg hgsgsyplpp gshphlphpl hhpqgqrpss
 481 gsasasgsps dpgfmsldey gsspgdlraf sshrsntpes iaeTppardg sggelygyms
 541 mdrplshcgr pyrrvsqdqa qdldrglrkr tysltTparq rqvpqpssas ldeytlmrat
 601 fsqssqrlcp sfpasspkva ynpypedygd ieigshksss snlqaddqym pmTpqaalrs
 661 ggpnscksdd ympmSptsvs apkqilqprl aaalppsqaa vpappsqvqr tfpvnqqqyk
 721 asSpaesspe dsqymrmwcq sklsmenpdp kllpngdyln mspseagtag tppdfsaalr
 781 ggseglkgip ghcysslprs ykapcscsgd ndqyvlmssp vgrileeerl epqatpgagt
 841 fgaaggshtq phhsavpssm rpsaiggrpe gflgqrcrav rptrlslegl qtlpsmqeyp
 901 lptepkspqe yinidfgeaq trlsppappl lasaassssl lsasspassl qsqtpqtssd
 961 srqrsplsdy mnldfsspkS pkpstrsgdt vgsmdgllsp easspypplp prpstspssl
1021 qqplppapgd lyrlppasaa tsqgptagss mssepgdngd ytemafgvaa tppqpivapp
1081 kpegarvaSp tsglkrlslm dqvsqveafl qvsqppdphr gakviradpq ggrrrhsset
1141 fsstttvtpv spsfahnskr hnsasvenvs lrkssegsst lgggdeppts pgqaqplvav
1201 ppvpqarpwn pgqpgaligc pggssspmrr etsvgfqngl nyiaidvrqe qqslaqsqpq
1261 pgdknswsrt rslggllgtv ggsgasgvcg gpgtgalpsa styasidfls hhlkeatvvk
1321 e
```

b. Ets variant gene 3 (PE1, METS) (GI: 134031941)

```
   1 mkaqcsivek pegqqqyqfp dwaykaesSp qsrqiqlwhf ilellqkeef rhviawqqge
  61 ygefvikdpd evarlwgrrk ckpqmnydkl sralryyynk rilhktkqkr ftykfnfnk
 121 vmpnypfini rssqvvpqSa ppvptassrf hfppldshSp tqdvqpqrfs asslsasqpe
 181 Sqvttdrkve psdledqsas dwhrqmdfmp srnalgggav ghqkrkpdil lplftrpamy
 241 pdphSpfaiS pvpqrqqvln vpispalslt ptmfsyspsp glspftsssc fsfnpeemkh
 301 ylhsqacsvf nyhlsprtfp rypglmvppl qcqmhpeeps qfsiklqppp aqrknrerve
 361 sreeavrqsv paSapvpsri kvepatekdp dslrqstqqk eeqtqevdsi rsrtieegkq
 421 tqfahpSptw psvsistpsd epleqtedse drsarepgvp ekkedalmpp klrlkrrwnd
 481 dpearelnkt gkflwnqaqp qqlattataa ada
```

c. CDC42 effector protein 1 (GI: 21312428)

```
   1 mpqpqqqtqa ptmslqklSp vqwvssshqk rrltadmisp plqdfrhtmh vqrqqdvfqd
  61 tsflsnhqqr sqnthrsprs flarklqqvr rvqvpprrma spaapspapp piSpiiknai
 121 slpqlnqaty dslvmqklsf dstpasstdg hsqyqlesqf ctisrlprve khsnrdrdrd
 181 pdhsqdreqs sfpsepTpnp elrrsdslls frfdldlgps llsellgvms lseapaaetp
 241 vptanppapa anpaptakpp ahaittldav tslpasavts lpapavassp srghfpngvt
 301 svlgpaaeak pspvgegpqv psnmafdrhq aswqasrasw qasrasrhyt emdarrelag
 361 vlpqvhgswe slnedwsapp asvrapvpts vqvnafefad aeeddevkv
```

Recombinant proteins were phosphorylated by ERK2 in *in vitro* kinase reactions, digested with trypsin or chymotrypsin, phosphorylated peptides enriched by IMAC and identified by HPLC-MS/MS.

Underline indicates peptides observed in phospho-mapping experiments and phosphorylated residues are highlighted. IRS2 and CDC42EP1 were analyzed with tryptic digest; ETV3 was analyzed with trypsin and with chymotrypsin. Phosphorylation sites detected on IRS2 are: T524, T576, T653, S675, S723, S980, S1089. Phosphorylation sites on ETV3 are: S29, S139, S159, S181, S245, S250, S373, S427. Phosphorylation sites on CDC42EP1 are: S19, S113, T197. None of these sites were detected in mock kinase reactions.

Appendix 3: Phosphorylation sites identified by phospho-motif antibodies for PXpSP and PXpTP

Experiment 1: SILAC-labeled 3T3-L1 fibroblasts with two conditions

Light SILAC cells are treated 20 minutes with EGF (100 ng/mL)

Lys6/Arg6 SILAC cells are pretreated with U0126 (5 µM for 10 minutes) then EGF

Known ERK1/2 substrate

Phosphorylation reduced by U0126 treatment (10 min prior to stimulation)

Protein	Site	EGF m/z (Light)	EGF+U0126 m/z (Heavy)	Ratio *	p-value†
chromodomain helicase DNA binding protein 8	2520	924.92	927.930	1.18	1.00
cytosolic phospholipase A2, group IVA	505	789.39	791.397	0.73	1.37E-07
dedicator of cytokinesis 7	862	926.76	928.767	0.92	0.09
dynein cytoplasmic 1 light intermediate chain 1	207	783.01	787.023	0.97	0.33
Ets2 repressor factor	529	738.33	741.340	0.77	8.79E-06
eukaryotic translation elongation factor 1 gamma	386	923.08	925.087	1.05	0.80
FCH and double SH3 domains 2	681	776.35	779.360	0.67	7.70E-12
HECT, UBA and WWE domain containing 1	3927, 3930	926.44	927.945	0.71	6.97E-09
hypothetical protein LOC105428 isoform 2	163	853.65	856.660	0.43	4.33E-44
mutS homolog 6	14	689.67	693.683	0.97	0.31
nuclear receptor coactivator 3	261	858.41	861.420	0.90	0.04
polynucleotide kinase 3'- phosphatase	187	636.28	639.290	0.92	0.09
rearranged L-myc fusion sequence	521	718.01	722.023	0.92	0.10
RNA binding motif protein 9 isoform 1/2	65	875.71	877.717	0.51	1.27E-29
zinc finger protein 828	394	679.83	682.840	0.88	0.01

Experiment 2: SILAC-labeled 3T3-L1 fibroblasts with two conditions

Light SILAC cells are treated 20 minutes with EGF (100 ng/mL)

Lys6/Arg6 SILAC cells are pretreated with U0126 (10 µM for 10 minutes) then EGF

Protein	Site (Mouse)	EGF m/z (Light)	EGF+U0126 m/z (Heavy)	Ratio *	p-value†
chromodomain helicase DNA binding protein 8	2520	924.92	927.930	1.18	0.99
DENN/MADD domain containing 2A	397	1044.98	1047.990	0.84	0.01
dynein cytoplasmic 1 light intermediate chain 1	207	783.01	787.023	0.91	0.10
Ets2 repressor factor	529	738.33	741.340	0.76	5.67E-05
homeodomain interacting protein kinase 1	1200	782.87	785.880	0.77	1.32E-04
melanoma antigen family D, 1	166	877.71	879.717	0.73	4.65E-06
nucleoporin 153	594	700.28	703.290	0.78	3.42E-04
phospholipase C, delta 3	31	1151.06	1154.070	0.60	3.80E-13
Ran-binding protein 10	365	779.33	782.340	1.09	0.88
rearranged L-myc fusion sequence	521	1076.51	1082.530	1.03	0.67
WD repeat domain 4	397	627.77	630.780	0.98	0.38
zinc finger protein 828	394	679.83	682.840	0.92	0.11

* U0126 with EGF divided by EGF alone. Data integrated over each 2 minute elution profile and weighted average of significant isotope peaks. Values are normalized to a panel of peptides identified in supernatant from the immunoprecipitation.

† One-tailed z-score with null hypothesis that U0126 does not reduce phosphorylation of the peptide (assuming ratios are distributed as a log-normal distribution, standard deviation derived from 20 supernatant peptides from 7 proteins)

Appendix 4: AS-ERK2 substrates identified in DLD1 cells and ratio of DLD1 to DKO

		Normalized Ratios: KRAS/WT	
Protein	Peptide	1	2
XRN2	DQPAFTPSGIL**pT**PHALGSR		1.84
AIM1L	DVPSPGGLSAPS**pS**PR		1.69
ADD1	AAVVT**pS**PPPTTAPHK	1.21	2.12
ADAR	TAESQ**T**P**p**TPSATSFFSGK	3.29	1.37
AKAP11	EDSGLF**pS**PIR		
ANKHD1	VFLQGPAPVG**pT**PSFNR	1.04	1.51
ANKLE2	TSLPL**pS**PVK	2.43	
BAG3	SSTPLH**pS**PSPIR	1.36	1.75
KIAA1543	APSPSGLM**pS**PSR	4.61	
CDC42EP3	ANSTSDSVFTETP**pS**PVLK	1.11	1.33
CDCA5	IVAHAVEVPAVQ**pS**PR	1.50	1.51
	THSVPA**pT**PTSTPVPNPEAESSSK	1.73	1.85
CCDC6	LDQPVSAPP**pS**PR	1.06	
CNKSR1	SPSLSLAPL**pS**PR		2.62
FAM120A	GVIS**pT**PVIR	1.36	
COBL	GPPS**pT**PVPTQTQNPESR		1.67
CTTN	TQTPPV**pS**PAPQPTEER	2.08	1.66
CXADR	TPQ**pS**PTLPPAK	2.67	2.15
CRB3	VPP**pT**PNLK		1.83
DOCK1	FSVSPSSPSSQQTPPPV**pT**PR	0.70	1.61
DOCK5	APEPDLM**pS**PTR		1.58
DENND4C	LF**pS**PVIAR		1.17
DLG7	TYQVTPM**pT**PR	1.46	1.74
DMAP1	DTIIDVVGAPL**pT**PNSR		2.44
EFHD2	RADLNQGIGEPQ**pS**PSR		2.89
ELK1	DLELPL**pS**PSLLGGPGPER		
EGFR	ELVEPL**pT**PSGEAPNQALLR	3.20	3.24
ERF	GEGPGEAGGPL**pT**PR	2.00	2.19
EIF4G1	EAALPPV**pS**PLK	1.83	1.97
FAM122B	RIDFTPV**pS**PAPSPTR		1.91
GPKOW	EGVLPL**pT**AASTAPISFGFTR		1.90
GRIN1	SVATGPM**pT**PQAAAPPAFPEVR		1.09
GULP1	ITQVSAPPAGSM**pT**PK	3.95	
SETD1B	LDTPNSYGQGTPL**pT**PR	1.60	1.45
BAT2	EAPPPVLL**pT**PK	2.06	1.67
	ETPPNGNL**pS**PAPR	0.79	1.23
FAM122A	HGLLLPA**pS**PVR		2.16
C19orf21	SSTVATLQGTPDHGDPR**pT**PGPPR	1.36	1.63
C19orf55	AWVPPL**pT**PALR		1.64
C10orf47	MAGNEALSPT**pS**PFR	2.61	2.24
FAM195B	SPPSSSEIF**pT**PAHEENVR	2.16	2.03
C15orf39	LEPPL**pT**PR	0.89	1.05
KIAA1522	SGPQIL**pT**PLGDR	1.60	1.33

C12orf52	ALLWTPPP**pT**PR		1.36
C14orf43	SFELPPY**pT**PPPIL**SP**VR		1.27
ITIH2	FYNQVS**pT**PLLR	48.72	48.80
JMJD1B	LPQ**pT**PLDTGIPFPPVFSTSSAGVK		1.33
KIF14	TADMPL**pT**PNPVGR	1.55	1.41
LARP	FDGVEGPR**pT**PK	1.35	1.89
	SLPTTVPE**pS**PNYR	1.88	2.03
	GPGLLE**pS**PSIFNFTADR		
LMNA	SGAQASSTPL**pS**PTR	1.71	1.46
	LRL**pS**PSPTSQR		2.74
LMNB1	AGGPTTPL**pS**PTR	2.96	2.42
LMNB2	AGGPATPL**pS**PTR	1.80	
MARK2	DQQNLPYGVTPA**pS**PSGHSQGR	2.21	1.71
	VPA**pS**PLPGLER		0.89
MAP7	LL**pT**PTHSFLAR		1.53
MFF	VLTLSERPLDFLDLERPPT**pT**PQNEEIR		1.51
MAPKAPK2	VPQ**pT**PLHTSR	0.66	0.84
MICALL1	KASESTTPAPP**pT**PRPR	3.29	
MLLT4	TPAAIPA**pT**PVAVSQPIR	1.13	1.24
MLL2	LSRPPPPA**pT**PSSMDVNSR	1.40	
	ASEPLL**pS**PPPFGESR		1.41
MLL4	LADVAPTPPK**pT**PAR	1.44	
MARCKS	GEPAAAAAPEAGA**pS**PVEK	4.04	
NDUFA10	KY**pS**PGYNTEVGDK	2.49	3.08
NDE1	TDTAVQATGSVPS**pT**PIAHR		1.64
TPR	TVPS**pT**PTLVVPHR	1.22	1.15
	TDGFAEAIH**pS**PQVAGVPR	1.70	1.89
	GLQL**pT**PGIGGMQQHFFDDEDR	2.30	2.22
NCOR2	GIITAVEPS**pT**PTVLR	1.35	1.31
NOP2	GTDTQTPAVL**pS**PSK	3.57	
NUP107	QPF**pT**PTSR	1.91	1.88
	NQVIPR**pT**PSSFR		3.91
NUP153	LM**pT**PKPVSIATNR	1.78	1.76
	VQMT**pS**PSSTGSPMFK	1.79	2.38
	SVYFKPSL**pT**PSGEFR		1.77
NUP98	NLNNSNLF**pS**PVNR	1.91	2.15
	DSENLA**pS**PSEYPENGER	3.77	
PAK1	SVIEPLPV**pT**PTR	0.85	0.92
PCNP	DTPT**pS**AGPNSFNK (non-canonical)	2.45	2.10
PHACTR4	IQQALTSPLPM**pT**PILEGSHR	0.98	1.17
PHLDB1	ELPPL**pS**PSLSR	1.41	1.54
PHLDB2	TSASEGNPYVSSTLSVPA**pS**PR	1.52	2.70
PCF11	IDGPP**pT**PASLR		1.92
PIAS4	AGAVPR**pT**PLAGPNIDYPVLYGK		2.61
PPP1R13L	AMLPG**pS**PLFTR	0.80	1.47
	IPMPPS**pS**PQPR	1.03	1.34
	QQGPP**pT**PFDFLGR		1.40

	G**pS**PLAEGPQAFFPER		1.39
PPP1R15A	APL**pS**PSLLIR		1.52
RANBP2	FESPATGIL**pS**PR	1.95	2.15
	ELVGPPLAETVF**pT**PK	3.00	1.94
	LLLDIPLQ**pT**PHK		2.01
RAVER1	LL**pS**PLSSAR	1.59	1.84
	EALGLGPPAAQL**pT**PPPAPVGLR		1.66
RELL1	SLMSVSGAETVNGEVPA**pT**PVKR	1.99	1.83
	EAAFGGGLL**pS**PGPEAT (terminus, no K/R)		
RAI1	SLTALSEPR**pT**PGPPGLTTTPAPPDK		1.07
	pTPGPPGLTTTPAPPDK	0.74	
RAI14	SITS**pT**PLSGK	1.17	1.10
ARHGEF11	SLENPTPPF**pT**PK	1.63	
RPS6KA1	DSPGIPP**pS**AGAHQLFR	1.29	1.71
RPS6KA4	LEPVYSPPG**pS**PPPGDPR	2.00	2.85
RBM14	ASYVAPL**pT**AQPATYR	2.35	2.02
RBM9	TEEAAADGGGGMQNEPL**pT**PGYHGFPAR		1.77
SCRIB	LPLLPPE**pS**PGPLR	1.43	1.40
	AFAAVPTSHPPEDAPAQPP**pT**PGPAASPEQLSFR		1.70
SEC16A	FANL**pT**PSR	1.25	1.43
	ADSGPTQPPLSL**pS**PAPETK	1.45	1.29
SHROOM1	SGPQPTEVPG**pT**PGPLNR	1.09	1.81
SIPA1L1	LIDLE**pS**PTPESQK	2.37	2.29
SNX2	EPILSSEPSPAVTPV**pT**PTTLIAPR		2.36
SPTBN1	TSSISGPL**pS**PAYTGQVPYNYNQLEGR		1.25
SRRM2	VPL**pS**AYER	2.55	2.13
	TPAAAAAMNLA**pS**PR	3.35	
	TAVAPSAVNLADPR**pT**PTAPAVNLAGAR		2.04
SF4	AQTSTDAPTSAPSAPPS**pT**PTPSAGK	4.67	
SFPQ	GMGPG**pT**PAGYGR	3.67	
STMN1	ESVPEFPL**pS**PPK	2.11	1.60
	ASGQAFELIL**pS**PR	1.95	1.68
SAMD4B	HALT**pS**PSLGGQGR		2.14
STIM2	SIVPS**pS**PQPQR	1.90	
SUPT5H	DVTNFTVGGFAPM**pS**PR		1.74
SYTL2	GILQVLPDEITFPL**pS**PLRK		0.54
TXLNA	RPEGPGAQAPS**pS**PR	2.02	
TPX2	SSDQPLTVPV**pS**PK	1.80	1.75
TMEM51	NVPPPSIEPL**pT**PPPQYDEVQEK	2.02	1.85
GIGYF2	ALSSGGSITSPPL**pS**PALPK	1.24	1.37
	ALSSGGSIT**pS**PPLpSPALPK		1.78
TUBB	IMNTFSVVP**pS**PK	1.01	
UBAP2L	SPAVATSTAAPPPPS**pS**PLPSK	2.00	2.06
	FPLDYYSIPFPTPT**pT**PLTGR		1.82
ZCCHC8	GTPPL**pT**PSDSPQTR	2.16	1.57