Functional Genomics Approaches to Identify and Characterize Oncogenic Signaling

Abstract

Oncogenes drive cancer by hijacking normal cellular functions involved in proliferation and survival. Suppression of the driving oncogene is highly effective for promoting tumor regression, a phenomenon termed “oncogenic addiction.” By using unbiased genetic tools to functionally probe oncogenic addiction, we can identify cancer dependencies and characterize aspects of oncogenic signaling.

We took an integrative approach to identifying oncogenes by merging loss-of-function RNAi screening data with gene expression and copy number for a panel of 102 cancer cell lines. We developed a computational approach, ATARiS, to improve the interpretation of RNAi screen data. We validated the method by comparing ATARiS gene phenotype values, which quantifies genetic dependency based only on RNAi reagents predicted to have good performance, to known oncogenic dependencies. Using this method, we identified *HNF1B* as a gene that is essential in cancer cells with high *HNF1B* expression. Furthermore, we find that *HNF1B* is amplified in tumors, and expression of *HNF1B* increases anchorage-independent growth. Thus, our approach improves identification of oncogenic drivers such as HNF1B.

In order to probe signaling pathways for a specific oncogene, *RAS*, we performed a genetic screen to rescue KRAS suppression in a KRAS dependent cell line. Using the Broad-CCSB library of ~15,000 open reading frames (ORFs), we identified YAP1. This finding was supported by a *Kras*-driven murine lung cancer model, in which acquired resistance to Kras suppression required Yap1. To assess the mechanism, we performed transcriptional profiling to identify gene sets and transcriptional motifs that are shared between KRAS and YAP1 regulated genes. We report an interaction between YAP1 and

FOS at the promoter regions of genes involved in the epithelial-mesenchymal transition (EMT). Together, these findings implicate transcriptional regulation of EMT by YAP1 as an essential component of oncogenic RAS signaling.

This thesis describes advances in genome scale gain- and loss-of-function technology to systematically identify and characterize cancer dependencies. We report novel biological findings such as oncogenic properties of HNF1B, and characterize YAP1 and EMT as a key component of RAS signaling. While specific genes were chosen for follow up studies, other genes that scored in our analyses may yield additional biological insights.

Table of Contents

Acknowledgements

Innumerable people have contributed to the success of this time while pursuing my Ph.D. Being part of the Hahn lab has been a great privilege, and I am honored to have met and befriended such talented, kind people. In particular, my mentor Bill Hahn has been a strong role model. He believes in us, and with that belief gives us the confidence to believe in ourselves and to trust our data. He seems to know exactly when his advice is needed and when to provide space, providing an atmosphere of support that I know I am lucky to have. Many people in the lab have supported me along the way, from Susan Moody and David Barbie and Xiaoxing Wang who were there even as I just started my Ph.D. and needed much guidance. And now, several years later, I enjoy discussions with newer members, Yaara Zwang, Elsa Krall, and William Kim, who have infused the lab with new ideas and new projects.

Not only does the lab itself have great people, but my collaborators have also been amazing. They open my eyes to different ways of thinking and doing science. From Wen Xue, a postdoctoral fellow in Tyler Jacks's lab, I have seen a different, deeper meaning of collaboration, as he provides not only ideas, but also the willingness to pitch in for the less glamorous side of experimental work. From Aviad Tsherniak, staff scientist at the Broad Institute working with Jill Mesirov, I am encouraged to think critically about the precise meanings of each sentence and word. I believe our conversations and our friendship have made me a better scientist.

Others that have influenced my time here include the M.D.-Ph.D. and Biological and Biomedical Sciences (BBS) program offices and advisors, who were there to discuss my choices in classes and laboratory and ensured that the past few years would run smoothly. My dissertation advisory committee, which includes Kevin Haigis, David Sabatini, and Peter Park, have shaped my thesis and offered numerous constructive

advice and support. I have been impressed and honored also by the willingness of my thesis committee members, Pasi Jänne, Robert Weinberg, and Adam Bass to make time in their schedule to accommodate my defense.

I want to thank the friends I have made over the past few years, most related to my Ph.D. program in some way. My classmates in the BBS program have become closer over time. Even as we all diverged into our respective laboratories, their availability for fun, commiseration, and weekly lunch through the past few years have made the graduate school experience great. In the Cancer Program at the Broad Institute, I have progressed through my training with a small cohort of graduate students, who I view as though they are graduate students in my own lab. As we all come to the conclusion of our projects, it has been impressive to reflect on our progress together. Finally, I would also at least in part credit the BBS program's extensive social calendar with the opportunity to meet my boyfriend Aleksandar Kostic – I cannot imagine anyone else on earth that understands me and complements me so well.

On a more personal note, I would like to thank my family who should receive credit for not only their support for these past few years, but rather for a lifetime. My brother Kevin, an elementary school student when I first started medical school, is about to become a junior in high school, and has become thoughtful and responsible, almost even mature. I know he brags about me (or at least he used to) to his friends, and I am motivated to live up to his accolades. Over the past few years, my mom Judy and my dad Mike have started to view me not only as their daughter but also as their friend. During this process of gaining a degree, I have also gained the maturity to take part in family decisions and provide source of support for them as well. Our family has become stronger over the years, together overcoming bad times and cherishing each other's joys and achievements.

Dedicated

In memory of my grandfather

Shuyao Shao

CHAPTER ONE

Introduction: Oncogenic Signaling and Genomic Technology

1.1 Overview

Cancer arises due to deregulation of cellular pathways in normal cells by somatic changes in the genome. The process of oncogenesis involves activation of oncogenes that drive cancer progression, and the repression of tumor suppressors that normally control proliferation. Through these mechanisms, tumor cells acquire cell intrinsic properties such as increased proliferation, the ability to evade cell death and tumor suppressors, and anchorage independent growth. In vivo, cancer cells must also be able to regulate cell extrinsic processes such as angiogenesis and the ability to evade the immune system (Hanahan and Weinberg, 2011).

The dogma of cancer as a disease caused by altered versions of somatic genes originally arose from the discovery of the *src* oncogene. Rous Sarcoma Virus (RSV) was known to acutely cause development of chicken sarcoma, though the transforming element, *v-src*, was dispensable for viral replication. When *v-src* was shown by molecular hybridization to be homologous to a gene in the chicken host genome, *c-src* (Stehelin et al., 1977), it became clear that the viral oncogene was an altered version of a gene originally acquired from the host chicken genome. Similar discoveries were made for a number of other viral oncogenes in different species (Maeda et al., 2008).

The first oncogenes from human cancer cells were identified by the observation that cDNA derived from human tumor cell lines caused transformation of mouse NIH-3T3 fibroblasts when introduced by chemical transfection (Shih et al., 1981). One of these oncogenes turned out to be a mutant version of cellular *RAS* (Parada et al., 1982), a potent oncogene later shown to drive tumorigenesis in a number of human cancers. The function of RAS involves response to mitogenic signals (Kamata and Feramisco, 1984), and mutant RAS drives constitutive signaling, leading to abnormal cellular proliferation by overcoming checkpoints that halt cell division in non-cancerous cells.

Oncogenes involved in this pathway are covered in detail in Chapter 1.2 of this Introduction.

While oncogenes like *RAS* alone are sufficient to drive transformation of immortalized cells such as NIH-3T3, additional genomic alterations are necessary for transformation of primary cells. Primary murine cells require RAS in combination with MYC signaling for transformation (Land et al., 1983), and transformation of primary human cells requires RAS in combination with SV40 large T antigen and Telomerase Reverse Transcriptase (TERT) to maintain telomere length (Hahn et al., 1999), and thereby bypass cellular senescence. This muti-step model of tumorigenesis was supported by efforts to characterize cancer genomes, which established a systematic acquisition of somatic DNA alterations in human tumors. For example, patients harboring adenomatous polyposis coli (*APC*) mutations have a highly proliferative colon epithelium that leads to increased risk for colorectal cancer. In these patients, as adenomatous lesions, precursors to colon cancer, progress into full carcinomas, the DNA acquires increasing numbers of defined tumorigenic changes including mutations in *RAS* (Fearon and Vogelstein, 1990).

As tools to delve deep into the cancer genome have steadily improved in quality and declined in price, our understanding of genetic changes in cancer has been steadily improving. Not only have sequencing efforts identified mutations, array comparative genomic hybridization (aCGH) techniques have revealed numerous copy number gains and losses across the genome (Beroukhim et al., 2010). These genomic changes involve not only oncogenes or tumor suppressors but also many genes in close proximity, thus obscuring the ability to identify the driving event. To add to the complexity, genomic gains, losses, and translocations have recently been shown to involve highly disruptive genomic events across multiple chromosomes through processes of chromothripsis or chromoplexy (Forment et al., 2012; Baca et al., 2013).

Given this genomic complexity, functional characterization is required to determine oncogenic drivers and understand their signaling pathways. In addition to genomic alteration, oncogenes can be defined by specific traits such as cellular dependency on the altered oncogene, and the ability to initiate tumor formation in immortalized models. RNA interference (RNAi), or the introduction of short interfering RNA (siRNA) into cells to degrade mRNA transcripts of complementary sequence, allows investigators to assess cellular requirement for the expression of specific genes. Expression libraries, a category that includes the cDNA libraries that initially identified the transforming properties of RAS (Shih et al., 1981), provides the opportunity to probe gain-of-function activity upon gene expression.

In this thesis, we are interested in systematic assessment of gene function in cancer. We improve analytical methods and apply our approach to identify novel oncogenic dependencies. We review existing knowledge of oncogenic signaling and the mitogenic pathway, and show how gain-of-function tools shed new light on well-established pathways. The work presented here contributes to the current exciting environment of cancer research, as analytic and technical tools advance our functional understanding of cancer biology.

1.2 Mitogenic signaling in cancer

1.2.1 RAS signaling

One of the first discovered human oncogenes was RAS. RAS is a guanosine-nucleotide binding protein (G protein) located at the cell membrane, and its activity is determined by GTP binding. GTPase Activating Proteins (GAPs) increase GTPase activity, decreasing RAS signaling; and Guanine Exchange Factors (GEFs) convert the GDP-bound state to a GTP-bound state, activating RAS. Oncogenic mutations in RAS occur at conserved sites: glycine-12, glycine-13, or glycine-61. Disruption of these amino acids prevents the activity of RAS GAPs. As a result, RAS is locked into a chronically active GTP-bound state and constitutively activates downstream genes driving tumor growth (Reviewed in Ahearn et al., 2011). A simplified diagram of the RAS pathway involved in cellular proliferation is shown in Fig. 1-1.

The RAS family includes 3 proto-oncogenes *KRAS, HRAS,* and *NRAS.* These genes encode 4 proteins – KRAS4A, KRAS4B, HRAS, and NRAS — all of which become oncogenic when activating mutations are acquired. *RAS* is the most commonly mutated oncogene in human cancer and is found in a variety of cell lineages. Despite structural similarities of the RAS proto-oncogenes, they show different distributions in human cancer. For example, *NRAS* mutations are primarily observed in hematopoietic malignancies and melanomas; *HRAS* mutations are observed primarily in bladder cancer; and *KRAS* mutations are predominant in lung, colon, and pancreatic tumors (Pylayeva-Gupta et al., 2011).

It is unclear whether differences in the distribution of *RAS* mutations occur due to differential lineage-specific gene expression or due to differences in gene function (Reviewed in Lau and Haigis, 2009). In colon cancer, *NRAS* and *KRAS* provide distinct functions that drive different responses in the colonic epithelium. Mice that express an

activated *Kras* allele in the colonic epithelium present a proliferative phenotype, whereas *Nras* primarily prevented susceptibility to apoptosis (Haigis et al., 2008). In lung cancer, KRAS and HRAS appear to play interchangeable roles. In murine models in which *Hras* was knocked in to the endogenous *Kras* locus, lung cancer induced by chemical carcinogens was equally likely to develop through *Hras* mutation (To et al., 2008), suggesting that lineage-specific expression differences rather than differential lineage-specific function. Furthermore, in development, *Hras* knock in fully rescued the embryonic lethality characteristic of *Kras* null embryos (Potenza et al., 2005). Thus, the context specificity of cellular response to RAS signaling is still being elucidated.

Models of RAS-driven tumorigenesis *in vivo* and *in vitro* have made characterization of the signaling pathway possible. The most convincing proof of *Ras*-driven oncogenesis *in vivo* involves expression of mutant *Ras* driven by the endogenous *Ras* promoter (Tuveson et al., 2004). In this model, conditional activation of the mutant allele is made possible by conditional Cre-mediated excision of a STOP element, leading to tissue-specific and/or temporally specific tumor development. *In vitro* transformation of human cells can be achieved by expression of RAS in immortalized cells. Immortalization is possible spontaneously (Soule et al., 1990), or by introduction of defined genetic elements SV40 Large T and TERT (Hahn et al., 1999). These models allow interrogation of RAS signaling, for example by replacing RAS with a combination of its downstream effectors AKT and MEK (Boehm et al., 2007). These effectors and the effects of RAS signaling are reviewed in Chapter 1.2.3 and 1.2.4.

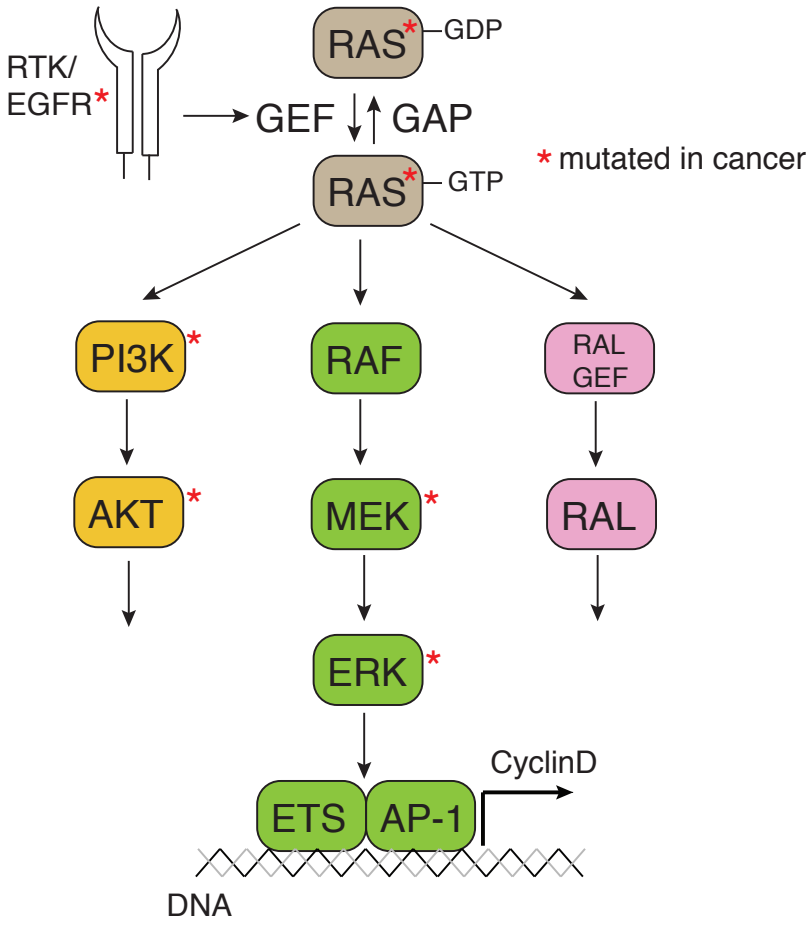

Figure 1-1. Mitogenic signaling through the RAS pathway.

1.2.2 RAS upstream signals

RAS has long been known to become activated upon stimulation by mitogens such as epidermal growth factor (EGF)(Kamata and Feramisco, 1984). Mitogens, extracellular signals that regulate growth, exert their activity through binding to receptor tyrosine kinases (RTKs) at the cell surface. Their role in cancer was quickly realized due to high homology of the RTK, epidermal growth factor receptor (*EGFR*), to a viral oncogene *v-erbB* (Downward et al., 1984). The concept that RAS activity was downstream of RTK signaling was supported by work showing that the oncogenic potential of RTKs required RAS signaling. When a monoclonal antibody targeting Ras was microinjected into NIH-3T3 cells, RTK-derived viral oncoproteins could no longer transform cells (Smith et al., 1986).

Conserved structural properties of RTKs include an extracellular binding domain, a transmembrane region, and intracellular catalytic regions for tyrosine kinase activity. The EGFR family initially described includes EGFR, ERBB2, ERBB3, and ERBB4. Later, RTKs including the insulin receptor (INSR), insulin-like growth factor receptor (IGF1R), platelet-derived growth factor receptor (PDGFR), and KIT, amongst others, were described. Of the 58 RTKs now identified, over half have been shown to be activated in cancer and play roles in increased cellular proliferation (Reviewed in Gschwind et al., 2004).

Mechanisms that activate EGFR, the prototypical RTK, are numerous. Amplifications of *EGFR* were initially identified in an epitheloid carcinoma cell line (Ullrich et al., 1984), and later found in many cancers of epithelial origin including in non-small cell lung cancer (NSCLC)(Hirsch et al., 2003). Sequencing projects from human cancer tissue have identified a number of mutations and deletions in the *EGFR* catalytic domain that cause increased receptor activity (Paez et al., 2004). Aberrant regulatory

mechanisms can also result in increased receptor expression and activity. One such mechanism includes downregulation of genes involved in targeting EGFR to the lysosome for ubiquitin-mediated degradation (Peschard and Park, 2003). While mechanisms of EGFR activation have been well characterized, analogous mechanisms of activation have also been described for other RTKs.

Signaling between RTK signaling and RAS requires a series of adaptor proteins that bind to specific binding domains. Adaptor proteins harbor SH2 domains that bind to regions on the receptor as well as SH3 domains that associate with RAS GEFs such as SOS1 and SOS2. Many adaptor proteins have been shown to drive cancer development. Adapter proteins such as insulin receptor substrate IRS1 and IRS2 increase breast cancer tumorigenesis and metastasis in mouse models (Dearth et al., 2006). The CRK family of adaptor proteins has been implicated in cancer through observations of increased gene expression, and amplification and functional requirement of *CRKL* in NSCLC (Bell and Park, 2012; Cheung et al., 2011b). Together, these observations confirm the relevance of upstream mitogenic signals in regulation of RAS signaling and cancer.

1.2.3 RAS effectors in regulation of cell proliferation

The ability of RAS to drive cancer growth likely involves the coordinated regulation of multiple effector pathways. Of the over ten RAS effectors that have been identified (Malumbres and Barbacid, 2003), effectors that are activated during tumorigenesis are the best characterized. First, the effector RAF was shown to directly interact with the GTP-bound form of RAS (Warne et al., 1993; Moodie et al., 1993). *RAF* mutations have been described in many human tumors, and though they often occur in cell lineages that also acquire *RAS* mutation, *BRAF* mutations are often exclusive of

mutations in RAS (Reviewed in Davies et al., 2002). Subsequently, RAS was shown to bind to the phosphoinositide 3-kinase (PI3K) catalytic subunit and modulate signaling through PI3K (Rodriguez-Viciana et al., 1994). *PI3K* is also mutated in many human cancers (Reviewed in Liu et al., 2009), but unlike RAF, PI3K is directly activated by RTK signaling in addition to RAS signaling. Discovery of RAL guanine nucleotide dissociation stimulator (RAL-GDS) interaction with RAS soon followed (Hofer et al., 1994). These pathways are depicted in Fig. 1-1. Genetic suppression of each effector alone prevents Ras-mediated transformation (Kolch et al., 1991; Ehrenreiter et al., 2009; González-García et al., 2005; Gupta et al., 2007). The contribution of each component to RAS signaling is still under investigation.

The relative contribution of each RAS effector to RAS signaling is likely context specific. Ras-mediated activation of Raf was sufficient for transformation of mouse NIH-3T3 cells whereas activation of Ral-GDS and PI3K were not. However, in immortalized human cells, activation of Ral-GDS was sufficient to drive transformation whereas activation of RAF and PI3K were not (Hamad et al., 2002). In contrast to models of transformation where activated PI3K was not sufficient for transformation, an activated form of PI3K can replace RAS to maintain viability *KRAS* mutant cancer cell lines (Lim and Counter, 2005). Thus, not only do different transformation models point to differential requirements of RAS signaling, different mechanisms may also exist between initiation versus maintenance of RAS signaling.

One major downstream effect of these pathways is control of cell proliferation. Proliferative signals involve activity of transcription factors such as FOS and activating transcription factor 2 (ATF2) (Mechta et al., 1997; Gutman et al., 1991). These genes are involved in the mitogenic response as components of activator-protein (AP-1) transcription factor complexes. ETS transcription factors in association with AP-1 sites have also been implicated in driving a RAS-related transcriptional program (Hollenhorst

et al., 2011). One result may involve increased expression of cyclin D1 (CCND1), a primary regulator of entry into the cell cycle (Filmus et al., 1994). In mouse models of breast cancer tumorigenesis, genetic ablation of cyclin D1 specifically prevented Ras-driven tumorigenesis, whereas tumor formation driven by Myc or Wnt signaling remained intact (Yu et al., 2001). However, while regulation of cell cycle entry is an important aspect of RAS signaling, additional functions of RAS are likely involved in its oncogenic properties.

1.2.4 RAS activates pleiotropic downstream signals

In addition to driving proliferation, RAS has also been implicated in regulation of metabolism, apoptosis, and metastasis. Together, these properties contribute to tumor growth and maintenance. The mechanistic details of RAS regulation of such diverse pathways are still under investigation.

RAS has been shown to regulate both signaling pathways that promote as well as suppress apoptosis. Although pro-apoptotic roles appear contradictory to the role of RAS in oncogenesis, both pro-survival and pro-apoptotic signals are kept in balance in normal cells. Even though oncogenic RAS increases both survival and apoptotic signaling, the balance is tipped toward uncontrolled cell proliferation except in specific contexts of cellular stress. Signaling through PI3K-AKT and RAL pathways activate NF-kB transcription factor activity, which increase the expression of pro-survival genes such as inhibitors of apoptosis (IAP) and BCL-2 (Chien et al., 2006; Reviewed in Mayo and Baldwin, 2000). Furthermore, expression of NF-kB suppressors such as IkBα reduces Ras-driven focus formation in NIH-3T3 (Finco et al., 1997). Pro-apoptotic effects of RAS are primarily attributed to its effectors in the Ras-association domain family (RASSF), RASSF1 and RASSF5 (Khokhlatchev et al., 2002). Complex formation of RASSF with

MST1 kinase increases caspase activity, sensitizing cells to apoptosis. *RASSF1* has been shown to epigenetically inactivated, particularly frequently in tumors that harbor *RAS* or *RAF* mutations (Richter et al., 2009). Thus, the dual effects of pro-apoptotic and pro-survival signals are tenuously balanced in cancer cells.

Metabolic pathways in cancer cells must be altered to support catabolic processes that enable high rates of cell division. Expression of RAS is associated with higher levels of enzymes supporting biosynthetic processes (Gaglio et al., 2011). RAS expression also increases expression of the glucose transporter (GLUT1), increasing glucose update (Flier et al., 1987). A recently described phenomenon attributed to RAS is autophagy, whereby a cell breaks down its organelles to generate energy and building blocks for cellular survival (Lock et al., 2011). Thus, a constellation of metabolic changes that are driven by RAS support the proliferation of tumor cells.

RAS activation leads to disruption of cell-cell contacts that are associated with the process of metastasis. Transcriptional regulators such as SNAIL and SLUG are upregulated. These genes repress cell junction proteins such as E-cadherin (CDH1) and promote widespread transcriptional changes that lead to an invasive phenotype (Schmidt et al., 2005; Giehl, 2005). This change is termed the epithelial-mesenchymal transition (EMT) and has been implicated in metastasis for many cancers. As will be discussed later, emerging paradigms suggest that RAS regulation of EMT may be important for cell survival in addition to metastasis. In order to bypass physical barriers such as the basement membrane, RAS regulates extracellular proteases to help degrade components of the extracellular matrix (Meade-Tollin et al., 1998). Furthermore, effector pathways such as RAF, PI3K, RAL, and RHO have all been implicated to impinge on processes for metastatic progression (Reviewed in Pylayeva-Gupta et al., 2011).

1.3 Oncogenic addiction

1.3.1 Definition and concept

Although tumorigenesis occurs through the acquisition of multiple genetic alterations, suppression of any single driving oncogene leads to tumor regression. Depending on the oncogene, this regression may be the result of apoptosis or cellular differentiation. This response was initially surprising because it was initially believed that removal of a single oncogenic stimulus would simply prevent tumor progression rather than lead to tumor regression; or that there would be no effect if the multiple co-existing oncogenic lesions were sufficient to drive continued growth. Furthermore, there is often a high differential dependency on the oncogenic gene in cancer cells as compared to their normal counterparts, leading to a large therapeutic window. The exaggerated response to inhibition of an oncogene in tumor cells was termed “oncogenic addiction” (Weinstein, 2000).

One of the first clinical proof-of-concept studies came from treatment of Chronic Myelogenous Leukemia (CML) with specific inhibitors of the driving oncogene BCR-ABL (Reviewed in Sherbenou and Druker, 2007). The *ABL* gene encodes a tyrosine kinase whose constitutive expression when translocated to the *BCR* locus specifically drives aberrant production of myeloid cells. When patients were treated with a kinase inhibitor targeting the ABL kinase, myeloid cell counts decreased to normal levels, and serum levels of the *BCR-ABL* translocation became undetectable, suggesting a specific depletion of cells harboring the translocation. Despite the presence of ABL protein in all cells, normal cells were resilient to ABL inhibition, and thus this kinase-specific treatment showed greatly reduced cytotoxicity compared to traditional chemotherapy.

This paradigm since has been extended clinically to a number of oncogenic drivers, particularly kinases for which pharmacologic inhibitors could be readily

developed (Weinstein and Joe, 2008). Inhibition of oncogenic forms of receptor tyrosine kinases has been successful in many different cancer types, and includes KIT in gastrointestinal tumors, EGFR in lung cancer, ERBB2 in breast cancer, and JAK in myelofibrosis. BRAF has also been successfully inhibited in melanoma causing tumor regression, and ALK in non-small cell lung cancer.

For many oncogenes in which direct pharmacologic inhibition is not possible clinically, mouse models and cancer cell lines suggest that tumors rely on continued oncogenic signaling. For example, efficacious RAS inhibitors have not been developed; however, oncogene addiction to RAS has been clearly illustrated by cancer models in mice and human cell lines (Fisher et al., 2001; Chin et al., 1999; Barbie et al., 2009). In fact, the concept of oncogenic addiction has become so prevalent that discovery of new oncogenes now often requires evidence that cancer cells require continued expression or activity of the putative oncogene for survival.

1.3.2 Mechanisms of gene dependency

Different mechanisms have been presented to explain the dependency on oncogenes. One mechanism of establishing dependency on continued oncogenic signaling, termed the “oncogenic shock” hypothesis, argues that oncogenic signaling may directly pre-dispose cells to cell death when oncogenic signals are removed. Similar to the dual pro-survival and pro-apoptotic signals provided by RAS (Chapter 1.2.4), many other oncogenes also elicit these opposing signals. The “oncogenic shock” hypothesis suggests that differential attenuation of pro-survival and pro-apoptotic signals when an oncogene is acutely removed leads to a cellular imbalance in favor of apoptosis (Sharma et al., 2006). Alternatively, but not exclusively, stochastic genetic alterations may drive cancer cells to depend on specific oncogenes. For example, expression of

one oncogene may set up a network that is permissive for stochastic loss of genes that would normally be required for cellular function (Kamb, 2003). Thus, cells no longer tolerate the loss of the oncogenic signal.

The two scenarios set up above – cellular rewiring caused by direct oncogenic signals, or accumulated changes due to changing cellular requirements – agree that one fundamental property is a cellular signaling network that differs between cancer cells and normal cells. One corollary to such different cellular networks is that cancer cells harbor differential dependencies than their non-cancerous counterparts. This phenomenon has been loosely termed "non-oncogene addiction" (Luo et al., 2009b), to suggest dependency on genes that are not oncogenic drivers. Non-oncogene addiction may also arise due to circumstantial susceptibilities that arise as tumor cells evolve. For example, heterozygous gene deletions often occur in cancer, sometimes including genes for which complete loss would be lethal (Nijhawan et al., 2012). Thus, even though these genes may not drive tumor growth, they also represent a class of dependencies in tumor cells that can be exploited. Understanding the unique networks established in tumor cells will inform ways to identify and target gene dependencies in cancer.

1.3.3 Strategies for leveraging oncogenic signals

While some cancer drivers are amenable to direct pharmacological inhibition, aberrant signaling in a tumor cells presents additional opportunities. Attempts to take advantage of these cellular vulnerabilities have yielded successful clinical inhibitors. For example, *BRCA1* and *BRCA2* are tumor suppressor genes that confer high risk of breast and ovarian cancer when deleted. Since their normal cellular function involves double-stranded break repair, tumor cells become more susceptible to DNA damage. PARP inhibitors, which increase the rate of double-stranded breaks and other DNA damage

that requires BRCA function for repair, show high efficacy in tumor cells with *BRCA* loss (Helleday, 2011). Targeted inhibitors, including those that inhibit histone deacetylases (HDACs) that globally modulate gene expression, and proteasome inhibitors that prevent protein turn over, are efficacious in multiple tumor types (Reviewed in Kisselev and Goldberg, 2001; Khan and La Thangue, 2011). While such inhibitors do not target a specific tumor driver, these results show that signaling networks in tumor cells are clearly highly susceptible to targeted inhibition of specific pathways.

Direct pharmacological inhibition of RAS has not been successful, though many models have shown that RAS-driven tumors require continued RAS expression. Inducible oncogenic *Kras* initiated lung adenomas in mice regressed rapidly upon removal of oncogenic Kras expression (Fisher et al., 2001), and this susceptibility holds true for Ras-driven tumors in other cell lineages (Chin et al., 1999; Jechlinger et al., 2009). Efforts to target RAS-driven cancers have focused on suppressing the activity of RAS effector kinases. However, pharmacologic inhibitors of BRAF or PI3K have not been individually successful in targeting RAS mutant cancers (Ihle et al., 2009; Solit et al., 2006). Regression was observed by inhibition of both genes combined in murine models of Kras-driven lung cancer (Engelman et al., 2008), and this combination is being investigated clinically.

One way that the RAS-driven cancers can be targeted is to identify gene dependencies specific for the RAS addicted state, or genes that are “synthetic lethal” to oncogenic RAS. The original concept of synthetic lethality was coined in yeast biology to describe pairs of genes that are individually dispensable for survival, yet cause lethality when both are lost. In the context of oncogenic signaling, the oncogenic signal is analogous to the first “lost” gene, which sets up a rewired cellular network with different genetic dependencies. Synthetic lethal genes to KRAS such as polo-like kinase 1 (PLK1), serine/threonine protein kinase 33 (STK33), TANK-binding kinase 1 (TBK1), and

snail homologue 2 (SNAI2) have been identified by RNAi viability screens in human cancer cell lines (Barbie et al., 2009; Scholl et al., 2009; Luo et al., 2009a; Wang et al., 2010). Additionally, WT1 dependence has been implicated using screens in Kras-driven mouse models (Vicent et al., 2010). These studies point to potentially susceptible points of KRAS signaling, but also reveal biology around the signals important for KRAS addiction. The mechanism of TBK1 dependence supports the RAL pathway as a major requirement in RAS oncogene addiction, and SNAI2 dependence points to the role of EMT as not only a mediator of tumor progression but also a mediator of cell survival.

1.3.4 Cellular mechanisms to bypass oncogene suppression

Despite the initial success of targeted kinase inhibitors, resistance mechanisms to therapy were quickly arose. Cancer cells can be resistant to therapy *de novo*, in which no tumor response is detected, or acquire resistance after initial response to therapy. While some mechanisms are shared between *de novo* and acquired resistance, additional non-cell autonomous mechanisms apply to *de novo* resistance such as drug metabolism by the patient or poor drug permeability to target sites. Studies of acquired resistance reveal mechanisms that pertain to both acquired and *de novo* resistance.

Many mechanisms of resistance re-establish the signaling pathways that are activated by the targeted oncogene. For example, *EGFR* mutant lung cancer treated with EGFR inhibitors commonly acquire resistance via amplifications in *MET* or *ERBB3* (Reviewed in Jänne et al., 2009). These receptor tyrosine kinases signal through similar, parallel pathways. Resistance also occurs when activating mutations in the *RAS* oncogene are acquired, thus activating signals downstream of EGFR. Correspondingly, treatment with EGFR inhibitors is contraindicated in patients whose tumors already harbor mutations in *RAS de novo*. In addition to activating alternate routes of signaling,

tumor cells have found creative ways to increase EGFR activity directly. Inhibition can be avoided through resistance mutations in *EGFR*, upregulation of EGFR ligands such as amphiregulin, and deregulation of EGFR degradation or internalization (Wheeler et al., 2010). While these varied responses highlight challenges to effective therapy, they also confirm the critical importance of maintaining the driving oncogenic signal.

One resistance mechanism that has been attributed to resistance to a number of chemotherapies is the epithelial-mesenchymal transition (EMT). EMT is associated with a less differentiated state often implicated in stem-cell-like states in cancer (Singh and Settleman, 2010). Cancer therapies fail to eradicate cancer stem cells due to their properties of quiescence, upregulation of drug membrane transporters, and resistance to apoptosis (Reviewed in Dean et al., 2005). While these properties clearly might contribute to drug resistance, the intrinsic properties of stem cells may not fully account for mechanisms of resistance. In fact, EMT may mediate resistance through specific signaling pathways. For example, increased E-cadherin (CDH1) expression, a marker of an epithelial state, correlates with increased sensitivity to EGFR inhibition; and restoring CDH1 expression in cells with low CDH1 levels can restore sensitivity to inhibitor treatment (Witta et al., 2006). In hepatocellular carcinomas which harbor *EGFR* amplifications, cells classified as epithelial or mesenchymal based on EMT marker expression predicted sensitivity or resistance to EGFR inhibitors, respectively (Fuchs et al., 2008).

Diverse mechanisms enable tumor cells to bypass oncogene dependence. The concepts of re-activation of downstream or parallel signaling driving resistance are conserved through studies of many oncogenes. Together, studies of resistance to oncogenic suppression increase understanding of the initiating oncogenic signal, inform decisions about therapy, and provide strategies for avoiding the development of resistance.

1.4 Applying genomic tools to identify cancer drivers

1.4.1 Characterizing structural changes in the cancer genome

While cancer cells are known to harbor multiple genetic alterations, comprehensive characterization of all genomic changes in cancer is challenging. Cancer cells harbor a gamut of structural alterations in DNA including chromosomal translocations, point mutations, and somatic copy number alterations. These alterations may occur any number of times in a cancer cell, ranging from affecting a single gene in the most stable cancer genomes, to affecting thousands of genes in the most unstable cancer genomes. Tools have been developed to distinguish alterations in genes that contribute to cancer progression from those that are altered as byproducts of genomic instability.

Somatic copy number alterations (SCNAs) can be annotated by using microarrays that span the genome. A number of oncogenes are primarily activated through amplification, including *MYC* and yes-associated protein 1 (*YAP1*) (Deming et al., 2000; Overholtzer et al., 2006). However, since any genomic region might be altered by chance, statistical measures must be used to determine whether a region of copy number alteration is recurrent, i.e. likely to harbor genes that drive cancer development. One method commonly applied, Genomic Identification of Significant Targets in Cancer (GISTIC), identifies regions that harbor potential cancer drivers as those that have high amplitude of copy number gain or deletion, and recurrence across multiple samples (Beroukhim et al., 2007; Mermel et al., 2011). This algorithm also favors focal genetic events, due to a hypothesis that broad events may occur through distinct mechanisms. This method has identified oncogenes such as *MCL1* and *BCL2L1* (Beroukhim et al., 2010).

Genetic translocations and mutations also contribute to the landscape of cancer. Mutations may arise to activate oncogenes, as described earlier in this this introduction, or to disrupt tumor suppressor function. While mutations within gene coding regions are best characterized, recent reports have also identified recurrent mutations in gene promoter regions that increase the expression of telomerase reverse transcriptase (*TERT*)(Huang et al., 2013), important for bypassing cellular senescence. To distinguish spurious mutations from those that drive cancer progression, algorithms such as Mutational Significance (MUTSIG) have been developed to statistically account for background rates of mutation (Banerji et al., 2012). Genetic translocations often join a proto-oncogene with a genomic region of high transcriptional activity, leading to hyperactive transcription of the oncogene. This is the case with *BCR-ABL*, and this mechanism drives activation of many other oncogenes as well (Reviewed in Nambiar et al., 2008).

A number of mechanisms contribute the rampant DNA alterations that affect genes other than the oncogenic driver. Chromothripsis, “chromosome shattering,” is a dramatically disruptive cellular process that results in a large number of chromosomal rearrangements and SCNAs in a localized region of the genome (Forment et al., 2012). Normal rates of cellular mutation are not enough to account for the high numbers of mutations observed in cancer cells. Thus, many cancers have been attributed with a “mutator phenotype,” at least partially attributed to loss of DNA mismatch repair genes (Prindle et al., 2010). Finally, for genetic translocations, the recently described phenomenon of chromoplexy suggests that translocations across multiple chromosomes occur simultaneously through a single genomic event (Baca et al., 2013). These effects suggest that many disrupted genes are bystanders of an unstable cancer genome, and additional computational and experimental tools must be used to identify cancer drivers.

1.4.2 Gene expression changes in cancer

Gene expression in cancer cells can be characterized at the RNA-level by microarray hybridization or RNA-sequencing technology. Both of these techniques enable identification of genes that are differentially expressed in cancer cells compared to normal counterparts. Since gene expression of oncogenes may be increased and expression of tumor suppressors may be decreased in cancer cells compared to normal cells, integration of expression and copy-number data can shed light on functionally relevant copy number alterations (Bussey et al., 2006). Changes in cellular signaling often involve changes in networks of gene expression rather than in expression of single genes. Thus, statistical methods such as gene set enrichment analysis (GSEA) help to discover sets of genes that are enriched (Subramanian et al., 2005). Gene expression profiling is commonly performed in cell lines and primary cancer tissue, and provides valuable information on cellular function.

Protein expression profiling is less common due to the high cost of mass spectrometry and the inherent noise in mass spectrometry data. One method to interrogate protein expression rapidly and at low cost involves reverse phase protein lysate microarrays (Paweletz et al., 2001). In this technique, whole cell lysates are immobilized in array format and are probed using 50-100 antibodies that bind to specific proteins. By measuring intensity of antibody binding, levels of each protein in the lysate are quantified. While this method is limited by the number of proteins that can be assessed due to antibody quality, this technology has allowed determination of expression of multiple proteins across many cell lines in high-throughput (Nishizuka et al., 2003).

1.4.3 Functional characterization: Loss-of-function tools

RNA interference (RNAi) screening involves short interfering RNAs (siRNAs) that are introduced into cells either directly as a duplex or by expression of short hairpin RNA (shRNAs) to be processed into active siRNAs. SiRNAs are designed to specifically degrade mRNA transcripts of complementary sequence to reduce the expression of gene products (Elbashir et al., 2001; Root et al., 2006). By systematically suppressing each gene, its functional role in a given cellular context can be determined.

An early application of arrayed RNAi screens in mammalian cells was to identify essential genes in a specific genetic context. One such screen identified gene dependencies specific to colon cancer cells that harbor activation of the Wnt pathway (Firestein et al., 2008), a developmental pathway strongly implicated in pathogenesis of many cancers. Two screens were performed in arrayed format to assess the effect of ~1000 genes in the kinome – one screen assayed for genes that affected proliferation, and the other identified genes that modulated Wnt reporter activity. The results of these screens were overlaid with results from an independent analysis to identify recurrently amplified genomic regions in colon cancer cells. The overlap of these lists resulted in the identification of *CDK8* as a colon-cancer specific oncogene.

A transition from plate-based screening to barcode screening greatly increased the scale of screened genes as well as the number of samples that could be screened in parallel (Silva et al., 2008; Luo et al., 2008). Genome-scale RNAi libraries provide an unbiased way to identify new pathway components and functional cancer dependencies. In barcode screening, the entire RNAi library is introduced simultaneously into a large population of cells at a rate of a single RNAi reagent per cell. At the assay end point, the relative abundance of each reagent is deconvoluted using the unique RNAi reagent sequences as “barcodes.” The effect of suppressing each gene can be inferred from the relative abundance of RNAi reagents, i.e. reagents targeting essential genes will be

depleted from the population. This method can be readily expanded to screen large numbers of cell lines in parallel (Cheung et al., 2011a; Marcotte et al., 2012).

In a more recent analysis of genes that are essential in Wnt active cancer cell lines, these improved technological advances enabled interrogation of ~11,000 genes across 102 cancer cell lines across different lineages (Rosenbluh et al., 2012). The results revealed not only single genes that were synthetic lethal, but multiple genetic dependencies in the same pathway including YAP1, YES1, and TBX5. Specifically, YAP1 was shown to be essential for Wnt active cancers and utilized a transcriptional complex including YES1 and TBX5. These results highlight how the ability to functionally dissect signaling pathways can be improved through advances in functional screening.

1.4.4 Functional characterization: Gain-of-function tools

Complementary to the loss-of-function approaches of RNAi, gain-of-function approaches can directly assess the role of specific genes in oncogenesis. To explore gain-of-function approaches at large scale, common techniques include introduction of a pooled cDNA library, such as the initial transformation screens that identified in 3T3s that identified human oncogenes (See Chapter 1.1)(Shih et al., 1981), or by random insertion of viral promoters to increase expression of endogenous genes (Dupuy et al., 2006).

While cDNA libraries can be derived from any source of RNA, their application is limited and biased without the development of libraries that are sequence verified and easily transferred into plasmid backbones that can be introduced into a wide variety of cell types at adequate expression levels. The human ORFeome collection was created to satisfy these goals by sequence verification of genetic open reading frames (ORFs) and cloning them into a Gateway cloning system (Lamesch et al., 2007). An initial

assessment of such an ORFeome collection used 354 kinase ORFs that were myristoylated (increases activation by localization to the cell membrane) and Flag-tagged to identify those that can transform immortalized cells (Boehm et al., 2007). This approach revealed *IKBKE* as a transforming oncogene in breast cancer by substitution for the AKT pathway in transformation.

Cooperation between the Broad institute and the Center for Cancer Systems Biology (CCSB) enabled the expansion of this library to 597 kinases. An analogous screen for transformation using this library identified *PAK1* as an oncogene in breast cancer (Shrestha et al., 2012). In addition to studying tumor initiation, this gain-of-function library was applied to identify resistance mechanisms to RAF kinase inhibition in melanoma cell lines (Johannessen et al., 2010). By assessing the effect of expressing each ORF on cellular susceptibility to RAF inhibition, ORFs that conferred resistance to RAF inhibition could be identified. Indeed, this approach confirmed the importance of MAPK reactivation for resistance to BRAF inhibitors in melanoma, and specifically identified COT as a driver of resistance.

This ORFeome collection has since been expanded to genome-scale and includes over 15,000 ORFs (Yang et al., 2011). Two features of this library facilitate use in many situations. Lentiviral delivery confers the ability to infect all cell types regardless of proliferative status, and a Gateway vector aids in versatility of cloning to different backbones with various tags or selection markers. In Chapter 4, we describe our usage of this library to probe RAS signaling.

1.4.5 Enabling integrative genomics

Rich datasets have been generated to systematically characterize cancer cells in terms of structural changes, expression changes, and functional response. Large-scale efforts have been made by many institutions to gather data from cancer cell lines as well as primary patient samples. Some of these ongoing efforts are summarized in Table 1-1.

Comprehensive datasets enable exciting opportunities to deeply interrogate signaling pathways in cancer. Over time, improvements in technology as well as improvements in computational methods help to make sense of complex data. By understanding underlying genetic aberrations in cancer, rational efforts can be made to thwart cancer progression. In this thesis, we illustrate advances in loss-of-function and gain-of-function genetic tools at the genome scale. Integrative use of these tools provides insight into novel oncogenic dependencies and allows us to probe more deeply into established oncogenic pathways.

Table 1-1. Databases that characterize genomic alterations in cancer cells. Deep characterization of cancer genomes provides a rich source of data for understanding cancer progression. These ongoing projects are often updated with new data.

	Genome Characterization	**Expression Characterization**	**Functional Characterization**	**Website**
Cancer Genome Project	Copy-number data for projected 800 cell lines and sequencing for selected gene mutations.		Compound sensitivity for some cell lines.	http://www.sanger.ac.uk/genetics/CGP/
National Cancer Institute NCI-60	60 cells lines deeply characterized by karyotyping, copy-number, and sequencing.	Gene expression profiling and protein lysate array profiling for all 60 cell lines.	Compound sensitivity data for all 60 cell lines.	http://dtp.nci.nih.gov/branches/btb/characterizationNCI60.html
The Cancer Genome Atlas	Data from 6000 patient tumor and normal tissues. Includes SNP, copy number, DNA methylation, and somatic mutations.	Gene expression and miRNA expression profiling.	Clinical outcome for a subset of patients.	http://cancergenome.nih.gov/
Cell Line Encyclopedia	Copy number and sequencing of selected gene mutations across several hundred cancer cell lines.	Gene expression profiling. Protein lysate array in progress.	Sensitivity to 24 cancer drugs across 504 cell lines. 102 cell lines functionally characterized by pooled RNAi screening.	http://www.broadinstitute.org/ccle/home
Cancer Gene Anatomy Project	SNP database from patient tissue.	Gene expression profiling.		http://cgap.nci.nih.gov/

CHAPTER TWO

Novel method for interrogation of genome-scale loss-of-function screens

This chapter has been adapted from:

Shao, D. D., Tsherniak, A., Gopal, S., Weir, B. A., Tamayo, P., Stransky, N., Schumacher, S. E., Zack, T.I., Beroukhim, R., Garraway, L.A., Margolin, A.A., Root, D.E., Hahn, W.C., Mesirov, J.P. (2013). ATARiS: Computational quantification of gene suppression phenotypes from multisample RNAi screens. *Genome Research*.

Contributions:

Diane Shao, Aviad Tsherniak, Shuba Gopal, Barbara Weir, Adam Margolin, Pablo Tomayo, David Root, William Hahn, and Jill Mesirov conceptualized the method (Fig. 2-1, 2-2). Shuba Gopal, Barbara Weir, Nicolas Stransky, Steven Schumacher, Levi Garraway, and David Root contributed datasets. Aviad Tsherniak implemented the method and analyzed high throughput qRT-PCR data (Fig. 2-5). D.D.S. performed computational analyses and experimental validation experiments.

2.1 Introduction

RNAi screening is a powerful approach that facilitates the systematic assessment of the effect of gene suppression on cell phenotypes such as cell death or the activity of a signaling pathway. The development and availability of genome-scale RNAi libraries provide the tools to identify new pathway components and context-specific cancer dependencies(Luo et al., 2008; Cheung et al., 2011a; Kittler and Pelletier, 2008; Hirsch, 2010). Technological and analytical advances will provide further opportunities for the application and interpretation of functional screens.

For screens in mammalian cells, a short interfering RNA (siRNA) is introduced into cells either directly as a duplex or by expression of a short hairpin RNA (shRNA) that is processed into active siRNA. This siRNA is designed to specifically degrade mRNA transcripts of complementary sequence to reduce the expression of gene products(Elbashir et al., 2001; Root et al., 2006). In practice, these reagents exhibit a variable degree of suppression of the targeted gene, and may also suppress genes other than the intended target(Jackson et al., 2003; 2006; Birmingham et al., 2006). Here we refer to a reagent's phenotypic effects resulting from suppression of unintended genes as *off-target effects*. Analytical approaches to identify specific types of off-target effects in siRNA (e.g., seed sequence similarity) have been previously developed(Marine et al., 2012; Sigoillot et al., 2012). However, we currently lack the ability to systematically characterize both the on-target and off-target effects of siRNAs.

To identify candidate genes that produce a desired phenotype based upon imperfect reagents, multiple distinct RNAi reagents targeting each gene are often screened(Echeverri et al., 2006; Cullen, 2006). Analyzing data from multiple reagents per gene has the potential to (1) increase the power to detect candidate genes, and (2) decrease the false-positive rates. For example, the "frequency approach" considers a

gene a candidate in a sample if several of its reagents induce a desired effect. This effect is usually measured by deviation from the experimental, or *de facto*, negative control effects(Chung et al., 2008; Bard et al., 2006; Müller et al., 2005). A variation of this approach is to assign a gene score by using a simple function, such as the average, of a few reagents with the most desired effects(Marcotte et al., 2012). More recent methods for scoring genes in individual samples, such as 'redundant siRNA activity'(RSA, König et al., 2007) and 'strictly standardized mean difference'(SSMD, Zhang et al., 2007; 2011a) have further decreased false-positive rates. For each sample, they consider the phenotypes produced by reagents for all of the screened genes simultaneously (RSA) or by all reagents for each gene separately (SSMD).

As RNAi screens are being performed in increasing numbers of samples(Cheung et al., 2011a; Brough et al., 2011; Collinet et al., 2010; Marcotte et al., 2012), a common analytical approach has been to segregate samples into two pre-defined classes in order to identify genes with differential effects. By summarizing data within each class, aberrant reagent effects in individual samples are less likely to impact the final result. The "second best" method assigns scores to genes based on each gene's second most differentially scoring reagent between classes(Cheung et al., 2011a), requiring – similar to the "frequency approach" – that favorable genes have at least two high-scoring reagents. Alternatively, RNAi Gene Enrichment Ranking (RIGER) ranks all the reagents by their differential effects and generates a gene-level score for each gene based on the rank distribution of its reagents(Luo et al., 2008; Barbie et al., 2009; 2009), analogous to RSA. However, the requirement of two pre-defined classes can limit full interrogation of the data.

Currently, RNAi analysis methods do not attempt to assess the performance of individual reagents. Thus, there is an opportunity to further improve analysis of RNAi data by harnessing the statistical information across many samples to identify and avoid

data from problematic reagents when determining gene-level effects. An analogous approach is employed by dChip(Li and Hung Wong, 2001) and RMA(Irizarry et al., 2003), two widely used methods for mRNA abundance quantification in microarray data. Given a set of samples, these algorithms quantify a probe set's overall abundance level in each sample from a set of multiple, distinct, complementary probes. In the case of RNAi data, one must also consider additional factors such as off-target effects thought to exist for a subset of reagents, the dramatically greater biological variability, and the possibility of multiple phenotypic effects for a single gene (e.g., due to different levels of on-target gene suppression).

Here we introduce ATARiS (Analytic Technique for Assessment of RNAi by Similarity), a novel computational approach to the quantification of gene-specific suppression phenotypes. ATARiS uses patterns in the data from multi-sample RNAi screens to estimate the performance of individual RNAi reagents targeting each gene and generates a per-gene value for each sample that quantifies the phenotypic effect of gene suppression. A schematic comparison of ATARiS to other methods of RNAi analysis is made in Fig. 2-1. We used data from two recent large-scale shRNA screens of 102 and 72 cancer cell lines(Cheung et al., 2011a; Marcotte et al., 2012), respectively, to demonstrate the performance of ATARiS.

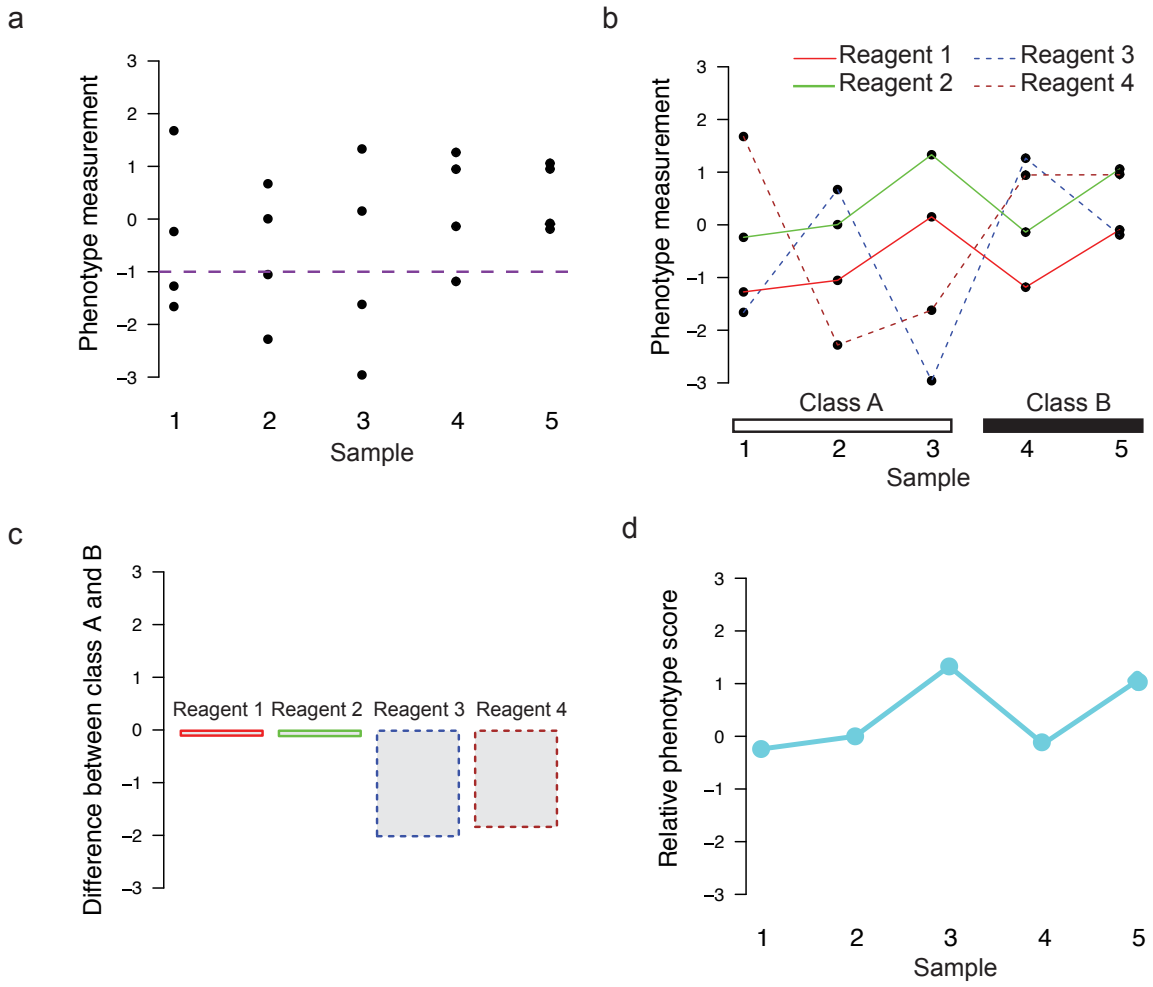

Figure 2-1. ATARiS accounts for patterns in RNAi reagent data in order to quantify phenotypic effect of gene suppression in each sample. Hypothetical phenotypic data from four RNAi reagents, all designed to target the same gene, in five independent samples from two classes A and B. (**a**) Samples 1, 2, and 3 each have at least two reagents that score below a desired threshold (purple dotted line), thus, according to "frequency approach" methods, this gene may be a "hit" in those samples. (**b**) A line connecting each reagent's effects across the samples reveals additional information. Specifically, we note that it is possible (as in this scenario) that different shRNAs drive the determination of hits in each sample when samples are each analyzed separately as in (a). (**c**) For each reagent, the difference between its mean values in class A and class

Figure 2-1 (Continued).

B is shown, reducing much of the noise from individual samples. Reagents 3 and 4 both show differential effects between the classes and would suggest that two-class-based analytic methods select this gene as a hit. (**d**) ATARiS phenotype values for each of the screened samples. Phenotype values represent relative gene-level effects in each individual sample by incorporating information from trends across all samples, favoring reagents that produce correlated effects (i.e., reagents 1 and 2 from (b)). If the user chooses to assess whether differential effects exist between class A and B, this example would show no significant difference by avoiding uncorrelated reagents 3 and 4.

2.2 ATARiS: Computational method overview

ATARiS is a computational method to assess gene suppression effects in each sample of multi-sample RNAi screens that include at least two RNAi reagents (siRNA or shRNA) designed to target each gene. Our method uses only data from reagents determined to have primarily on-target effects, discarding data from reagents with off-target effects. To identify on-target reagents, we noted that in an RNAi library, reagents are designed to target distinct sequences. Thus, it is unlikely that any two reagents – including those targeting the same gene – will suppress the same set of off-target genes. We therefore concluded that when RNAi reagents designed to target the same gene behave similarly *across the screened samples*, the observed effects are likely due to suppression of the intended gene rather than off-target suppression. For each gene in a screen, ATARiS identifies sets of reagents with similar behavior across all samples in order to produce two types of results:

1. A ***gene solution*** that summarizes the observed effects produced by identified on-target reagents into quantitative values across all screened samples (the value for an individual sample is called a *phenotype value*). We account for potential multiple phenotypic outcomes after suppression of a given gene, possibly due to different degrees of gene suppression, by allowing for multiple solutions comprised of disjoint sets of consistent reagents.
2. A ***consistency score*** for each RNAi reagent that represents the confidence that the reagent's observed phenotypic effects are the result of on-target gene suppression. ATARiS assigns higher consistency scores to reagents whose profiles (i.e., the observed effect of that reagent in every screened sample) exhibit higher correlation to a larger number of reagent profiles within the same solution.

We give a general description of ATARiS here (see schematic Fig. 2-2), and provide technical details in Appendix 1.

First, to construct a *gene solution* for a given gene *G*, ATARiS considers the observed data of all RNAi reagents designed to target *G*. The data for each reagent *r* are median-centered, as we are interested in reagents whose *relative* effects across the samples are similar. For each sample *s*, ATARiS calculates a value c_s that summarizes the effects produced by all the reagents targeting *G* in *s*. We refer to the vector **c** of all c_s values as the *consensus profile*. To estimate c_s, ATARiS models each data point $x_{r,s}$ (i.e., the observed effect induced by reagent *r* in sample *s*) as a product of two unknown quantities: e_r, representing the relative magnitude of the effects of reagent *r,* and c_s. ATARiS estimates the values for e_r and c_s by minimizing an L_1-norm objective function using the method of alternating minimizations(Csiszar and Tusnady, 1984). An L_1 norm makes the optimization more robust to outliers, which are common in this type of data.

Next, ATARiS iteratively refines the considered set of reagents by evaluating the similarity of each reagent profile to the consensus profile. If, for any reagent profile, the significance of the Spearman correlation (calculated using an empirical null distribution) is lower than a pre-defined threshold, the reagent whose profile is least similar to the consensus profile is discarded from further analysis. ATARiS then repeats the process of computing a consensus profile and discarding dissimilar reagents until either only one reagent remains — in which case no gene solution is generated — or until all remaining reagents have profiles significantly similar to the consensus profile. The consensus profile **c** for the retained reagents is then used as a *gene solution* for that gene, and we refer to its elements c_s as the gene's phenotype values. The entire process is then repeated for any remaining reagents not yet contributing to a solution until no more solutions are found. A greedy approach to refinement, rather than an exhaustive one, allows scaling to larger numbers of reagents per gene.

After generating all gene solutions for gene *G*, ATARiS computes a consistency score for each of its reagents. The consistency score for reagent *r* is based on the negative log_{10} of the integrated *p*-values of the Spearman correlation coefficients of *r*'s profile to each of the other reagent profiles within the same solution. Thus, the consistency score may be interpreted as a *p*-value, i.e., a consistency score of 1.3 corresponds to $-log_{10}$ *(p-value of 0.05)*. For RNAi reagents that do not participate in any solution, *all* reagents targeting *G* are considered in computing the consistency score. Thus, even for a reagent that is excluded from a solution depending on the pre-defined threshold, ATARiS still provides an assessment of the confidence in its functional effects.

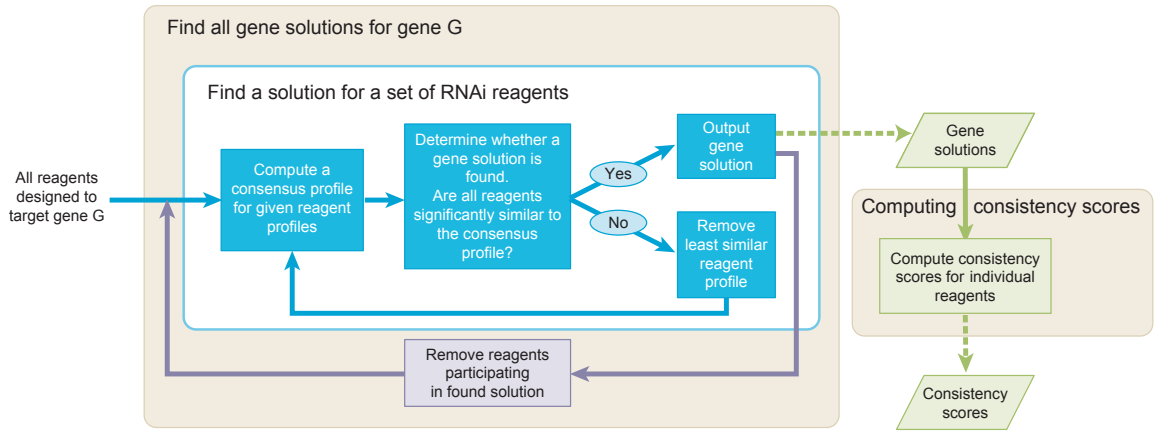

Figure 2-2. A schematic diagram of the ATARiS algorithm.

2.3 Experimental methods

Analysis of Project Achilles dataset

Genome-scale pooled shRNA screens to identify genes essential for proliferation in 102 cancer cell lines were performed using a lentivirally delivered pool of 54,020 shRNAs targeting 11,217 genes(Cheung et al., 2011a). Each cell line was infected in quadruplicates and propagated for at least 16 population doublings. The abundance of shRNA constructs was measured by microarray hybridization and raw .CEL files from custom Affymetrix barcode arrays were processed with a modified version of dCHIP software. ShRNAs that had an overlap of more than 3 base pairs to other screened shRNAs were removed (n=679). The log_2 fold change in shRNA abundances for each cell line at the conclusion of the screening relative to the initial plasmid DNA reference pool was calculated(Cheung et al., 2011a). The log_2 fold change data were then normalized by a robust Z-score normalization (i.e., centering around the median and scaling by the Median Absolute Deviation). The median value was used to collapse data from replicates. Each data point represents the abundance of one shRNA construct within one cell line as compared to the initial abundance of that shRNA construct in the initial plasmid DNA pool (See supplementary data from Shao et al., 2013). The ATARiS analysis ignored data for shRNAs targeting non-human genes (n=4) and genes targeted by only one shRNA (n=8).

Analysis of Marcotte et al. dataset

Marcotte et al. (2012) performed genome-wide pooled shRNA screens to identify genes essential for cancer cell survival and proliferation in 72 breast, pancreatic, and ovarian cancer cell lines (Marcotte et al., 2012). They used a lentiviral shRNA library targeting ~16,000 genes with 78,432 shRNAs, of which 50,981 shRNAs were also used

in the Achilles screens. We obtained shRNA-level shARP (shRNA Activity Ranking Profile) scores for all the shRNAs and cell lines screened through the COLT-Cancer database (Koh et al., 2012) and considered them to represent the observed phenotypic effects. We computed a robust Z-score for each cell line separately and discarded data for two cell lines (OVCA1369_TR, HPDE) that showed aberrant score distributions. We ran ATARiS on the normalized values using the same parameters used for the analysis of the Achilles dataset. ATARiS found gene solutions for 8406 (54.4%) of the genes using data from 29731 (39.2%) of the shRNAs.

Two-Class comparisons

For each 2-class comparison, ATARiS gene phenotype values were used to calculate a mean for each class for each gene solution. The difference of means between the classes was used as a scoring metric and *p*-values were estimated based on a null distribution generated by 50,000 class label permutations. *Q*-values were generated by the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

Annotation of cell line genomic features

We constructed a matrix of genomic features for cell lines that had matched genomic data from the Cancer Cell Line Encyclopedia (CCLE)(Barretina et al., 2012). As previously described, features include mutational status, tumor tissue lineage, regions of recurrent copy-number gain or loss (derived from GISTIC), and combined gene mutation and copy number amplification (for oncogenes) or combined mutation and copy number deletion (for tumor suppressors). GISTIC regions were assessed across all available CCLE cell lines, of which 76 were screened in Achilles. Amplification and deletion of specific genes were defined by relative log fold copy-number value greater than 0.25 or less than -0.25, respectively. All data are represented as binary values, with 1

representing presence of the indicated feature in the sample. See supplementary data in Shao, et al. (2013) for full feature matrix.

Cell culture

All cancer cell lines were cultured in RPMI-1640 (Invitrogen) supplemented with 10% FBS (Sigma) and 1% streptomycin and penicillin.

Low-throughput assessment of cell viability

Cells were replated at 50,000 cells/well post-infection and post-selection in triplicate in 12-well plates post-selection with puromycin. Wells were counted 4 days later by ViaCell.

Lentiviral infection

Lentivirus containing shRNA targeting *BRAF*, *PIK3CA*, *KRAS*, *MYC*, and controls targeting *GFP* or *LacZ* for validation of ATARiS consistency scores were purchased directly from The RNAi Consortium (Root et al., 2006) for gene suppression validation studies. Lentivirus for *KRAS* dependency experiments were produced as previously described (Barbie et al., 2009). Cells were infected in media containing 8ug/ml polybrene and 1:10 dilution of virus. Infected cells were selected with 2ug/ml puromycin for 48 hours. ShRNA identities are provided in Table 2-1.

Table 2-1. Identities of shRNA used in Chapter 2.

shRNA	**TRC Identifier**	**NM number**	**Target** (5'-3')
shKRAS-1	TRCN0000033263	NM_033360.2-269s1c1	GACGAATATGATCCAACAATA
shKRAS-2	TRCN0000033260	NM_033360.2-407s1c1	GAGGGCTTTCTTTGTGTATTT
shKRAS	TRCN0000033262	NM_033360.2-509s1c1	CCTATGGTCCTAGTAGGAAAT
shKRAS	TRCN0000033261	NM_033360.2-667s1c1	GATCCGACAATACAGATTGAA
shKRAS	TRCN0000040149	NM_004985.3-641s1c1	GATGCCTTCTATACATTAGTT
shKRAS	TRCN0000040148	NM_004985.3-3896s1c1	CCTCGTTTCTACACAGAGAAA
shKRAS	TRCN0000018337	NM_004985.x-204s1c1	TAGTTGGAGCTGGTGGCGTAG
shKRAS	TRCN0000010369	NM_004985.x-1160s1c1	CAGTTGAGACCTTCTAATTGG
shKRAS	TRCN0000040150	NM_004985.3-570s1c1	CTCAGGACTTAGCAAGAAGTT
shKRAS	TRCN0000040152	NM_004985.3-492s1c1	AGGACTCTGAAGATGTACCTA
shKRAS	TRCN0000040151	NM_004985.3-297s1c1	CTACAGGAAGCAAGTAGTAA
shKRAS	TRCN0000033259	NM_033360.2-4328s1c1	GCAGACGTATATTGTATCATT
shBRAF-1	TRCN0000006293	NM_004333.2-304s1c1	CTATGAAGAATACACCAGCAA
shBRAF-2	TRCN0000006292	NM_004333.2-1538s1c1	CAGCAGTTACAAGCCTTCAAA
shBRAF-3	TRCN0000006291	NM_004333.2-2267s1c1	GCTGGTTTCCAAACAGAGGAT
shBRAF-4	TRCN0000006290	NM_004333.2-838s1c1	CCGCTGTCAAACATGTGGTTA
shBRAF-5	TRCN0000006289	NM_004333.2-1106s1c1	GCAGATGAAGATCATCGAAAT
shPIK3CA-1	TRCN0000010407	NM_006218.x-3234s1c1	AATGAAAGCTCACTCTGGATT
shPIK3CA-2	TRCN0000039607	NM_006218.1-2145s1c1	GCTCATTAACTTAACTGACAT
shPIK3CA*	TRCN0000039603	NM_006218.1-3251s1c1	GATTCCACACTGCACTGTTAA
shPIK3CA-3	TRCN0000039604	NM_006218.1-2368s1c1*	CCAGACATCATGTCAGAGTTA
shPIK3CA-4	TRCN0000039606	NM_006218.1-924s1c1	GCCATCTTATTCCAGACGCAT
shPIK3CA-5	TRCN0000039605	NM_006218.1-1057s1c1	CGAGACATTGACAAGATTTAT
shMYC	TRCN0000010391	NM_002467.x-1970s1c1	CAACCTTGGCTGAGTCTTGAG
shMYC	TRCN0000039638	NM_002467.2-1828s1c1	CCATAATGTAAACTGCCTCAA
shMYC	TRCN0000039639	NM_002467.2-1552s1c1	CCCAAGGTAGTTATCCTTAAA
shMYC	TRCN0000039642	NM_002467.2-1377s1c1	CCTGAGACAGATCAGCAACAA
shMYC	TRCN0000039641	NM_002467.2-408s1c1	CAGGAACTATGACCTCGACTA
shMYC	TRCN0000039640	NM_002467.2-1657s1c1	AATGTCAAGAGGCGAACACA
shGFP	TRCN0000072181	clonetechGfp_437s1c1	ACAACAGCCACAACGTCTATA
shLacZ	TRCN0000072231	lacZ_1650s1c1	CGCTAAATACTGGCAGGCGTT

*No data. Virus obtained from the RNAi platform for this shRNA could not infect cells.

Immunoblotting

Cell lysates collected 72 hours post-infection were run on 4-12% Bis-Tris gel (Invitrogen NuPAGE) and transferred to nitrocellulose membrane for immunoblotting. Primary antibodies were obtained from Santa Cruz (KRAS sc-30, BRAF sc-5284, MYC sc-764, ß-actin sc-1615) and Cell Signaling (PI3Kinase 110alpha #4255). Immunoblots for BRAF and PIK3CA protein were visualized by infrared imaging (LICOR). Quantification was performed by ImageJ software (http://rsb.info.nih.gov/ij).

Data Access

ATARiS can be run online on user-provided data through the GenePattern computational genomics suite (Reich et al., 2006) accessible on the ATARiS website (http://broadinstitute.org/ataris). The website also includes all datasets used to obtain the results described in this manuscript.

2.4 Validation of ATARiS results

2.4.1 Datasets used to validate ATARiS

To test and validate ATARiS, we primarily used the data produced by Project Achilles – a dataset produced from massively parallel screening of 102 cancer cell lines with a genome-scale pooled shRNA library targeting more than 11,000 human genes with an average of 5 shRNAs per gene (Cheung et al., 2011a). The final abundance of each shRNA after propagation of the cell line was determined with respect to the initial reference shRNA pool to assess cellular dependency on each shRNA's target (i.e., shRNAs that target essential genes will be depleted). See Chapter 2.3: Experimental methods for a full description of additional data pre-processing and normalization steps.

The application of ATARiS to this dataset yielded gene solutions for 7,250 genes, and incorporated data from 49.5% of the screened shRNA reagents when using a 0.15 significance threshold. With this threshold, we would expect ATARiS to generate solutions, on average, for 15% of the genes using randomly permuted data. In our dataset, 6,233 genes had one associated gene solution, 1,017 genes had two or more solutions, and 3,955 genes had no solutions (Table 2-2). We illustrate the type of correlated reagent data that becomes incorporated into a gene solution by using an example gene *BRAF* (Fig. 2-3). A consistency score was generated for every screened shRNA, including those that do not participate in any gene solution.

We evaluated whether ATARiS results are robust when applied to samples screened independently of Project Achilles by analyzing results from a set of independently performed genome-scale pooled shRNA viability screens (Marcotte et al., 2012). We used the RNAi screening data from Marcotte et al. (2012), comprised of 72 cancer cell lines screened using a comparable shRNA library. For genes that have a solution in both datasets, we found that the shRNAs targeting those genes are more likely to participate in a solution in both datasets than in one dataset but not in the other (odds ratio = 2.1; 95% confidence interval [1.96, 2.26]; *p*-value < 2.2×10^{-16}; Fisher's exact test). Furthermore, the Pearson correlation coefficient for ATARiS shRNA consistency scores in the two datasets is 0.46 (95% confidence interval [0.45, 0.47]; *p*-value < 2.2×10^{-16}). These observations suggested that ATARiS produces robust results between independent screens.

Table 2-2. Results for genes calculated from Achilles RNAi dataset. We account for all genes screened in terms of number of shRNAs used to target that gene and the resulting number of gene solutions identified by ATARiS.

		Number of shRNAs per gene															
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**19**	**20**	**Totals**
Number of gene solutions per gene	**0**	53	256	1041	2577	12	8	0	3	4	1	0	0	0	0	0	0	**3955**
	1	24	230	1346	4549	20	24	14	10	12	2	0	0	1	0	1	0	**6233**
	2	0	0	98	844	5	14	7	9	18	4	3	1	0	2	0	0	**1005**
	3	0	0	0	0	0	0	2	1	5	1	0	0	1	0	0	1	**11**
	4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	**1**
																		11205

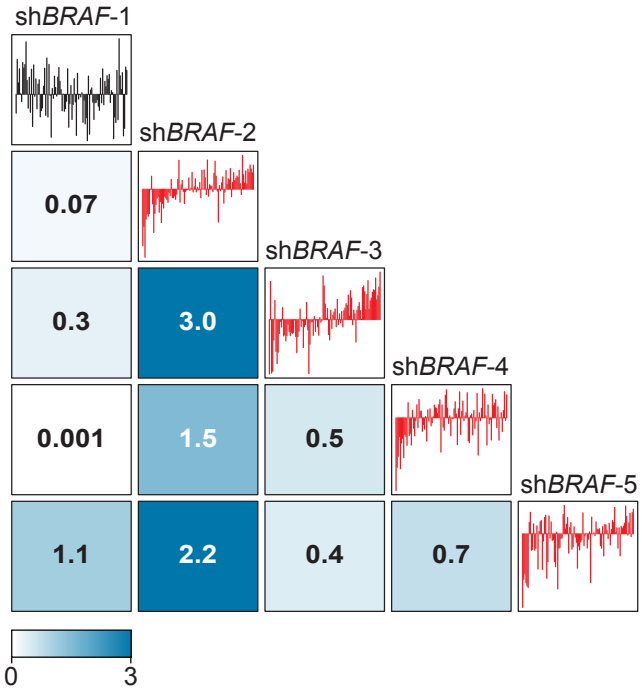

Figure 2-3. *BRAF* example for shRNA data integration into gene solutions. Data from the Achilles dataset for shRNAs targeting *BRAF*. Median-normalized screening data across 102 samples are displayed as barplots in sample order of ascending *BRAF* phenotype value. Boxed numbers display $-log_{10}$ *p*-value of the Spearman correlation coefficient for the two shRNA labeled in the corresponding row and column. Red, shRNAs with correlated effects that are incorporated into the *BRAF* gene solution.

2.4.2 Validation of shRNA consistency scores

ATARiS shRNA consistency scores are intended to reflect our confidence in the specificity of each reagent. However, validation is challenging since the currently accepted standards for evaluating reagent performance, i.e., immunoblotting and quantitative RT-PCR for on-target gene suppression, cannot assess off-target effects whereas ATARiS consistency scores attempt to encompass both on- and off-target aspects. Since a greater degree of target gene suppression does not necessarily amplify functional outcome, and effective on-target gene suppression does not equate to lack of off-target effects, we did not expect high correlation between immunoblotting results and ATARiS scores. We expected, however, that shRNAs with high consistency scores have some degree of detectable on-target gene suppression in order to produce correlated profiles.

We validated consistency scores on a few selected genes – *BRAF*, *PIK3CA*, *KRAS*, and *MYC* – chosen for their importance in cancer, availability of reagents to assess the expression of these genes, and known functional effect of their shRNAs in a subset of Project Achilles cell lines (Cheung et al., 2011a). We introduced individual shRNAs into the A549 cancer cell line and performed immunoblotting on cell lysates to determine changes at the protein level. For *BRAF*, we observed that ATARiS consistency scores are high for shRNAs that reduce BRAF protein levels (Fig. 2-4A). We note that for sh*BRAF*-3 and sh*BRAF*-4, which have similar consistency scores but different degrees of protein suppression, 40% protein suppression may be sufficient to produce functional effects, and the effects may not be enhanced by increased protein suppression. For *PIK3CA*, only two shRNAs (sh*PIK3CA*-1 and sh*PIK3CA*-2) have high consistency scores and both result in increased suppression of PIK3CA protein levels (Fig. 2-4b). Our interpretation for the low consistency score of sh*PIK3CA*-3, which

effectively suppresses PIK3CA at the protein level, is that it may also have significant off-target effects. For *KRAS* and *MYC*, the effects of expressing individual shRNAs on protein levels also agreed with ATARiS consistency scores (Fig. 2-4c, 2-4d).

To test whether consistency scores reflect on-target gene suppression for many more genes, we compared ATARiS consistency scores to gene suppression assessed by qRT-PCR for 9,050 of the shRNAs from the screening library (manuscript in preparation, data available on request). We found that shRNAs with significantly high consistency scores (corresponding to FDR < 0.1) suppress target gene mRNA levels to a greater degree than other shRNAs targeting the same gene (*p*-value = 0.003, $\chi 2$ test; Fig. 2-5). We therefore concluded that genes with high consistency scores are likely to have a functionally relevant degree of gene suppression.

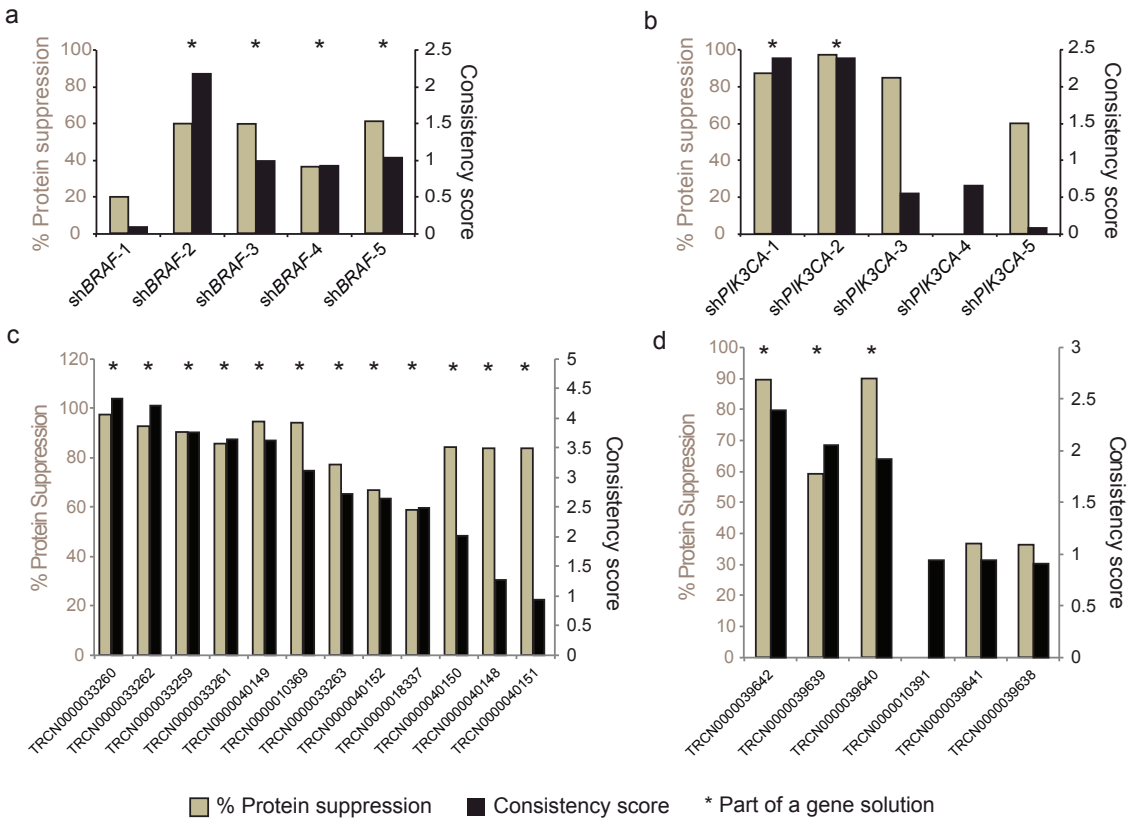

Figure 2-4. ATARiS shRNA consistency scores are associated with on-target gene suppression. Consistency scores computed by ATARiS and corresponding protein suppression levels by immunoblotting are shown for shRNAs targeting (**a**) *BRAF* (**b**) *PIK3CA*, (**c**) *KRAS*, and (**d**) *MYC*. A higher consistency score represents greater confidence that the effects produced by the shRNA are due to suppression of the target gene. Immunoblotting for the effect of each shRNA compared to control shRNA was performed in cell line A549 and percent suppression compared to control shRNA was calculated based upon quantification by ImageJ software. Shading of axis labels correspond to data bars of the same type. *Asterisks indicate reagents used in the gene's ATARiS gene solution.

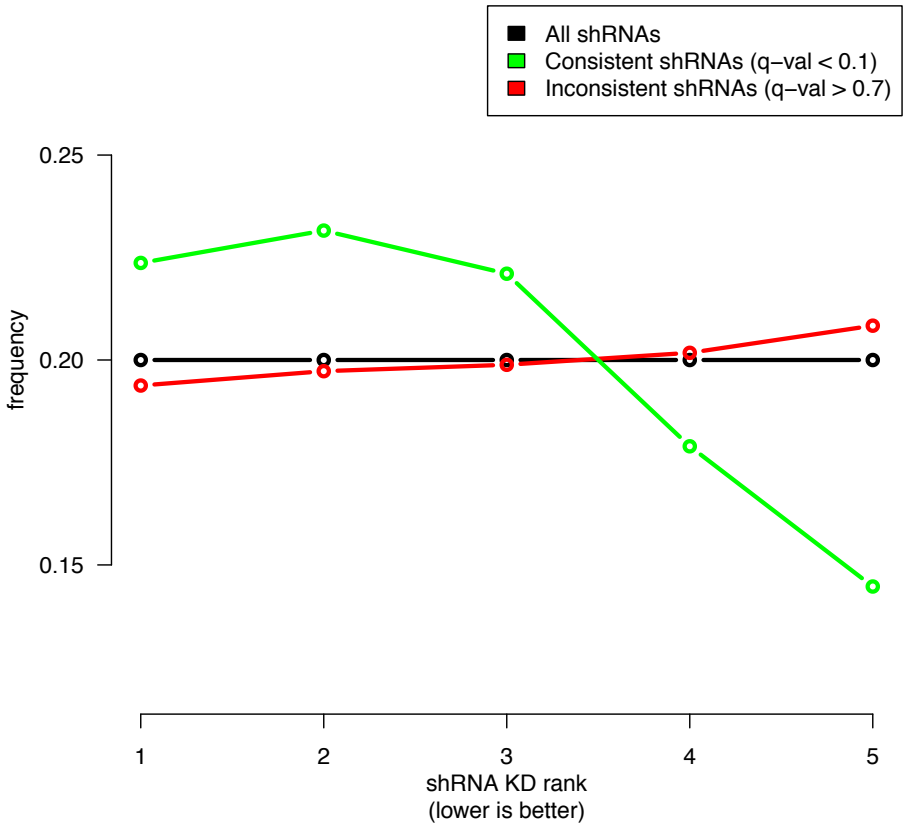

Figure 2-5. On-target gene suppression measured by high-throughput qRT-PCR is associated with higher consistency scores. Using high-throughput qRT-PCR data of shRNAs (manuscript in preparation) we analyzed screening data for genes with exactly five shRNAs with high confidence qRT-PCR data (n=9,050 shRNAs). For each gene, we ranked the level of mRNA suppression of each of its shRNAs from 1 to 5 (1, most suppressed; 5, least suppressed), and assessed the frequency of each rank for those shRNAs predicted to perform well by ATARiS (consistency score q-value < 0.1). For comparison, we show the frequency of mRNA ranks when using shRNAs that have low consistency scores (consistency score q-values > 0.7), or all shRNA.

2.4.3 Validation of ATARiS gene phenotype values

We first validated individual ATARiS gene phenotype values, representing degree of dependency on each gene in the Achilles dataset, by assessing whether they recapitulate known dependencies for the oncogenes *BRAF*, *PIK3CA*, and *KRAS*. For each oncogene, we calculated the area under Receiver Operating Characteristic curve (AUC) statistic (Mason and Graham, 2002) to measure the degree to which the gene phenotype values discriminate between cell lines harboring a mutation versus those without mutation. We confirmed that cell lines harboring a mutation have significantly lower phenotype values for the respective gene, i.e., are more sensitive to gene suppression (Fig. 2-6a; p-value < 0.01; Mann-Whitney test). To show that our phenotype values may be meaningful for individual cell lines, we performed low-throughput viability assays on cell lines that span a range of *KRAS* phenotype values. We introduced two *KRAS*-specific shRNA or a control shRNA into three *KRAS* wild-type and three *KRAS* mutant cell lines and measured cell proliferation/viability after six days using an ATP-luminescence assay. Indeed, the cell lines most sensitive to KRAS suppression were the ones that received the lowest *KRAS* phenotype values (Fig. 2-6b). Thus, we affirmed that ATARiS phenotype values reflect the relative effects of gene suppression between individual samples.

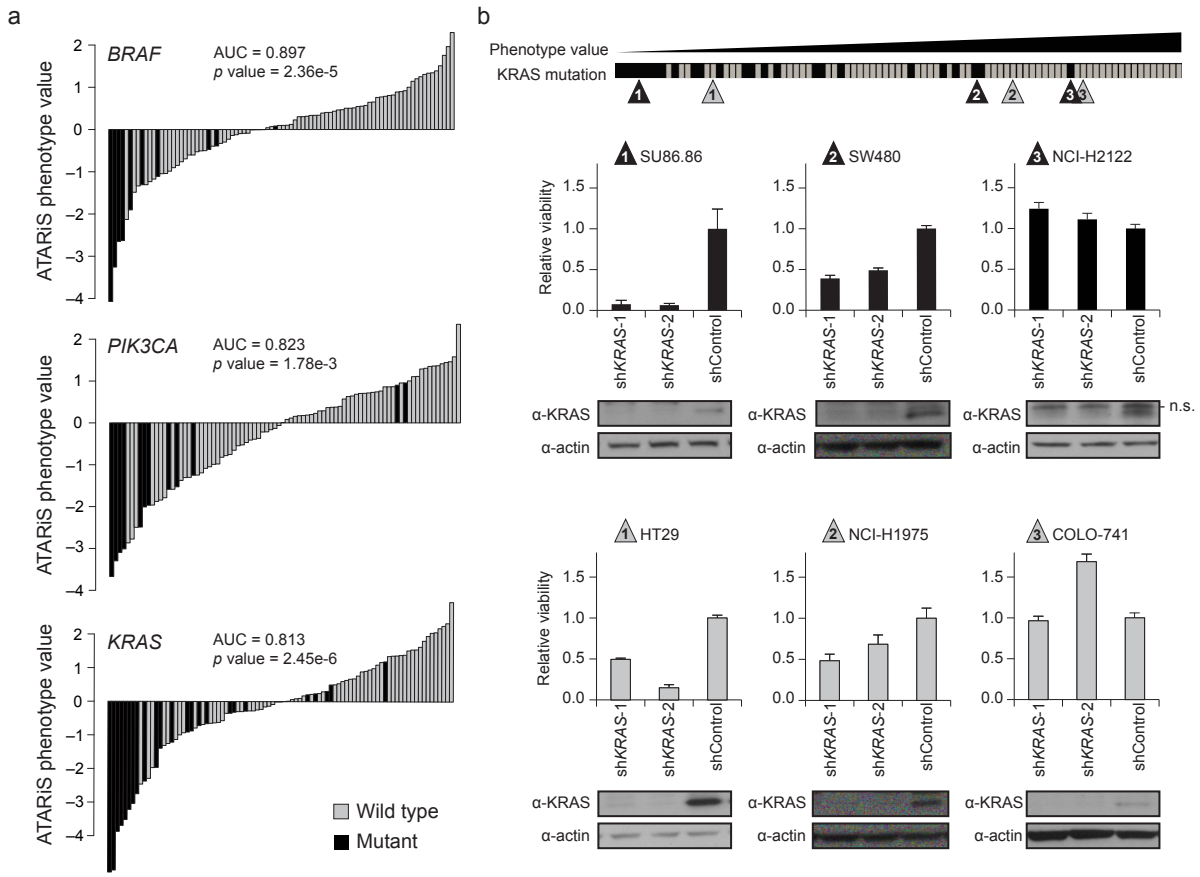

Figure 2-6. ATARiS gene phenotype values reflect biological dependencies. (**a**) Correspondence between gene mutation status and ATARiS phenotype values for *BRAF, PIK3CA,* and *KRAS*. Each vertical bar represents a single screened sample, colored by mutation status. In each plot, samples are ordered by increasing phenotype values. AUC, area under receiver operating characteristic curve. *P*-value, assessed by Mann-Whitney test. (**b**) Low-throughput validation of the relationship between gene phenotype scores and gene dependency. Immunoblot lanes correspond to bars in the graph directly above. Horizontal bar orders all cell lines with known *KRAS* mutation status in increasing order by ATARiS phenotype value, with validated samples marked by corresponding triangles. Grey, *KRAS* wild-type. Black, *KRAS* mutant. Error bars, ± one standard deviation (*n*=3). n.s., non-specific band.

We reasoned that if ATARiS solutions are meaningful, then we should be able to "rediscover" the above oncogenic dependencies. Two-class comparisons between groups of cell lines with defined properties are currently a common application of this type of RNAi proliferation screen data, so we defined classes based on mutation status for each of *BRAF*, *PIK3CA*, and *KRAS* to identify differentially required genes. For each analysis, we calculated the mean difference between mutant versus wild-type classes for each gene solution and estimated *p*-values from an empirically calculated null distribution by class permutation. *KRAS*, *BRAF*, and *PIK3CA* are each ranked 1st for being differentially required in their respective mutant class and remained significant after Benjamini-Hochberg adjustment for multiple hypothesis testing (Benjamini and Hochberg, 1995)(*q*-value < 0.25, Fig. 2-7a). The fact that each class comparison yielded statistically meaningful results lends more validity to ATARiS phenotype values overall.

Finally, we sought to show that ATARiS phenotype values are valid for more than the specific oncogenes described above by defining classes using recurrent genomic alterations. Since commonly amplified or deleted regions in cancer are believed to include drivers that require unique cellular networks, we reasoned that more genes should be differentially essential when classes are defined by significant genomic alteration than when defined randomly. We defined significantly amplified and deleted peaks based on application of the Genomic Identification of Significant Targets in Cancer (GISTIC) method (Beroukhim et al., 2007) to copy number data from the Cancer Cell Line Encyclopedia (CCLE)(Barretina et al., 2012), a large collection of cancer cell lines annotated with genomic data (Table 1-1), of which 76 were screened in Project Achilles. For each peak present in at least six Achilles cell lines, we defined two classes based on the peak's presence or absence (101 total peaks), and calculated the difference in means between classes for every ATARiS gene solution to identify differentially essential genes (see Chapter 2.3: Experimental methods). Only 5% of

analyses using randomly defined classes yielded more than 4 significantly differential genes while 16% of analyses using GISTIC peaks do. We showed that significantly more essential genes were in classes defined by GISTIC peaks compared to random classes (*p*-value = $6x10^{-6}$, Wilcoxon rank sum; Fig. 2-7b), supporting the idea that ATARiS gene phenotype values likely reflect underlying biology.

a

Classes	Gene	Rank	P.Val	Q.Val
KRAS Mutation	*KRAS*	1	2.00E-05	0.053
BRAF Mutation	*BRAF*	1	2.00E-05	0.158
PI3Kinase Mutation	*PIK3CA*	1	2.00E-05	0.147

b

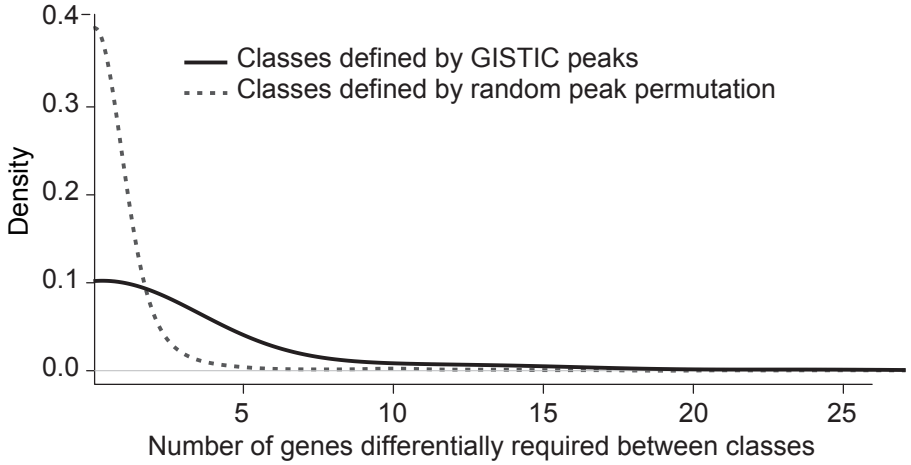

Figure 2-7. Gene solutions confirm known oncogenic dependencies and identify statistical enrichment in gene dependencies at recurrently altered genomic loci. (**a**) Rank of dependency phenotype value in two-class comparison by mutation status for common oncogenes. (**b**) Genes differentially required in sample classes defined by recurrent amplification or deletion peaks. Recurrent genomic peaks were identified by GISTIC analysis across genomic data for samples from the Cancer Cell Line Encyclopedia. For each peak existing in at least six samples (*n*=101), classes were defined based on presence or absence of the peak. Genes that are differentially required in samples harboring the peak (FDR < 0.25) were determined. The distribution of the number of significantly differential genes is shown. For comparison, the same analysis was performed using classes defined by random permutation of peak assignments.

2.3.4 Application of ATARiS for biological discovery

In Figure 2-7, we showed that ATARiS gene phenotype values could be effectively used for two-class comparisons, the focus of current analytic methods for multi-sample RNAi data. Additionally, ATARiS per-sample phenotype values expand the repertoire of downstream analyses from comparison between two classes to a range of additional possibilities. These include integrated analysis with other types of genomic data, e.g., gene expression, mutations, genome copy number, which provide quantitative information for genes in each sample. We describe illustrative examples here.

1. *Using phenotype values to identify genomic predictors of gene dependency.* Genes that control the cell cycle G1 restriction point are commonly altered in the cancer genome. Thus, we focused on ATARiS solutions for E2F transcription factors, well characterized in checkpoint regulation, to determine whether we could identify known (and unknown) genetic alterations related to E2F activation. We used an annotated sample feature list that includes significant amplification and deletion peaks, cell lineage, mutation, and copy number alterations of major oncogenes/tumor suppressors (see Chapter 2.3: Experimental methods). When we examined *E2F3*, we found that *E2F3* dependence is significantly associated with an *E2F3*-containing amplification peak 6p22, a *MYC*-containing amplification peak at 8q24.1, as well as *RB1* copy-number loss (Fig. 2-8) — all mechanisms that lead to *E2F3* activation(Leone et al., 2001; Dyson, 1998; Oeggerli et al., 2006). Similar analyses can be applied to other gene phenotype scores to elucidate genomic relationships with functional data.
2. *Identifying functional relationships between gene phenotype scores.* We evaluated the potential of using correlations between different gene solutions to yield functionally meaningful gene relationships. In particular, we focused on the

ATARiS solution for Cyclin D1 (*CCND1*) since this gene serves as critical mediator between the mitogenic pathway and cell cycle progression. The genes whose solutions most significantly correlated to *CCND1* solution included a gene that encodes the CCND1 binding partner CDK6 (no solution was available for the other cyclin D binding partner CDK4) and also members of the mitogenic pathway, including *KRAS* and *RAF1* (Musgrove et al., 2011; Liu et al., 1995, Fig. 5b)(Fig. 2-9). Thus, we showed that using statistical relationships between ATARiS phenotype values allowed us to assess functional gene networks.

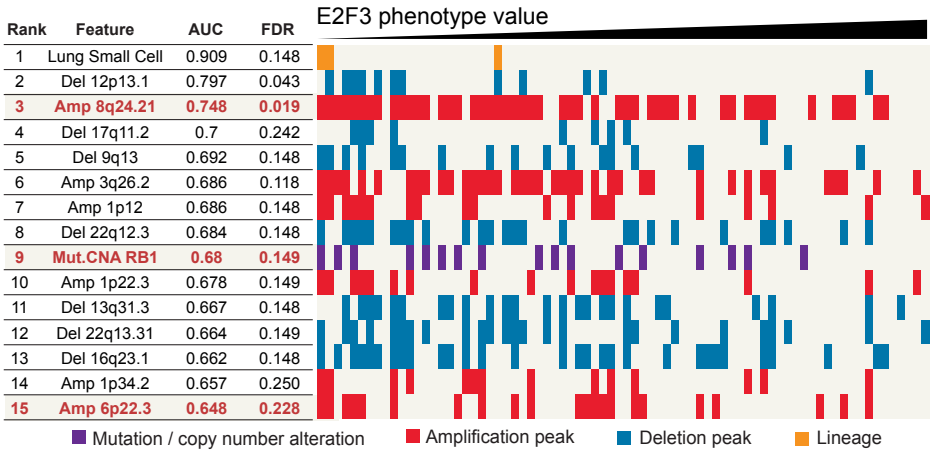

Figure 2-8. E2F phenotype values predict genomic features associated with dependency. Genomic features are shown ranked by their correspondence to E2F3 phenotype values as measured by area under receiver operating characteristic curve (AUC). Amplification and deletion peaks, as determined by GISTIC, are denoted by "Amp" and "Del," respectively. Columns correspond to individual cell lines.

Rank	Solution	Correlation	FDR
1	CCL8_1	0.590	0.010
2	KRAS_2	0.566	0.010
3	64748_1	0.565	0.010
4	KRAS_1	0.557	0.010
5	CRX_1	0.543	0.010
6	RPS11_1	0.543	0.010
7	Gpr12_1	0.538	0.012
8	CDK6_2	0.530	0.012
9	RAF1_1	0.524	0.010
10	HIATL1_1	0.516	0.012

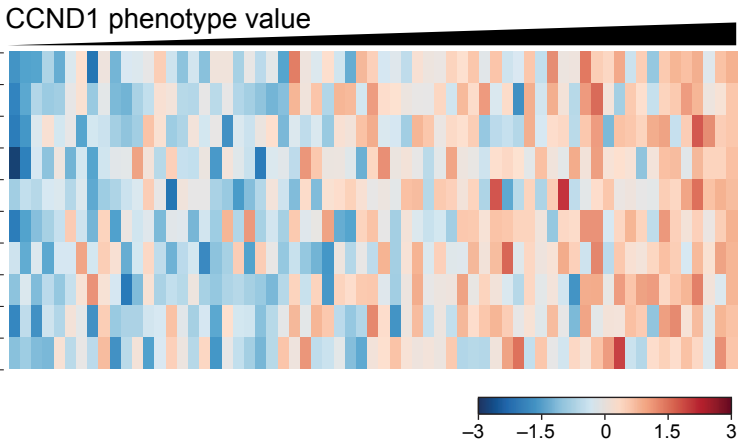

Figure 2-9. Correlations between gene phenotype value profiles to CCND1 gene solution. Gene solutions are ranked by their similarity to the CCND1 gene solution using Pearson correlation coefficient. *P*-values were generated by permutation of sample labels. Numbers following gene names in indicate gene solution number.

2.5 Discussion

One key advance of ATARiS lies in the ability to distinguish reagents with on-target effects and reject reagents with significant off-target effects by mining patterns across multi-sample screens. ATARiS reagent consistency scores may be interpreted as a *p*-value that estimates the confidence in each reagent, and thus enables the use of ATARiS in the selection of reagents for validation studies and as an aid in the design and refinement of RNAi libraries. Based on our work with these shRNAs, we anticipate that we lack more than 1 effective shRNA for a fraction of targeted genes. ATARiS will allow us to interrogate these situations and to develop improved libraries in the future. For example, one might create additional shRNA reagents for genes that lack solutions or create sub-libraries only containing shRNAs involved in ATARiS solutions. As RNAi libraries include more reagents per gene, and as screens include more samples, the ability of ATARiS to correctly identify on-target reagents will also improve.

ATARiS gene phenotype values are an inherently different metric than previous gene scores for RNAi. Existing methods determine gene candidates in a manner that is dependent on a user-defined desired phenotype whereas ATARiS aims to summarize the data available for each gene in an unbiased way. For example, with a "frequency approach," it is theoretically possible for a single gene in a sample to be a candidate for both a positive and a negative phenotype, whereas the ATARiS phenotype value provides a single metric for that gene. For methods such as RSA and SSMD, gene scores are influenced by the distribution of reagents towards or away from a *desired* phenotype. In comparison, ATARiS attempts to determine the best subset of reagents that describe the actual gene-level effect. Furthermore, it incorporates information across all screened samples instead of using data from each sample independently. Another major difference between ATARiS and previous work is that gene phenotype

values are *relative* to the samples screened instead of absolute, as is the case in RSA and SSMD. Finally, unlike RNAi analysis methods for two-class comparisons such as RIGER(Barbie et al., 2009) and "second best,"(Cheung et al., 2011a) which are primarily used to determine a single value representing each gene's differential effect across classes, ATARiS phenotype values describe the effect of each gene in each individual sample.

ATARiS is analogous to the approaches used by methods such as RMA (Irizarry et al., 2003) and dChip (Li and Hung Wong, 2001) for microarray data analysis in that gene scores are evaluated by incorporating multiple probes/reagents and excluding problematic ones. It is similar to dChip specifically in that a multiplicative model is fit to the set of probes/reagents. However, ATARiS differs from both methods in order to account for the unique attributes of RNAi data. For example, while most microarray probes are assumed to generally agree, the majority of RNAi reagents do not. Figure 2-10 shows how the correlation coefficients between data from shRNA pairs targeting the same gene are only marginally higher than the coefficients from random shRNA pairs. Thus, ATARiS implements an empirical null distribution to determine correlations that are significantly above background. Another difference is that ATARiS considers multiple solutions for each gene, as varying degrees of gene suppression by distinct reagents may produce different effect profiles across samples.

Although high-throughput shRNA viability screens are discussed here, ATARiS can be applied to any screen where multiple, redundant reagents produce different observed outcomes in multiple samples. For example, ATARiS would apply to a screen that uses siRNA reagents or measures a phenotype other than viability. A similar approach can potentially be applied to small molecule screening where multiple target-specific compounds are assayed across different samples. In addition, since ATARiS analyzes the data of each gene independently, it can be effectively applied to screens

that target a small number of genes as long as multiple samples are screened. One such example is validation screens, where screeners have prior expectation that the reagents screened will produce an effect. By contrast, methods such as RIGER (Barbie et al., 2009) construct a null distribution from all screened reagents, requiring many reagents to be screened while the majority of them may have no effect on the measured phenotype.

Despite the fact that parallel screens continue to grow in size, we recognize that not all datasets will be as large as Achilles. When the sample size is small, the number of genes with solutions decreases. This is due to the loss of genetic heterogeneity between samples and reduced statistical power to discriminate true effects from noise. The user-defined significance threshold allows for tailoring to specific applications. In our analyses, we chose a relatively lax threshold (0.15) such that greater numbers of gene solutions will be available for analysis. As our significance calculations are based on an empirical null distribution, a more stringent threshold will provide increased confidence in the solutions generated (i.e., lower false-positive rate). Furthermore, the compatibility of results from independent datasets suggests that investigators screening single or few samples may use ATARiS results (such as consistency scores) derived from larger datasets to improve their ability to assess reagent performance and gene effects in their screened samples.

We hope that by providing a foundation for interpreting RNAi gene suppression effects as quantifiable values in individual samples, we will aid functional genomics in reaching its full potential. We illustrated a number of analytic methods that are enabled by ATARiS, and note that investigators have already begun to map phenotype-based gene networks (Horn et al., 2011). Nevertheless, much remains to be explored. ATARiS results from the Achilles and Marcotte et al. (2012) datasets will be useful for deeper analysis, but we also believe that the application of ATARiS to other screening datasets, large and small, will yield novel insights.

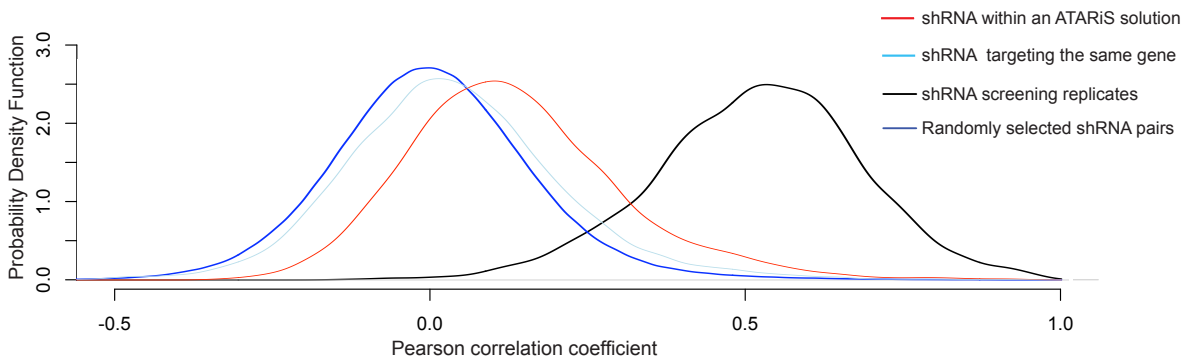

Figure 2-10. Similarity between effects produced by shRNAs across 102 screened samples. ShRNAs targeting 500 randomly selected genes were used to calculate Pearson correlation coefficients for screening data between all pairs of shRNA within each indicated set. Density distributions (Probability Density Function) of the correlation coefficients for each set are displayed in the indicated color. As expected, the correlations between shRNA profiles within ATARiS solutions are significantly higher than those between randomly selected pairs of shRNA profiles (p-value < 2.2×10^{-16}, Welch's t-test) and were also significantly higher than the correlations between profiles of shRNAs targeting the same gene (p-value < 2.2×10^{-16}, Welch's t-test), demonstrating that ATARiS identifies shRNAs with consistent effects.

CHAPTER THREE

Identification of HNF1B oncogenic function through integration of genomic data

This chapter has been adapted from:

Shao, D. D., Tsherniak, A., Gopal, S., Weir, B. A., Tamayo, P., Stransky, N., Schumacher, S. E., Zack, T.I., Beroukhim, R., Garraway, L.A., Margolin, A.A., Root, D.E., Hahn, W.C., Mesirov, J.P. (2013). ATARiS: Computational quantification of gene suppression phenotypes from multisample RNAi screens. *Genome Research*.

Contributions:

Barbara Weir, Shuba Gopal, Nicolas Stransky, Steven Schumacher, Travis Zack, Rameen Beroukhim, Levi Garraway, and David Root contributed datasets. Diane Shao performed computational analyses and experiments.

3.1 Introduction

The integration of analytic results from multiple genomic datasets has yielded many novel insights in cancer (See Chapter 1.4). The majority of such integration uses implied relationships, e.g. using recurrent amplifications identified in one dataset to filter functional results acquired from a different experimental model. The acquisition of multiple data types within the same samples provides direct insight to the relationships between DNA structural alteration, gene expression, and functional outcome.

RNAi data from the ATARiS analysis of the Achilles screen provides a source of quantitative functional data for individual cancer cell lines. The cell lines used in Achilles have also undergone gene expression profiling and aCGH through the efforts of the Cancer Cell Line Encyclopedia. *A priori*, we consider genes that are highly amplified, highly expressed, and highly dependent in cancer to be likely cancer drivers. The ability to assess these attributes in genomic data from matched samples provides a powerful opportunity.

3.2 Experimental methods

Cell culture

All cancer cell lines were cultured in RPMI-1640 (Invitrogen) supplemented with 10% FBS (Sigma) and 1% streptomycin and penicillin. HA1E and HA1EM immortalized lines were cultured in alpha-MEM (Invitrogen) supplemented with 10% FBS.

Lentiviral infection

Lentivirus for *HNF1B* dependency experiments were produced as previously described (Barbie et al., 2009). Cells were infected in media containing 8ug/ml polybrene and 1:10 dilution of virus. Infected cells were selected with 2ug/ml puromycin for 48 hours. Detailed shRNA identities are provided in Table 3-1.

Table 3-1. Identities of shRNA used in Chapter 3.

shRNA	**TRC Identifier**	**NM number**	**Target** (5'-3')
shHNF1B-1	TRCN0000017508	NM_000458.1-2162s1c1	CCGTACTGTCTATGTTGTGAT
shHNF1B-2	TRCN0000017509	NM_000458.1-734s1c1	CCGACAATTCAACCAGACAGT
shHNF1B-3	TRCN0000017510	NM_000458.1-751s1c1	GCAAATCTTGTACCAGGCCTA
shHNF1B-4	TRCN0000017511	NM_000458.1-800s1c1	CCGACAATTCAACCAGACAGT
shHNF1B-5	TRCN0000017512	NM_000458.1-923s1c	CAGTCCAGAGTTCTGGAAATA
shGFP	TRCN0000072181	clonetechGfp_437s1c1	ACAACAGCCACAACGTCTATA
shLacZ	TRCN0000072231	lacZ_1650s1c1	CGCTAAATACTGGCAGGCGTT

Low-throughput assessment of cell viability

Cells were replated at 50,000 cells/well post-infection and post-selection in triplicate in 12-well plates post-selection with puromycin. Wells were counted 4 days later by ViaCell.

Immunoblotting

Cell lysates collected 72 hours post-infection were run on 4-12% Bis-Tris gel (Invitrogen NuPAGE) and transferred to nitrocellulose membrane for immunoblotting. Primary antibodies were obtained from Santa Cruz (HNF1B sc-7411, ß-actin sc-1615)..

Anchorage Independent Growth Assay

HA1EM cells infected with lentiviral expression plasmid pLX-304 with desired genes were selected for 5 days in 10ug/ml blasticidin. Cells were seeded in triplicate at $2.5x10^4$ cells per well in 0.4% top agar (Difco) in 6-well plates. Bottom agar was 0.6% agar (Difco) supplemented with 20% FBS. Macroscopic images were collected of each well and colonies were counted using CellProfiler (http://cellprofiler.org).

Xenograft Assay

HT29 infected with lentiviral plasmid PLKO.1 sh*HNF1B*-1 or shControl were expanded for 4 days before subcutaneous implantation into immune-compromised mice (Taconic, CrTac:NCr-Foxn1nu). $2x10^6$ cells were implanted into each of three sites per mouse. Tumor growth was monitored every two weeks by digital caliper measurement of tumor diameter. Approximate cross-sectional area was calculated.

3.3 HNF1B is essential in HNF1B amplified tumors

3.3.1. Identification of HNF1B

We evaluated the correlation between each Achilles gene solution (described in Chapter 2) and the corresponding gene's expression values across 83 Project Achilles cell lines for which expression microarrays are available, and found that the gene with the highest correlation is *HNF1B* (Table 3-2). We noted that previously reported cancer dependencies or oncogenes such as *PAX8* (Cheung et al., 2011a; Li et al., 2011), *BCL2L1* (Beroukhim et al., 2010), *E2F3* (Oeggerli et al., 2006), and *MYB* (Ramsay, 2008) are also significantly essential in samples that express the gene highly. Furthermore, we evaluated whether oncogenes, a subset of genes that might be expected to be essential in highly expressed cell lines, are enriched in our results. We determined that the list of known amplified oncogenes reported in Beroukhim et al. (2010) was significantly over-represented towards the top of our list (*p*-value = 5.38e-06; Wilcoxon rank sum). For comparison, as expected, known tumor suppressors were not enriched (*p*-value = 0.84). These results suggested that other statistically significant genes from this analysis might contribute to malignant transformation, in particular *HNF1B*.

A common mechanism for increased gene expression in cancer is genomic amplification; thus we also analyzed which genes involved in recurrent, focal genomic amplifications specifically scored as dependent in those samples (See Chapter 2.3: Experimental methods). *HNF1B* again ranked at the top of this analysis (Table 3-3), suggesting that *HNF1B* was one target of this amplification. We note that *HNF1B* is amplified in 23 percent of all cancers (http://broadinstitute.org/tumorscape).

Table 3-2. Table of genes that are essential in samples with high expression. Top 75 results are shown. Known cancer drivers are highlighted in red.

Rank	ATARiS Solution	Gene	Correlation	P-Value	FDR
1	HNF1B_1_11000	HNF1B	-0.553	2.00E-05	0.075
1	PAX8_1_10011	PAX8	-0.534	2.00E-05	0.075
3	E2F3_1_11111	E2F3	-0.427	4.00E-05	0.075
3	ELF3_1_01001	ELF3	-0.434	4.00E-05	0.075
5	SOX10_1_01111	SOX10	-0.436	6.00E-05	0.075
5	HIST1H4D_1_0101	HIST1H4D	-0.422	6.00E-05	0.075
7	NGEF_1_01101	NGEF	-0.433	8.00E-05	0.086
8	FERMT1_1_01010	FERMT1	-0.398	1.00E-04	0.094
9	BCL2L1_1_11100	BCL2L1	-0.373	1.40E-04	0.096
9	ASL_1_11111	ASL	-0.399	1.40E-04	0.096
11	POLE3_1_11010	POLE3	-0.379	1.40E-04	0.096
12	MYB_1_1111111	MYB	-0.370	3.40E-04	0.213
13	MPP6_1_0110	MPP6	-0.357	4.00E-04	0.225
14	PITX3_1_10111	PITX3	-0.370	4.20E-04	0.225
15	HNF4A_1_10101	HNF4A	-0.360	4.60E-04	0.23
16	DNAJB8_1_011	DNAJB8	-0.362	5.00E-04	0.23
17	PTBP2_1_01001	PTBP2	-0.350	5.20E-04	0.23
18	SOX9_1_11011	SOX9	-0.346	5.80E-04	0.242
19	ZNF573_1_1011	ZNF573	-0.335	6.40E-04	0.253
20	ACTN1_1_0111	ACTN1	-0.343	9.20E-04	0.345
21	ZNF695_1_11111	ZNF695	-0.337	1.06E-03	0.369
22	TNFSF10_1_11011	TNFSF10	-0.340	1.08E-03	0.369
23	PDE3A_1_11111	PDE3A	-0.326	1.14E-03	0.372
24	FUBP1_2_11001	FUBP1	-0.342	1.20E-03	0.372
25	PNLDC1_1_11000	PNLDC1	-0.348	1.24E-03	0.372
26	ODZ1_1_11110	ODZ1	-0.328	1.34E-03	0.384
27	CHI3L2_1_11000	CHI3L2	-0.307	1.38E-03	0.384
28	NRG2_1_01010	NRG2	-0.311	1.44E-03	0.386
29	POMGNT1_1_01111	POMGNT1	-0.323	1.56E-03	0.393
30	ADNP2_1_00011	ADNP2	-0.317	1.62E-03	0.393
31	TRADD_1_0111	TRADD	-0.314	1.64E-03	0.393
32	HDAC4_1_11100	HDAC4	-0.326	1.72E-03	0.393
33	RBM47_1_10010	RBM47	-0.316	1.74E-03	0.393
34	MAPT_1_111	MAPT	-0.333	1.78E-03	0.393
35	HOXA9_1_0111	HOXA9	-0.312	1.90E-03	0.401
36	KIAA0430_1_10100	KIAA0430	-0.313	1.92E-03	0.401
37	AGPAT3_1_0110	AGPAT3	-0.307	2.12E-03	0.417
38	E4F1_1_11110	E4F1	-0.308	2.18E-03	0.417
39	KRAS_1_001111101011	KRAS	-0.336	2.26E-03	0.417
40	CCNE1_1_0111	CCNE1	-0.310	2.30E-03	0.417
41	KPNA5_1_11111	KPNA5	-0.308	2.48E-03	0.417

Table 3-2 (Continued).

42	TEAD1_1_10110	TEAD1	-0.293	2.50E-03	0.417
43	FLNB_1_00111	FLNB	-0.299	2.58E-03	0.417
43	FGFR1OP_1_1101	FGFR1OP	-0.298	2.58E-03	0.417
45	WWTR1_1_11111	WWTR1	-0.307	2.62E-03	0.417
46	ADAM21_1_11110	ADAM21	-0.306	2.64E-03	0.417
47	GPR22_1_10110	GPR22	-0.305	2.72E-03	0.417
47	PLXDC2_1_11111	PLXDC2	-0.305	2.72E-03	0.417
47	LMNB2_1_0111	LMNB2	-0.300	2.72E-03	0.417
50	CTNNB1_1_0110	CTNNB1	-0.313	3.04E-03	0.451
51	FOXD2_1_0101	FOXD2	-0.307	3.08E-03	0.451
52	RALGPS2_1_10110	RALGPS2	-0.300	3.14E-03	0.451
53	STK31_1_10001	STK31	-0.310	3.18E-03	0.451
54	CHML_1_1110	CHML	-0.292	3.40E-03	0.463
55	SLC29A3_1_01110	SLC29A3	-0.303	3.50E-03	0.463
56	GYS2_1_11110	GYS2	-0.293	3.52E-03	0.463
57	GBE1_1_0111	GBE1	-0.289	3.66E-03	0.463
58	ITGAV_1_110	ITGAV	-0.284	3.78E-03	0.463
59	CHST2_1_11111	CHST2	-0.285	3.84E-03	0.463
60	ELOVL4_1_01001	ELOVL4	-0.302	3.88E-03	0.463
61	CMKLR1_2_11000	CMKLR1	-0.287	3.94E-03	0.463
62	SAMD4B_1_11111	SAMD4B	-0.295	3.96E-03	0.463
63	HMOX2_1_1011	HMOX2	-0.291	3.98E-03	0.463
64	MICB_1_01111	MICB	-0.293	4.04E-03	0.463
65	CCNB1_1_10111	CCNB1	-0.286	4.16E-03	0.463
66	KCNH4_1_10111	KCNH4	-0.292	4.18E-03	0.463
66	HS3ST5_1_10110	HS3ST5	-0.297	4.18E-03	0.463
68	LGALS13_1_1111	LGALS13	-0.295	4.32E-03	0.463
69	PTGFR_1_11001	PTGFR	-0.285	4.38E-03	0.463
70	NDOR1_1_01001	NDOR1	-0.289	4.42E-03	0.463
71	WT1_1_10001111000	WT1	-0.296	4.44E-03	0.463
71	TMUB1_1_11010	TMUB1	-0.286	4.44E-03	0.463
73	IRX1_1_11111	IRX1	-0.268	4.52E-03	0.465
74	ADARB1_1_01111	ADARB1	-0.282	4.68E-03	0.475
75	KLF16_1_11111	KLF16	-0.277	5.00E-03	0.497

Table 3-3. Table of genes that are essential in samples with focal gene amplification. Top 25 results are shown. Known cancer drivers are highlighted in red.

Rank	ATARiS Solution	Gene	Mean Difference	P-Value	FDR
1	HNF1B_1_11000	HNF1B	-2.81	1.20E-04	0.300
2	OR2T2_1_1001	OR2T2	-1.42	9.00E-04	0.700
3	E2F3_1_11111	E2F3	-1.14	1.22E-03	0.700
4	SRI_1_01011	SRI	-0.94	1.40E-03	0.700
5	GSK3B_1_001111011	GSK3B	-1.34	1.66E-03	0.700
6	ZAP70_1_11101	ZAP70	-0.83	1.80E-03	0.700
7	HOXC13_1_11111	HOXC13	-0.83	2.22E-03	0.700
8	PAX8_1_10011	PAX8	-1.07	2.42E-03	0.700
9	SLC35B3_1_01111	SLC35B3	-1.08	2.52E-03	0.700
10	GH1_1_11011	GH1	-0.89	3.44E-03	0.733
11	CACNG7_1_1111	CACNG7	-1.02	3.68E-03	0.733
12	JUN_1_111110	JUN	-0.86	3.88E-03	0.733
13	SELL_1_01111	SELL	-1.61	4.12E-03	0.733
14	AK5_1_01101	AK5	-1.14	4.18E-03	0.733
15	RPS6KC1_1_11101	RPS6KC1	-0.83	4.40E-03	0.733
16	TFAP2B_1_1111	TFAP2B	-0.97	5.32E-03	0.742
17	GLI1_1_10101	GLI1	-1.09	5.44E-03	0.742
18	TNNI3K_1_11101	TNNI3K	-1.20	6.04E-03	0.742
19	HECTD1_1_01001	HECTD1	-0.93	6.12E-03	0.742
20	RALGPS2_1_10110	RALGPS2	-1.54	7.08E-03	0.742
21	NFE2_1_01101	NFE2	-1.20	7.30E-03	0.742
22	GMFG_1_1110	GMFG	-1.37	7.38E-03	0.742
23	PRTFDC1_1_11110	PRTFDC1	-1.20	7.54E-03	0.742
24	NR1I2_1_11111	NR1I2	-1.04	7.58E-03	0.742
25	SP7_1_11110	SP7	-0.54	7.92E-03	0.742

3.3.2 Characterization of HNF1B dependency

Characterization of *HNF1B* dependency was made straightforward by directly applying ATARiS results to reveal which shRNAs are on-target and, specifically, which samples show higher dependency. We confirmed that the two *HNF1B*-specific shRNAs receiving the highest consistency scores suppressed HNF1B levels as assessed by immunoblotting (Fig. 3-1a). Furthermore, exogenous expression of HNF1B in cells expressing a *HNF1B* 3'UTR-specific shRNA (sh*HNF1B*-1) abrogated the cell death induced by expressing the *HNF1B* 3'UTR-specific shRNA alone (Fig. 3-1b), confirming that the observed shRNA effects were specific. We used a panel of cell lines to confirm that HNF1B protein expression was correlated to *HNF1B* phenotype values (Fig. 3-2a). Finally, we used cell lines that expressed high levels of *HNF1B* to confirm that they were indeed sensitive to *HNF1B* suppression by the two *HNF1B*-specific shRNAs as compared to control shRNA. For comparison, we showed that DLD-1 and an immortalized cell line HA1E(Hahn et al., 1999), neither of which harbor amplifications involving *HNF1B* nor express high levels of the gene, are insensitive to *HNF1B* suppression (Fig. 3-2b).

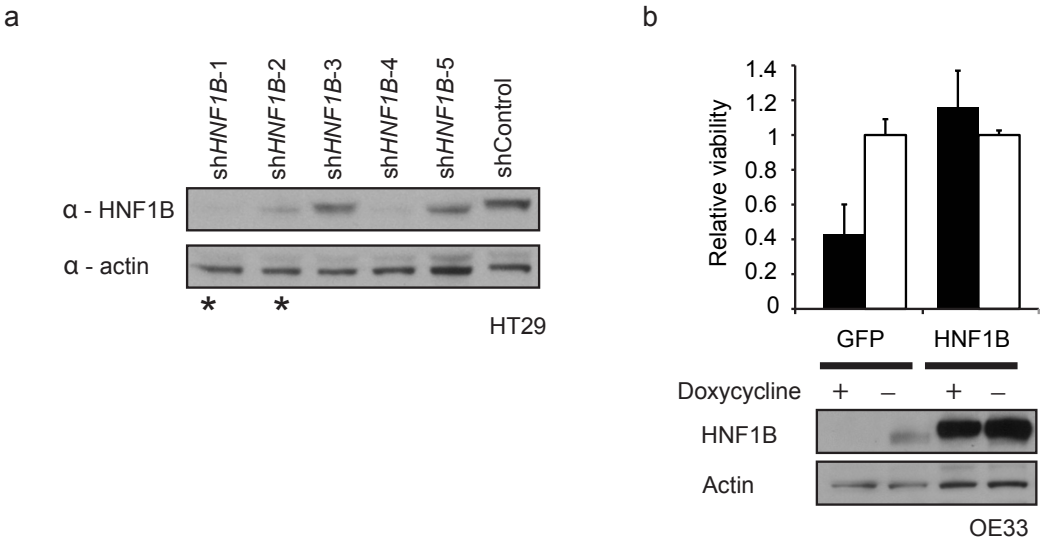

Figure 3-1. Validation of HNF1B shRNA. (**a**) Immunoblot of HNF1B after expression of five independent shRNA designed to target HNF1B. The two shRNA incorporated into the ATARiS solution, which also have the highest consistency scores, are indicated by asterisks. (**b**) Cell viability upon exogenous expression of HNF1B or GFP in an HNF1B-sensitive cell line OE33 with stable integration of doxycycline-inducible expression of shHNF1B-1. Each bar in the graph corresponds to the immunoblot lane directly below.

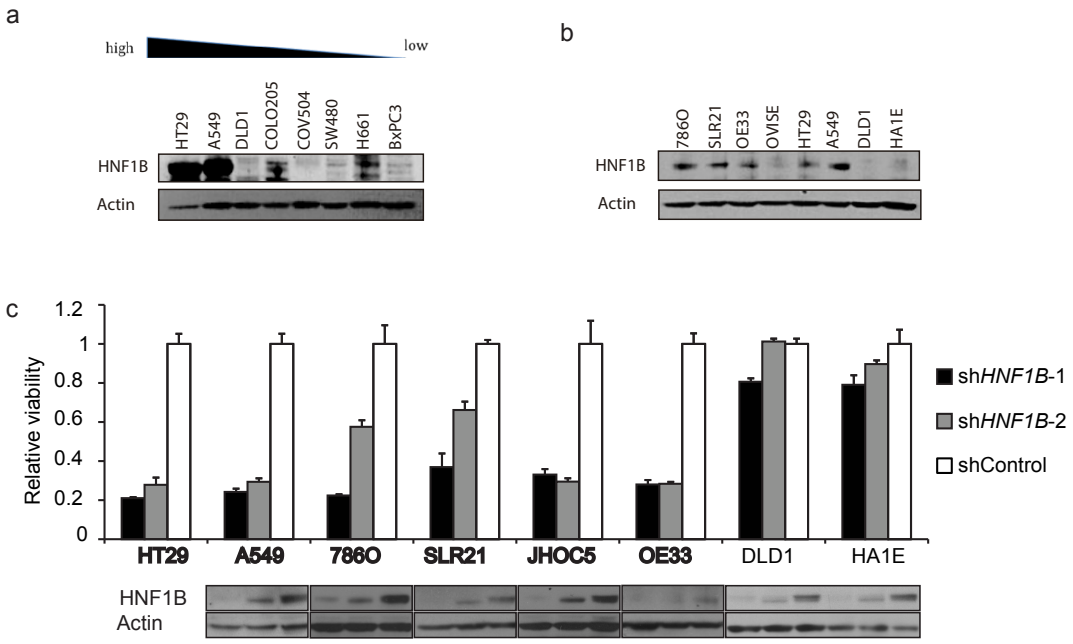

Figure 3-2. High expression of HNF1B corresponds to HNF1B dependency in cancer cell lines. (**a**) Relationship between HNF1B gene phenotype value and expression in a panel of cell lines ordered from high to low dependence. (**b**) Specific cell lines used in panel (c). (**c**) Relative viability of a panel of cell lines upon suppression of control or two *HNF1B*-specific shRNAs. Cell lines with high levels of HNF1B are shown in bold text. Each bar in the graph corresponds to the immunoblot lane directly below. Each boxed image derives from a separately exposed gel, as the *HNF1B*-amplified samples express much higher endogenous levels of HNF1B (Supplementary Fig. 10). Data for HT29 is shown in panel Fig. 3-1a.

ATARiS phenotype values allowed us to identify two additional cell lines SLR-21 and 786-O that had low phenotype values for HNF1B, but for which we did not have corresponding copy-number data. We confirmed that these cell lines were also dependent on *HNF1B* (Fig. 3-2b) and had corresponding genomic copy-number gain (Fig. 3-3).

To determine whether HNF1B expression is essential for tumor maintenance in vivo, we performed xenograft experiments by implanting HT29 colon cancer cells subcutaneously after expression of control or *HNF1B*-specific shRNA (Fig. 3-4a). In the initial two weeks, the xenografts with suppressed HNF1B showed marked growth impairment (n=3; p<0.01, one-tailed Student's t-test). Four weeks post-injection, their growth increased, likely due to re-activation of HNF1B expression (Fig. 3-4b), suggesting that HNF1B expression was critical for growth.

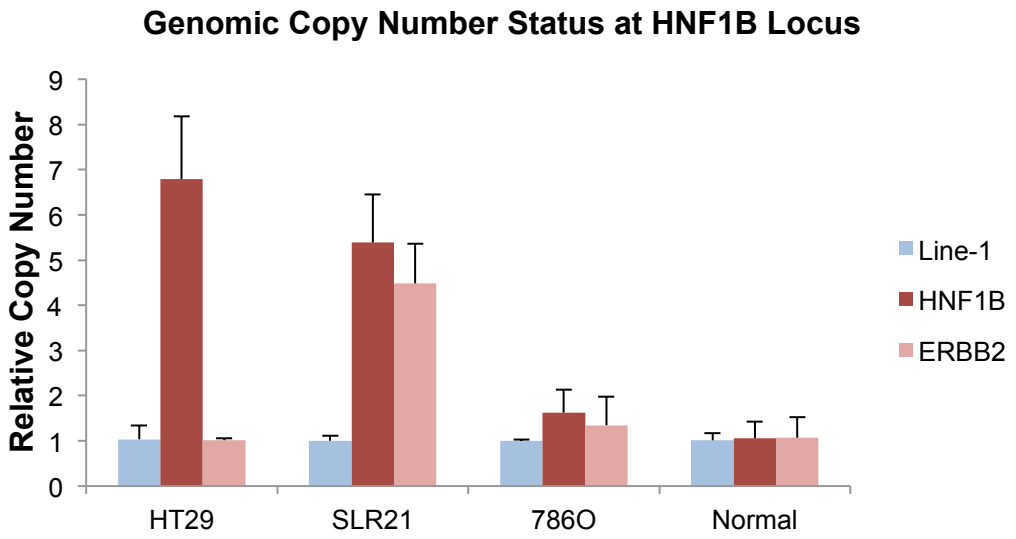

Figure 3-3. Genomic copy number status at the HNF1B locus. Quantitive RT-PCR of genomic DNA at HNF1B locus and adjacent ERBB2 locus. HT29 harbored known amplification of HNF1B. SLR21 and 786O had unknown copy number status but were identified as dependent by ATARiS phenotype values and subsequent validation. Error bars, +/- one standard deviation (*n*=3). Primers complementary to Line-1 genomic repetitive elements were used for normalization, and normal human DNA (Applied Biosystems) was used as reference.

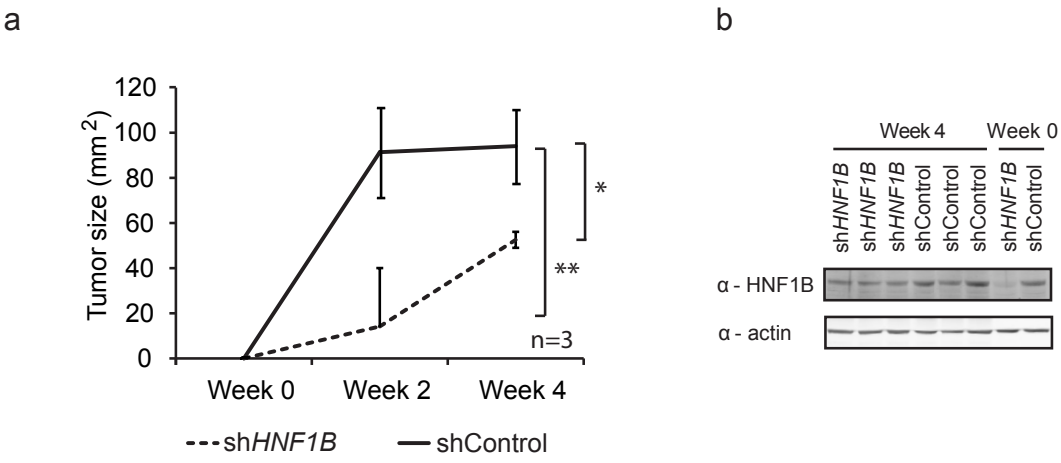

Figure 3-4. HNF1B sensitive cell line HT29 expressing sh*HNF1B*-1 or shControl was implanted subcutaneously into immune-compromised mice. ShHNF1B-1 was used for all experiments since it has potent effects and is specific for *HNF1B*, as shown in panels **a** and **b**. Tumor volume was monitored bi-weekly and lysates were collected pre-implantation and from tumors formed at 4 weeks. * indicates *p*-values < 0.05, ** indicates *p*-value < 0.01 (one-tailed Student's *t*-test)

3.3.3 HNF1B drives anchorage independent growth

Finally, we sought to determine whether expression of HNF1B transforms human cell lines. Specifically, we introduced *HNF1B* or *LacZ* cDNA into HA1EM cells, which are immortalized, non-tumorigenic human embryonic kidney cells that are transformed upon addition of oncogenes AKT or IKKε (Boehm et al., 2007). Expression of HNF1B conferred the ability for anchorage independent growth, a marker of cell transformation (Fig. 3-5). Together these observations – that *HNF1B* is amplified in human cancers, transforms immortalized cells, and is essential for those cancer cell lines that harbor increased *HNF1B* copy number – provide strong evidence that *HNF1B* is an oncogene.

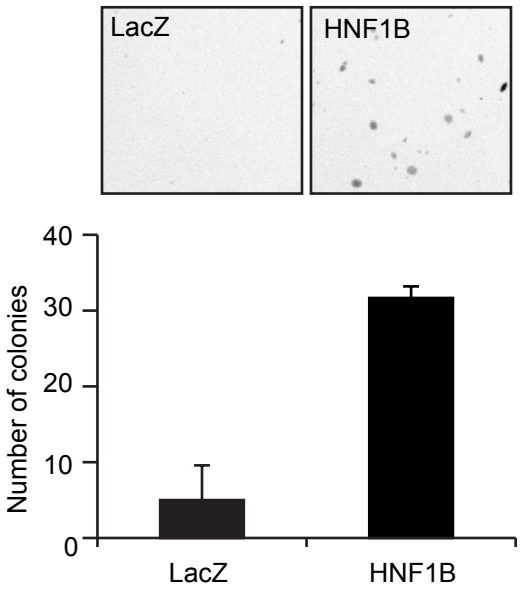

Figure 3-5. HNF1B expression increases anchorage independent growth. HNF1B or LacZ was expressed in HA1EM cells and anchorage independent growth determined. Representative photos shown after six weeks. Error bars, ± one standard deviation (*n*=3).

3.4 Discussion

The role of HNF1B in cancer has been controversial, with conflicting evidence supporting both oncogenic and tumor suppressor roles. In ovarian clear cell carcinoma cell lines, *HNF1B* is highly expressed and HNF1B suppression decreases cell viability (Tsuchiya et al., 2003). In other ovarian cancer studies, epigenetic inactivation of *HNF1B* has been reported (Terasawa et al., 2006). Studies of cancer risk SNPs across different ovarian cancer cell types reveal differential risk conferred in different subtypes, leading to increased HNF1B expression in ovarian clear cell, and increased methylation of *HNF1B* in serous subtypes (Shen et al., 2013). Genome-wide association studies have associated SNPs in the *HNF1B* locus with risk for prostate and endometrial cancers (Spurdle et al., 2011; Schumacher et al., 2011).

Our results support a functional role for HNF1B expression across multiple cancer cell types. While ovarian samples were included in our analysis, the samples that had increased HNF1B expression also encompassed other lineages such as colon and esophageal. Prostate and endometrial cancer cell lines were not well represented in our dataset. Developmentally, *HNF1B* is required for visceral endoderm formation (Barbacci et al., 1999) and proper development of the genitourinary tract (Bellanne-Chantelot et al., 2005; Ryffel, 2001), but appears to be dispensable in adult tissue (Verdeguer et al., 2009), making it a reasonable candidate for therapeutic targeting.

HNF1B is located near the known oncogene *ERBB2*. However, our observations indicate that HNF1B independently induces anchorage independent growth. It remains possible that HNF1B may cooperate with *ERBB2* to drive transformation in a manner analogous to what has been observed for *YAP1* and *CIAP1*, which reside in a single amplicon in hepatocellular cancer (Zender et al., 2006).

While the mechanism of HNF1B oncogenic activity has not been examined here, we propose several mechanisms. One putative mechanism involves CDK8, a lineage-specific oncogene in colon cancer (Firestein et al., 2008). We find that gene expression changes upon HNF1B suppression strongly match expression changes upon CDK8 suppression in the same cell lines. Co-immunoprecipitation of CDK8 and HNF1B in endogenous samples has been observed (data not shown). A second putative mechanism involves the tumor suppressor HNF1A (Bluteau et al., 2002). HNF1B and HNF1A commonly form heterodimers. Thus HNF1B overexpression may titrate away available HNF1A, thus preventing tumor suppressor function of HNF1A. These hypotheses for the mechanism of HNF1B merit further exploration.

CHAPTER FOUR

Functional expression screen reveals YAP1 convergence with KRAS signaling to regulate EMT

This chapter has been adapted from:

Shao, D. D., Xue, W., Krall, E.B., Bhutkar, A., Piccioni, F., Wang, X., Schinzel, A.C., Sood, S., Rosenbluh, J., Kim, J.W., Zwang, Y., Root, D.E., Jacks, T., Hahn, W.C. KRAS and YAP1 converge to regulate EMT and tumor survival. *In preparation*.

Contributions:

Xiaoxing Wang, Joseph Rosenbluh, Joon Kim, Yaara Zwang, David Root, Tyler Jacks, and William Hahn contributed to experimental design. Federica Piccioni and Anna Schinzel were involved in design and execution of screening. Elsa Krall contributed EMT experiments (4-21a, 4-22). Wen Xue and Sabina Sood performed mouse experiments (Fig. 4-10, 4-11, 4-12). Arjun Bhutkar analyzed RNA-seq data from mouse experiments (4-10d). Diane Shao performed screening, *in vitro* experiments, and data analysis.

4.1 Introduction

4.1.1 Mitogenic signaling pathways

As reviewed in the Introduction (Chapter 1.3), activation of proto-oncogenes induces a state in which cancers are dependent on continuous signaling from the oncogene for survival. This state, sometimes referred to as oncogene addiction (Weinstein and Joe, 2008), has been observed in cancers harboring mutations, amplifications, or translocations of *BRAF*, *EGFR*, *HER2* and *BCR-ABL* (Sharma and Settleman, 2007). Although the mechanisms that lead to oncogene addiction remain poorly understood, pharmacologic inhibition of these oncogenes results in significant clinical responses. Tumors that become resistant to these therapeutic interventions often exhibit re-activation of the signaling pathways regulated by these oncogenes. For example, EGFR addicted cancers that relapse have been found to harbor *MET* copy-number amplification or *KRAS* mutation, which mediate resistance by activating downstream effector pathways independent of EGFR (Jänne et al., 2009).

Activating mutations of the *KRAS* proto-oncogene occur in a substantial fraction of human pancreatic, lung and colon cancers, and those cancers that express the mutant allele exhibit oncogenic addiction to KRAS (Sharma and Settleman, 2007; Lau and Haigis, 2009). Oncogenic KRAS activates pleiotropic signaling pathways that contribute to tumor initiation and maintenance including the Mitogen-activated protein kinase (MAPK), Phosphoinositide 3-kinase (PI3K), and Ral guanine nucleotide exchange factor (RalGEF) signaling pathways (Pylayeva-Gupta et al., 2011). Suppression or inhibition of each of these pathways prevents tumor initiation and slows the growth of established tumors (Kolch et al., 1991; Ehrenreiter et al., 2009; González-García et al., 2005; Gupta et al., 2007). One consequence of mutant KRAS signaling is aberrant activation of the AP-1 family transcription factors, important for promoting

cellular responses to mitogenic signaling (Karin, 1995). Specifically, KRAS has been shown to increase FOS and JUN activation through MAPK-dependent and -independent mechanisms (Deng and Karin, 1994). Despite extensive studies of KRAS signaling pathways, our understanding of the components specifically important for KRAS oncogenic addiction remains incomplete.

4.1.2 YAP1 signaling in development and cancer

YAP1 is a transcriptional co-activator that participates in several context-dependent transcriptional programs that regulate organ size and promote cell proliferation (Wang et al., 2009). Recurrent *YAP1* amplifications are observed in hepatocellular cancers, and YAP1 is essential for tumors that harbor amplifications involving *YAP1* (Zender et al., 2006). In addition to regulating proliferation, YAP1 has also been reported to drive an epithelial-to-mesenchymal transition (EMT) and to increase the metastatic potential of mammary epithelial cells (Lamar et al., 2012; Zhao et al., 2008b; Overholtzer et al., 2006).

Regulation by YAP1 includes Hippo (Hpo)-dependent and Hpo-independent factors. The mammalian Hippo pathway involves a core kinase cascade in which MST1/2 phosphorylates LATS and MOB, which in turn inhibit YAP1 entry into the nucleus by phosphorylation of YAP1 on serine-127 (Pan, 2010). Phosphorylated YAP1 is sequestered in the cytoplasm, and thus cannot activate downstream transcriptional programs. Cell membrane inputs to YAP1 signaling include cell density sensing likely through response to mechanical stress (Dupont et al., 2011), and G-protein coupled receptor (GPCR) signaling through the Gs component (Yu et al., 2012). One Hpo-independent signaling mechanism implicated in cancer development involves the

phosphorylation of YAP1 at tyrosine-357 by the tyrosine kinase YES1 to promote YAP1 interaction with β-catenin and modulation of Wnt signaling (Rosenbluh et al., 2012).

In the nucleus, YAP1 activates Transcriptional Enhancer Activator Domain (TEAD) family transcription factors, particularly during development (Pan, 2010). In mammalian cells, additional YAP1 transcriptional partners such as SMAD, RUNX, TBX5, and the ERBB4 internal cytoplasmic fragment have been described (Wang et al., 2009). These observations suggest that YAP1 interacts with specific transcription factors in different cellular contexts to promote cell proliferation, organ growth, or survival.

4.1.3 Rationale

Because genes that promote clinical resistance to targeted therapies provide direct mechanistic insight regarding the targeted oncogene, we applied a similar concept to systematically probe KRAS oncogenic addiction pathways. We performed a genetic screen to identify open reading frames (ORFs) that are able to sustain the survival of KRAS-dependent cancer cell lines in the setting of KRAS suppression. Here we report that YAP1 is required for RAS induced cell transformation and that YAP1 and KRAS converge at the transcriptional level to regulate genes involved in the epithelial-mesenchymal transition.

4.2 Experimental methods

KRAS rescue screen

HCTtetK cells were seeded at 300 cells per well in 50 ul of media in 384-well plates. The next day, cells were infected using 1 ul of virus ($1.4x10^8$ infectious particles per ml per RNAi consortium virus tittering protocol; http://www.broadinstitute.org/rnai/public/resources/protocols) in 25 ul media supplemented with 8ug/mL Polybrene. The media was exchanged the following day. Two days after infection, 500 ng/mL doxycycline was added to a total volume of 50 ul media per well. For 10% of plates, additional replicates received treatment with blasticidin or no treatment for confirmation of overall infection efficiency. Seven days after infection, viability of each well was determined by CellTiterGlo (Promega). B-score adjustment was performed for each plate (Brideau et al., 2003), and final score for each ORF was normalized to approximately 40 negative control values on each plate (uninfected wells and wells infected with HcRed, eGFP, BFP, LacZ, or Luciferase).

Generation of HCTtetK cells

HCT116 cells were infected with lentivirus to integrate a doxycycline inducible KRAS shRNA using the pLKO-Tet-On backbone (Novartis). Selection for cells harboring stable integration was achieved using 2 ug/mL of puromycin. Cells were seeded at 0.3 cells per well in 96-well plates to allow selection of clones. We assessed thirty clones and selected HCTtetK based on effectiveness of KRAS suppression upon doxycycline treatment.

In-Cell Western

Cells were seeded at 1000 cells per well in black with clear-bottom 384-well plates in 30 ul media. The following day, 30 ul media supplemented with 16ug/mL Polybrene was added to each well for infection with 1.5 ul virus containing each ORF. One day post-infection, 45 ul media was exchanged for 25 ul fresh media. Two days post-infection, an additional 30 ul media supplemented with 500 ug/mL doxycycline was added. After 30 hrs, cells were fixed using 25 ul 4% formaldehyde and 0.1% TX-100 in PBS for 30 min at RT. Fixative was washed off using 50 ul PBS. Blocking was performed using 25 ul Odyssey Blocking Buffer (LI-COR) overnight at 4°C. 18 ul of primary antibody mix (1:500 dilution of anti-phospho-S6 (Cell Signaling #2211) + 1:1000 dilution of anti-S6 (Cell Signaling #2317); or 1:500 dilution of anti-phospho-ERK (Cell Signaling #9101) + 1:1000 dilution of anti-ERK (Santa Cruz sc-135900) in 0.1% Tween-20 in Odyssey Blocking Buffer) was incubated for 1 hr at RT. Washes were performed 3 times, 10 minutes each, using 0.1% Tween-20 in PBS. 18 ul of secondary antibody mix (1:800 IRDye 680RD goat anti-mouse (LI-COR) +1:800 IRDye 800CW goat anti-rabbit (LI-COR) in 0.1% Tween-20 in Odyssey Blocking Buffer) was incubated for 1 hr at RT. Washes were again performed. A final rinse was performed using 50 ul PBS. Scanning was performed using LI-COR Odyssey (680RD at intensity 7.0, 800CW at intensity 9.0) and quantification was performed using LICOR Image Studio software.

Rescue experiments in vitro

ORFs were introduced by lentiviral delivery and cells expressing each ORF were selected with 10ug/mL Blasticidin for at least 4 days. For cell lines harboring a tet-inducible shRNA, cells expressing each ORF were then seeded in replicates in 96-well plates and treated with media supplemented with doxycycline or media alone over 5 days. Results were quantified by CellTiterGlo (Promega). For cell lines that do not

harbor a tet-inducible shRNA, the desired shRNA was introduced by lentiviral delivery and selected with Puromycin for 48 hours before replating at 10,000 cells per well in replicates in 12-well plates. Results were quantified by Vi-Cell Cell Viability Analyzer (Beckman Coulter).

Mouse lung transplant model

$5x10^4$ cells were transplanted into NCr-nu/nu recipient mice (Taconic) by tail vein injection and mice were treated with Doxycycline diet *(Harlan* Laboratories). Bioluminescence imaging was performed as previously described (Xue et al., 2011). Luciferase signal in the lung was quantified using Xenogen software and normalized to tumors on Day 0, before doxycycline treatment.

Generation of KP-KrasA cells and derivative lines

$Kras^{G12D};p53^{fl/fl}$ lung adenocarcinoma cells were infected with retroviral vectors TRE-GFP-miR30 shKras-PGK-Puro (Zuber et al., 2011), rtTA3-PGK-Hygro and MSCV-luciferase-IRES-GFP. GFP^+ cells were sorted into single cell clones to screen for cells showing efficient doxycycline-inducible Kras knockdown, resulting in KP-KrasA, KP-KrasB, and KP-KrasC lines from independent clones. KP-KrasA were stably infected with TRE-dsRed-miR30 shYap1-PGK-Venus-IRES-NeoR (Zuber et al., 2011) to simultaneously express shKras and shYap1 from both TRE promoters upon doxycycline treatment.

Anchorage independent growth assays

Cells were seeded in triplicate at $2x10^4$ cells per well in 0.4% top agar (Difco) in 6-well plates. Bottom agar was 0.6% agar (Difco) supplemented with 20% FBS. Macroscopic images were collected of each well and colonies were counted using CellProfiler (http://cellprofiler.org).

Chromatin Immunoprecipitation

Cultured cells were fixed for 15 min at room temperature with 1% formaldehyde in PBS. Crosslinking was stopped using 2.5 M glycine. Cells were collected in RIPA buffer (Sigma) containing protease inhibitors (Roche). Sonication was performed for 23minutes at 40% intensity and samples were centrifuged to remove insoluble materials. 1 mg lysate was incubated with 10 ug of either YAP1 (Santa Cruz) or FOS (Santa Cruz) antibody overnight at 4°C. The next day 50 µl Protein G Sepharose (Sigma) was added for 2 hr at 4°C. The beads were then washed twice with cold RIPA followed by 5 washes with wash buffer (100 mM Tris-HCl pH = 8.5, 500 mM LiCl, 1% NP-40 (v/v), 1% deoxycholic acid (v/v), and another two washes with RIPA. Beads were resuspended in 100 µl of TE buffer, and DNA was reverse cross-linked by adding 200 µl of Talianidis buffer [70 mM Tris-HCl pH = 8, 1 mM EDTA, 1.5% SDS (w/v)] and incubating overnight at 65°C. Supernatant containing DNA was collected. DNA yield was quantified by Qubit fluorometer (Invitrogen). Equal amount of DNA (approx. 0.1ng) of each sample was used for qPCR. Primers were designed using UCSC Genome Browser (http://genome.ucsc.edu) for reported FOS binding sites upstream of VIM and SNAI2, and a genomic locus downstream of each gene was used as control. Detailed primer sequences are provided in Table 4-1.

Protein Co-immunoprecipitation Assay

Nuclear Complex Co-IP kit (Thermo Fisher) was used for co-immunoprecipitation. 200ug lysate was used per IP reaction. Immunoprecipitation was performed using 10ul YAP1 antibody (Cell Signaling #4912), 50ul V5-tagged magnetic beads (Fisher Scientific #5050600), or the corresponding amount of control antibodies (anti-GFP, Cell Signaling #2555; Dynabeads protein G, Life Technologies #10004D). Anti-V5 antibody directly conjugated to HRP (Life Technologies #R96125) was used in cases where cross-reactivity to immunoglobulin heavy chains needed to be avoided.

Collection of RNA

RNA from cell lines were extracted using PerfectPure RNA Cell Kit (Fisher Scientific #2302820). RNA from mouse tumor tissue was extracted using TRIzol (Invitrogen) after tissue homogenization.

Reporter Assay

293T were transfected using TransIT-LT1 reagent (Mirus). 1 ug of each ORF was transfected together with 0.5 ug reporter (or control reporter) and 0.5 ug pLX-GFP construct. 24 hours post-transfection, cells were seeded in 30 ul media in 384-well plates. Reporter activity was assessed using Steady-Glo Luciferase Assay System (Promega), and normalized to total GFP fluorescence. TEAD reporter (8xGT-IIC-LucII) and control (delta-51-LucII) were obtained from RIKEN Bioresource Bank (http://dna.brc.riken.jp/), and AP1 reporter (pAP1(PMA)-TA-Luc) and control (pTA-Luc) were purchased from Clontech.

Microarray Processing

HCTtetK cells expressing LacZ, KRAS, or YAP1 were treated with doxycycline for 30 hours to suppress KRAS. Untreated cells were used as control. RNA was collected using PerfectPure RNA Cultured Cell Kit (5Prime) and expression profiling was performed on Human Genome U133A 2.0 Array (Affymetrix) using the Dana Farber Cancer Institute Microarray Core. Signal intensities were estimated by Affymetrix® "Statistical Algorithm" for absolute analysis. Relative fold change in intensity between treated and untreated samples were calculated.

Gene Expression Analysis

To identify enriched gene sets, we used genes co-regulated by both KRAS and YAP1 to query the MSigDB gene set database C2 collection of chemical and genetic perturbations (CGP) version accessed July 2011 (http://www.broadinstitute.org/gsea/msigdb). Statistical enrichment was measured by hypergeometric test per MSigDB website. To identify enriched transcription factors, we input the same gene list to query into the TransFind algorithm (Kielbasa et al., 2010). As a list of "unregulated" gene sets, we input the list of genes regulated by KRAS alone but not by YAP1. Promoter regions of 1000 nucleotides were considered for each gene using TRANSFAC highest information database version 2009.4.

RNA-Seq Processing

Paired-end Illumina mRNA sequencing for nine samples yielded 51mer reads in the range of 38.5M to 88.6M pairs per sample. Reads were trimmed to remove traces of adapter sequence using the FASTX-Toolkit from the Hannon Laboratory (CSHL, http://hannonlab.cshl.edu/fastx_toolkit). Adapter-stripped reads were aligned with Tophat (ver. 2.0.5) (Trapnell et al., 2009) using mouse (mm9) transcriptome and genome

annotation from the UCSC genome browser (Kent et al., 2002). Approximately 86%-94% of reads were aligned across samples. Transcript assembly, abundance estimation, and differential expression analysis were performed using Cufflinks (ver. 2.0.2)(Trapnell et al., 2010). Three replicates for each time-point (day 0, day 21), corresponding to each of three clones (KP-KrasA, KP-KrasB, and KP-KrasC), were grouped to derive significance of differential expression across experimental conditions.

Immunoblotting

Immunoblotting was performed using 40 ug cell lysate per sample on 4-12% Bis-Tris gels (Invitrogen NuPAGE) and transferred to nitrocellulose membrane. Primary antibodies were obtained from Santa Cruz (KRAS sc-30, ß-actin sc-1615, FOS H-125, MYC sc-764, ERK1/2 sc-135900) and Cell Signaling (PI3Kinase 110alpha #4255, p44/42 MAPK #9107, Phospho-S6 Ribosomal Protein #2211, S6 Ribosomal Protein #2317). Immunoblots were visualized by infrared imaging (LI-COR) with the exception of KRAS and MYC, visualized by chemiluminescence.

Cell Culture

All cancer cell lines were cultured in RPMI-1640 (Invitrogen) supplemented with 10% FBS (Sigma) and 1% streptomycin and penicillin. HA1E immortalized cell line was cultured in alpha-MEM (Invitrogen) supplemented with 10% FBS. Mouse cell lines and 293T cells were cultured in DMEM (Invitrogen) supplemented with 10% FBS.

Table 4-1. Primer sequences used for qRT-PCR in Chapter 4.

Gene	Target	Species	Sequence
SNAI2 (Slug)	Promoter DNA	Human	5'-TGGGGCATGTCATTACACAG 3'-GGCTTGCGTTTTTACCACAT
SNAI2 (Slug)	3' Control DNA	Human	5'-CCCCTTGCCAGATGTTTCTA 3'-TGGCTGATAGCTTGACTGGA
VIM	Promoter DNA	Human	5'-GGCCCAGCTGTAAGTTGGTA 3'-CCTAGCGGTTTAGGGGAAAC
VIM	3' Control DNA	Human	5'-TTCACTGCAGCCAACAACTC 3'-GCCAGGTGTGTGGCTAGTTT
VIM	mRNA	Human	5'-GTGGACCAGCTAACCAACGACAAA-3 3'-TTCAAGGTCAAGACGTGCCAGAGA-3
FN1	mRNA	Human	5'-CGGTGGCTGTCAGTCAAAG 3'-AAACCTCGGCTTCCTCCATAA
SNAI2 (Slug)	mRNA	Human	5'-ACCTTGTGTTTGCAAGATCTGCGG-3 3'-TGCAAATGCTCTGTTGCAGTGAGG-3
ZEB1	mRNA	Human	5'-GATGATGAATGCGAGTCAGATGC 3'-ACAGCAGTGTCTTGTTGTTGT
CDH1	mRNA	Human	5'-CTGGGACTCCACCTACAGAAAGTT-3 3'-GAGGAGTTGGGAAATGTGAGCA-3
OCLN	mRNA	Human	5'-ACAAGCGGTTTTATCCAGAGTC 3'-GTCATCCACAGGCGAAGTTAAT
FOS	mRNA	Human	5'-CCGGGG ATAGCCTCTCTTACT 3'-CCAGGTCCGTGCAGAAGTC
SNAI1	mRNA	Human	5'-TCGGAAGCCTAACTACAGCGA 3'-AGATGAGCATTGGCAGCGAG
SOX10	mRNA	Human	5'-CCTCACAGATCGCCTACACC 3'-CATATAGGAGAAGGCCGAGTAGA
CDH2	mRNA	Human	5'-TCAGGCGTCTGTAGAGGCTT 3'-ATGCACATCCTTCGATAAGACTG
ZEB2	mRNA	Human	5'-CAAGAGGCGCAAACAAGCC 3'-GGTTGGCAATACCGTCATCC
TWIST1	mRNA	Human	5'-CTGCCCTCGGACAAGCTGAG 3'-CTAGTGGGACGCGGACATGG
Yap1	mRNA	Mouse	5'-ATGACAACCAATAGTTCCGATCC 3'-CAGGGTGCTTTGGCTGAT
Actb	mRNA	Mouse	5'-CTTTGCAGCTCCTTCGTTG 3'-GATGGAGGGGAATACAGCCC

Lentiviral Infection

shRNA in pLKO.1 backbone was obtained directly from the RNAi Consortium (Root et al., 2006). Detailed shRNA identities are listed in Table 4-2. ORFs were cloned into plx-304 and do not express C-terminal V5 tag unless otherwise noted. Lentivirus was produced in 293T cells using the three-vector system as described (Moffat et al., 2006). Cells were infected in media containing 8ug/ml polybrene and 1:20 dilution of virus for shRNA or 1:5 dilution of virus for ORFs. Infected cells were selected with 2 ug/ml puromycin for 2 days or 10 ug/ml blasticidin for 5 days depending on the selection marker.

Table 4-2. Identities of shRNA used in Chapter 4.

shRNA	Source	Species	Target sequence
shKRAS	Broad (TRCN_010369)	human	CAGTTGAGACCTTCTAATTGG
shYAP1-1	Broad (TRCN_107266)	human	GCCACCAAGCTAGATAAAGAA
shYAP1-2	Broad (TRCN_107265)	human	CCCAGTTAAATGTTCACCAAT
shMYC-1	Broad (TRCN_039639)	human	CCCAAGGTAGTTATCCTTAAA
shMYC-2	Broad (TRCN_039640)	human	AATGTCAAGAGGCGAACACA
shFOS-1	Broad (TRCN_016004)	human	GCGGAGACAGACCAACTAGAA
shFOS-2	Broad (TRCN_273940)	human	TCTCCAGTGCCAACTTCATTC
shSLUG-1	Broad (TRCN_15389)	human	CCCATTCTGATGTAAAGAAAT
shSLUG-2	Broad (TRCN_15388)	human	GCCAAATCATTTCAACTGAAA
shLuciferase	Broad (TRCN_072261)	control	CACTCGGATATTTGATATGTG
shYap1-2		mouse	TGCTGTTGACAGTGAGCGCAAGATACTT CTTAAATCACAATAGTGAAGCCACAGAT GTATTGTGATTTAAGAAGTATCTTTTGCC TACTGCCTCGGA
shYap1-1		mouse	TGCTGTTGACAGTGAGCGAATGAATTAA CTCTGTGTATAATAGTGAAGCCACAGAT GTATTATACACAGAGTTAATTCATGTGC CTACTGCCTCGGA
shKras		mouse	TGCTGTTGACAGTGAGCGACAGACCCA GTATGAAATAGTATAGTGAAGCCACAGA TGTATACTATTTCATACTGGGTCTGCTG CCTACTGCCTCGGA
shRenilla		control	TGCTGTTGACAGTGAGCGCAGGAATTAT AATGCTTATCTATAGTGAAGCCACAGAT GTATAGATAAGCATTATAATTCCTATGCC TACTGCCTCGGA

4.3 Identification of YAP1 function in KRAS rescue screen

4.3.1 Systematic identification of genes that rescue the loss of oncogenic KRAS expression

We performed a genome scale genetic rescue screen to identify genes that support the survival of *KRAS*-dependent cancer cells upon suppression of *KRAS*. We first generated a cell line for screening by stably introducing a doxycycline-inducible shRNA targeting the *KRAS* 3' untranslated region (UTR) into the HCT116 *KRAS*-mutant colon cancer cell line (referred to as HCTtetK). We introduced 15,294 ORFs from the Center for Cancer Systems Biology (CCSB)/Broad Institute lentivirally delivered expression library (Yang et al., 2011) into these cells in an arrayed format in triplicate under optimized conditions where assayed plates showed 98% transduction in each well on average. We induced the suppression of KRAS by doxycycline and assessed cell proliferation/survival after five days (Fig. 4-1a). As a positive control, we expressed a mutant $KRAS^{G13D}$ ORF, which lacks the *KRAS* 3'UTR and thus cannot be suppressed (Fig. 4-1b). We considered an ORF a 'hit' if it obtained a KRAS rescue score greater than 3, i.e. the viability in that well was at least 3 standard deviations above the mean of negative controls on each plate after B-score normalization (Brideau et al., 2003). All of the 150 wells containing cells expressing $KRAS^{G13D}$ scored above this threshold, and only 1 of the 1,119 negative control wells (0.05%) scored above this threshold.

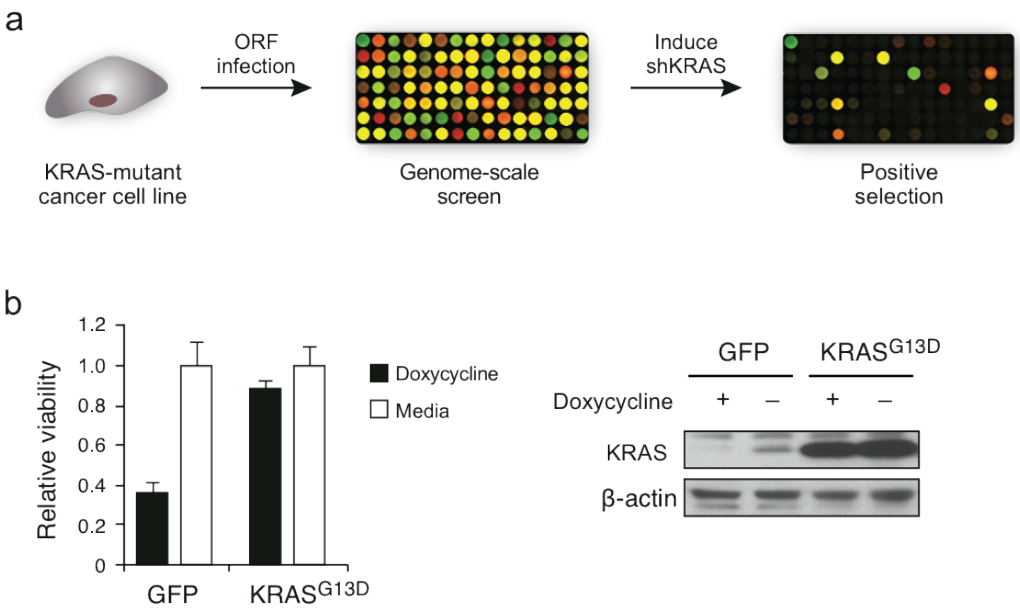

Figure 4-1. Systematic identification of genes that rescue loss of viability induced by KRAS suppression. (**a**) Schematic diagram of an arrayed format screen to identify ORFs that can rescue loss of cell viability induced by suppression of KRAS in KRAS-dependent cells. (**b**) Suppression of KRAS in HCTtetK cells, and rescue by KRAS ORF. Data represented as mean +/- SD normalized to cell viability in untreated conditions.

In total, we identified 147 genes that met this criterion (Appendix 2). Among the highest scoring candidates, we observed genes whose products included the sterile alpha motif (SAM) that function as post-transcriptional regulators (Baez and Boccaccio, 2005), the WW-domain binding proteins YAP1 and WWTR1, and several members of the FGF family (Fig. 4-2). In a separate screen in HCTtetK cells focused on 597 kinases (CCSB/Broad Kinase ORF Collection), we also identified FGFR1 as a kinase that was able to rescue the suppression of KRAS (Fig. 4-3).

Since we expected candidates that emerged from this screen to activate known pathways involved in KRAS signaling, we assessed the ability of each ORF to activate MAPK or PI3K signaling. Specifically, we expressed the 147 ORFs in HCTtetK cells in an arrayed format and quantified the activity of each candidate using an image-based approach (LICOR In-Cell Western). We assessed MAPK activity by measuring the ratio of phospho-ERK to total-ERK levels, and PI3K activity by measuring the ratio of phospho-S6 ribosomal protein to total S6 ribosomal protein levels (Fig. 4-4). We found that the majority of the candidate genes activated at least one of the two pathways (55.1%), with 16.1% activating MAPK only, 13.4% activating PI3K only, and 25.6% activating both pathways. Although we did not expect all of the candidate genes to activate these two KRAS effector pathways, the observation that a large proportion of these candidates activated these pathways increased our confidence in the biological relevance of the rescue screen.

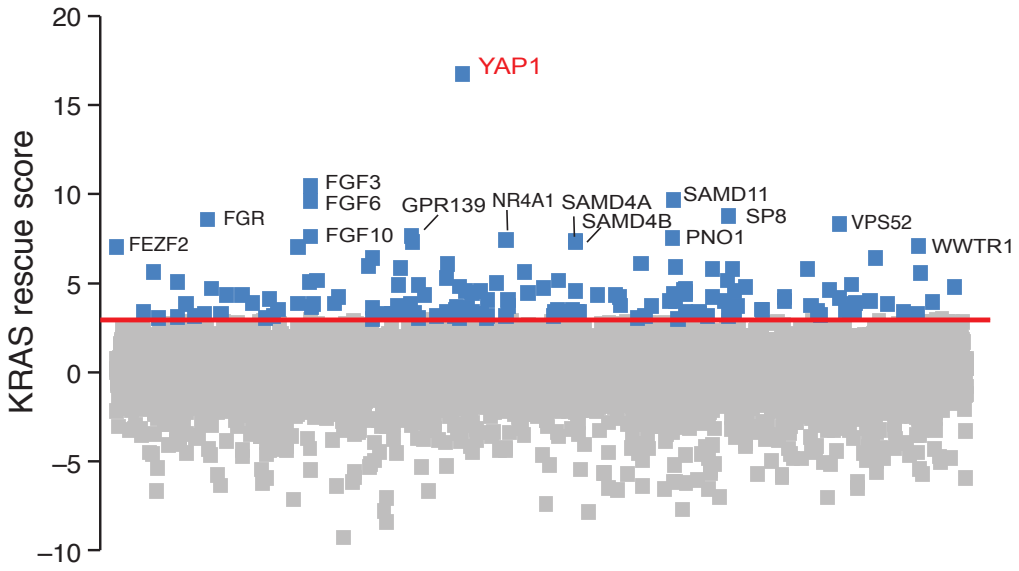

Figure 4-2. Results of rescue screen using genome-scale ORF library. Distribution of scores for all screened genes averaged across 3 replicates. KRAS rescue score indicates standard deviations from mean of negative control wells after B-score normalization by plate (Brideau et al., 2003). Red line, 3 SD. Blue points, gene 'hits.'

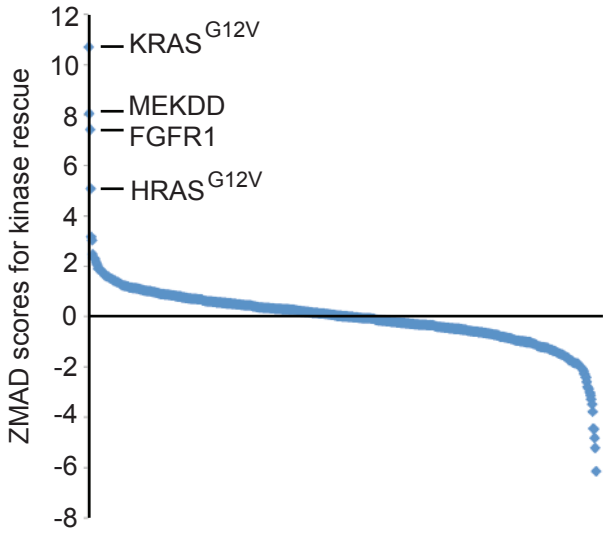

Figure 4-3. Results of rescue screen using kinase ORF library. KRAS rescue screen using kinase ORF library of ~600 kinases, including ORFs harboring common oncogenic activating mutations. Screen was performed in the same manner as described for Figure 1. Genes that score 3SD above negative controls are labeled.

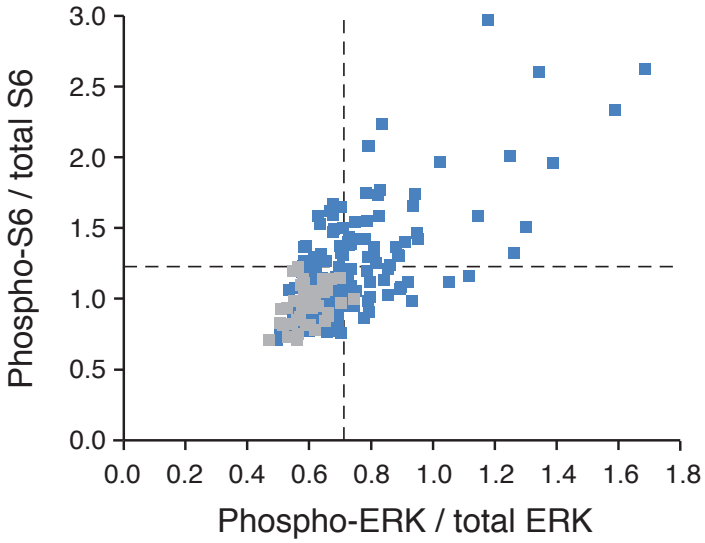

Figure 4-4. Characterization of 147 hits by assessment of ERK and S6 phosphorylation state. Each point represents the ratio of signal intensity of phosphorylated versus total protein by in-cell western. Duplicate wells were averaged. Lines indicate 2SD above mean of negative controls. Gray, negative controls.

4.3.2 YAP1 rescue oncogenic KRAS suppression in KRAS-dependent cancer cell lines

We focused our efforts on understanding *YAP1*, the highest scoring gene in the screen. Expression of the *YAP1*-2γ (hereafter referred to as *YAP1*)(Sudol, 2012) ORF in HCTtetK cells prevented the morphological changes that characterize suppression of KRAS (Fig. 4-5a), rescued the loss of viability induced by suppressing KRAS (Fig. 4-5b), and maintained phospho-ERK and phospho-S6 levels (Fig. 4-5c).

To ensure that the observed effects were not unique to HCT116 cells, we tested the ability of YAP1 to complement the loss of KRAS function in different cell contexts by using two additional *KRAS*-mutant colon cancer cell lines (SW480, LS513) and two additional *KRAS*-mutant pancreatic cancer cell lines (SU86.86, AsPC-1). We found that wild-type YAP1 rescued the loss of viability induced by KRAS suppression in the LS513, SU86.86, and AsPC-1 cell lines, and that a constitutively active version of YAP1, which harbored mutations on 5 serine residues to prevent serine phosphorylation ($YAP1^{5SA}$)(Zhao et al., 2007), rescued cell death in SW480 (Fig. 4-6). We concluded that YAP1 signaling functionally replaces KRAS in different *KRAS*-dependent cancer cells, although YAP1 itself may be differentially regulated in different cell lines.

To determine whether the survival phenotype observed upon expression of YAP1 was specific to KRAS, we assessed whether YAP1 expression also rescued suppression of the MYC oncogene, which is amplified in HCT116 cells. Suppression of MYC using two MYC-specific shRNAs reduced cell viability. Expression of MYC but not YAP1 rescued this phenotype (Fig. 4-7). This observation confirmed that YAP1 expression specifically prevents loss of viability induced by KRAS suppression in cancer cell lines that depend on KRAS.

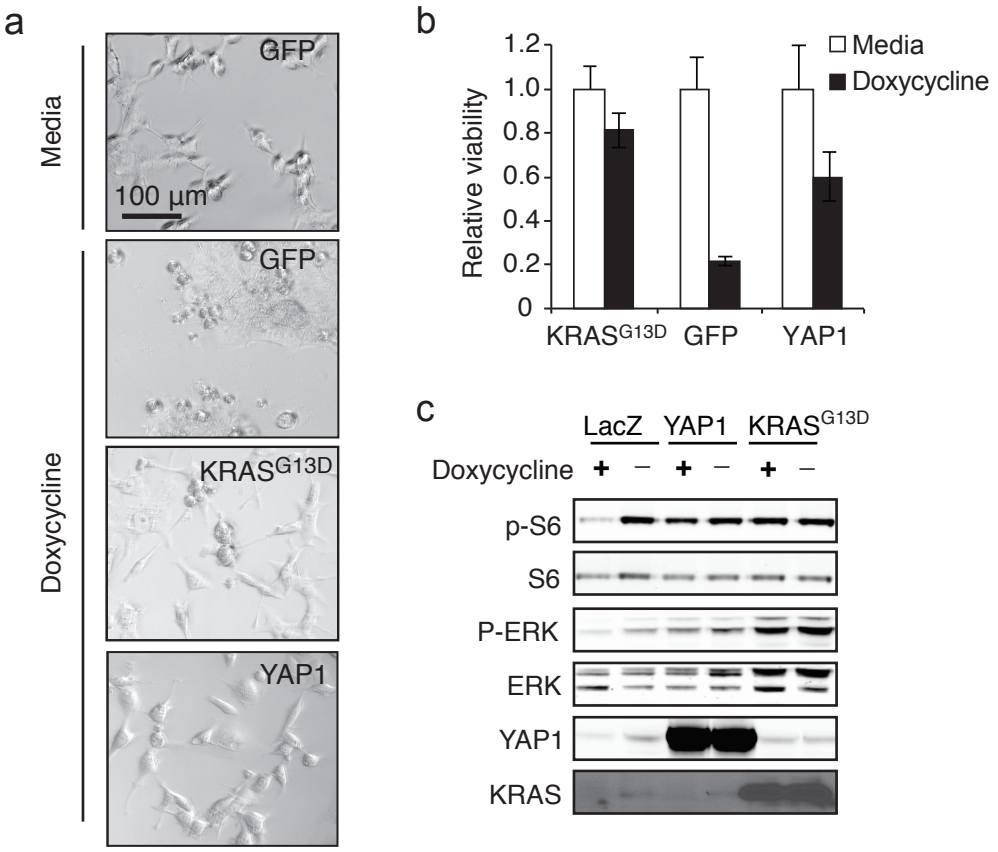

Figure 4-5. YAP1 rescues KRAS suppression in HCTtetK cells. (**a**) Morphology of HCTtetK cells expressing the indicated vectors at 20x magnification. The indicated ORFs were expressed and cells were treated with doxycycline (KRAS suppressed). (**b**) Viability of HCTtetK cells upon KRAS suppression in cells expressing the indicated genes. Data normalized to cell viability in media condition. Data from bar plot represents mean +/- SD for triplicate experiments. (**c**) Signaling changes upon suppression of KRAS and YAP1 rescue.

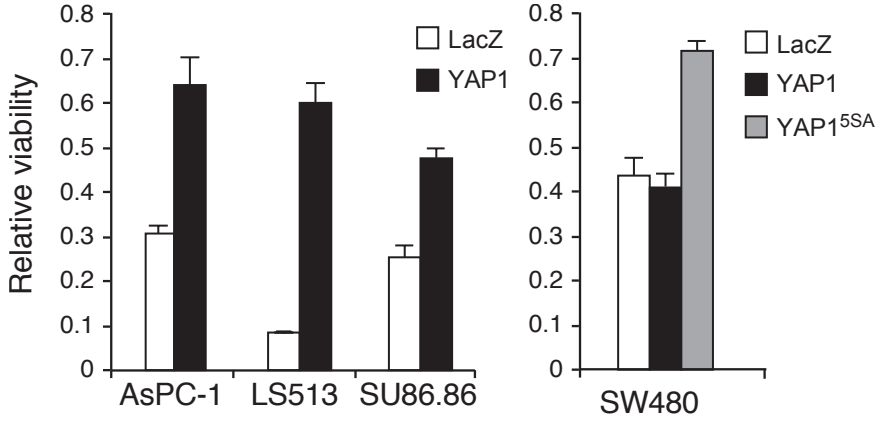

Figure 4-6. YAP1 rescues KRAS suppression in several *KRAS* mutant cancer cell lines. Consequences of expressing YAP1 in KRAS-mutant cell lines after KRAS suppression. Viability after expression of shKRAS normalized to shLuciferase control in the presence of each indicated ORF. Data from bar plots represented as mean +/- SD for triplicate experiments.

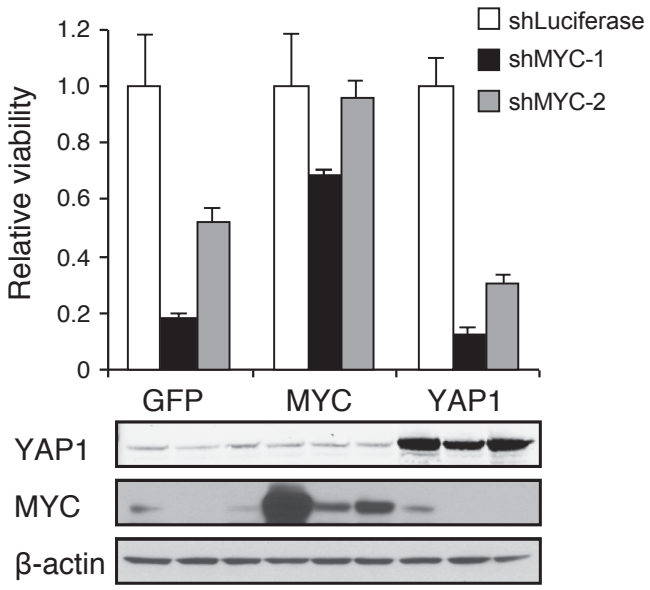

Figure 4-7. YAP1 expression does not rescue MYC suppression in HCT116. Response of HCT116 cells to MYC suppression in cells that express the indicated ORFs. Data from bar plot represents mean +/- SD for triplicate experiments.

To interrogate the functional relationship between KRAS and YAP1 further, we assessed whether YAP1 is required for KRAS-induced cell transformation. We used an anchorage independent growth assay to determine whether suppression of YAP1 using two YAP1-specific shRNAs abolished *KRAS*-driven transformation. We expressed $KRAS^{G13D}$ or *YAP1* cDNA in immortalized HA1E cells (Hahn et al., 1999) and confirmed that either $KRAS^{G13D}$ or YAP1 induced colony formation when a control shRNA was expressed. In this experimental model, we found that expression of both of the *YAP1*-specific shRNAs abrogated KRAS-driven anchorage independent colony formation (Fig. 4-8). Expression of shYAP1-2 targets the *YAP1* 3'UTR and, as expected, does not suppress anchorage independent colony formation driven by the YAP1 ORF.

Manipulating KRAS expression did not affect phosphorylation of YAP1 serine-127, a site previously implicated in regulation of YAP1 by Hpo signaling (Zhao et al., 2007), nor did these manipulations affect phosphorylation of components of the Hpo signaling cascade such as LATS-1/2 and MST2 (Fig. 4-9a). Furthermore, we found that YAP1 expression had no effect on the activation of KRAS as assessed by GTP-bound KRAS levels after pull-down with the Ras-binding domain of RAF1 (Fig. 4-9b). These observations indicated that although YAP1 expression can specifically rescue cells from the effects of KRAS suppression, KRAS and YAP1 do not directly activate each other.

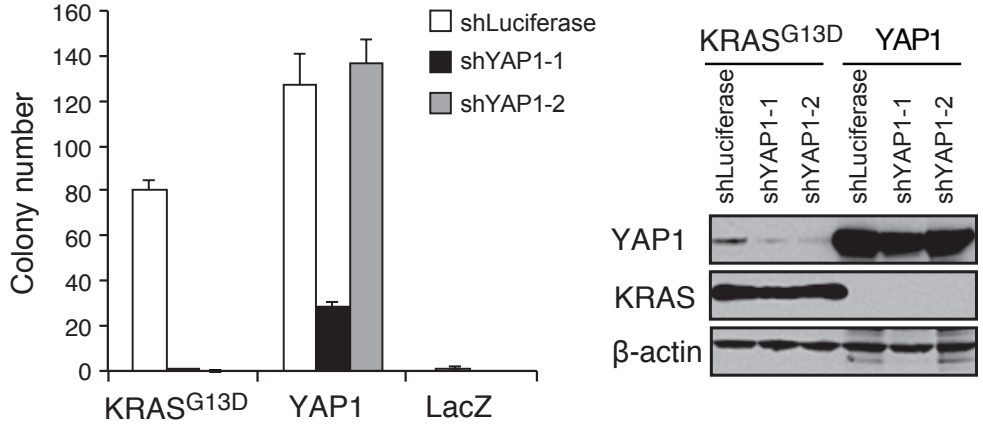

Figure 4-8. Effect of YAP1 suppression on anchorage independent growth. YAP1-specific shRNA were expressed in HA1E transformed with $KRAS^{G13D}$ or YAP1 ORF to assess effects on soft agar colony formation. Data from bar plot represents mean +/- SD for triplicate experiments. Corresponding immunoblots of YAP1 and KRAS expression are shown.

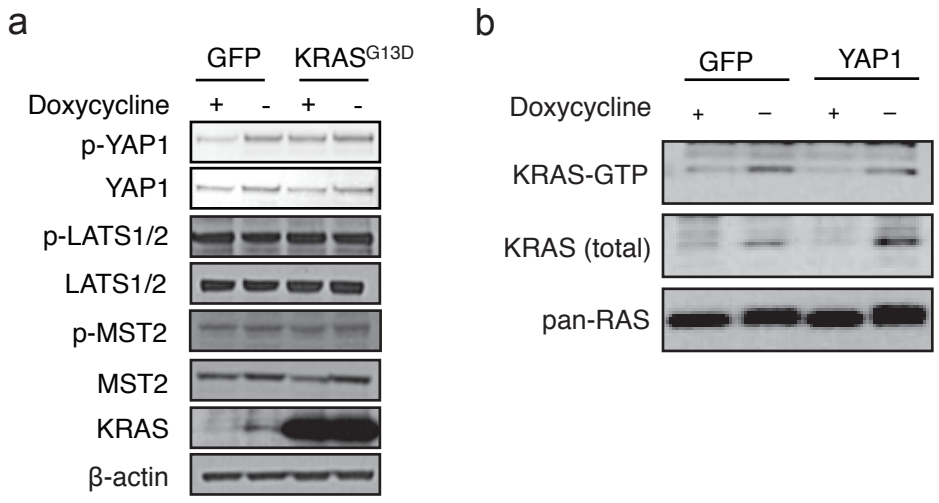

Figure 4-9. Biochemical assessment of Hippo cascade activity and KRAS activity. (**a**) Assessment of Hpo pathway or YAP1 phosphorylation by suppression or expression of KRAS. (**b**) Assessment of GTP-bound KRAS levels after YAP1 expression.

4.3.3 Yap1 mediates resistance to Kras suppression in a Kras^{G12D}-driven lung cancer mouse model

In order to investigate a role for the YAP pathway in mediating resistance to Kras suppression further, we examined the consequences of K-ras inhibition in a well-studied mouse model of lung adenocarcinoma (Tuveson et al., 2004; DuPage et al., 2009). We first used primary lung adenocarcinoma cells derived from the *Kras*$^{Lox\text{-}STOP\text{-}Lox\text{-}G12D}$*;p53*$^{flox/flox}$ (KP) mouse lung cancer model (DuPage et al., 2009) into which we introduced a doxycycline-inducible shRNA targeting *Kras* expressed from the 3'UTR of GFP (hereafter referred to as KP-KrasA cells). In this system, doxycycline treatment activates the GFP reporter as well as shKras, resulting in suppression of endogenous wild-type Kras and mutant Kras^{G12D}. After intravenous injection of tumor cells into recipient mice, tumor burden in the lung was monitored weekly by a constitutively expressed luciferase construct in the tumor cells (Fig. 4-10a). Seven days post-transplantation, doxycycline-containing diet was administered to the recipient mice. This treatment resulted in rapid lung tumor regression within 7 days (Fig. 4-10b), suggesting lung tumors in this model depend on ongoing oncogenic Kras signaling. Surprisingly, however, tumors recurred over the course of the next 2 weeks even though Kras remained suppressed in tumor tissue, as assessed by qRT-PCR of *Kras* mRNA from microdissected tumors (Fig. 4-10c). Thus, in this model as well some Kras-driven tumor cells can continue to proliferate despite Kras-independent after prolonged suppression of the oncogene.

To assess the molecular basis of this Kras-independent process, KP-KrasA cells were cultured for 21 days in the presence of doxycycline, resulting in cells that continued to grow despite suppression of Kras. RNA sequence profiling (RNA-seq) of these cells after 21 days on doxycycline compared to cells without exposure to doxycycline showed

significant up-regulation of a published Yap1 gene signature (Fig. 4-10d)(Taube et al., 2010). This observation suggests that Kras-independent mouse lung cancer cells similarly use YAP1 activity similarly to the effect that was observed when we expressed YAP1 to rescue KRAS suppression in human cancer cell lines.

To explore whether Yap1 signaling was involved in Kras oncogenic addiction *in vivo*, we performed both gain-of-function and loss-of-function experiments in this mouse model. We observed that forced expression of YAP1 partially prevented the tumor regression observed upon initial suppression of Kras (Fig. 4-11a, 4-11b). To test whether relapse from Kras suppression involves Yap1 signaling, we generated KP cells in which two shRNAs are expressed simultaneously. Specifically, in addition to the Kras-specific shRNA, we expressed a doxycycline inducible constructs that drives the expression of red fluorescent protein (RFP) carrying shRNAs targeting either Yap1 or Renilla luciferase within the 3'UTR (Zuber et al., 2011). While tumors with Kras suppression alone relapsed over time, the combination of Yap1 suppression delayed tumor relapse while the expression of the Renilla luciferase shRNA did not (Fig. 4-12a). We found that the residual tumors that formed in the presence of the Yap1-specific shRNA after 28 days no longer suppressed *Yap1* as assessed by qRT-PCR (Fig. 4-12b), suggesting that suppression of Kras *in vivo* selects for cells with less effective Yap1 suppression. These results confirm that upregulation of Yap1 signaling plays a compensatory role *in vivo* upon loss of Kras signaling.

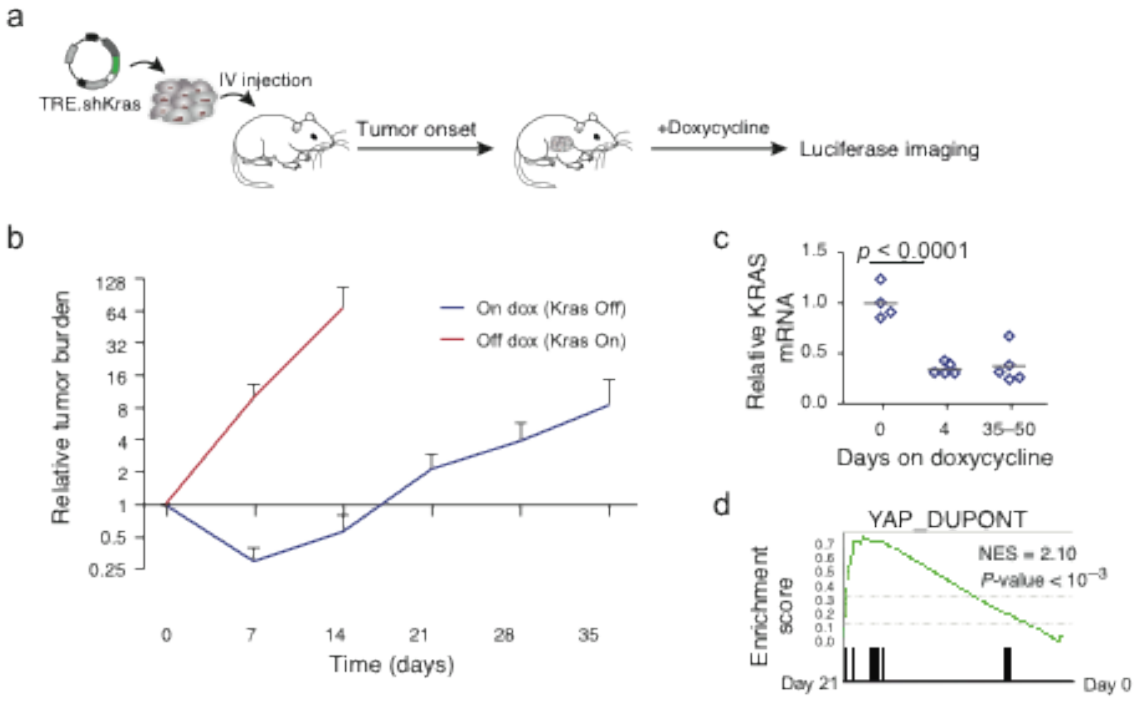

Figure 4-10. Yap1 activity is increased in a mouse model of Kras-independent lung cancer. (**a**) Schematic of mouse transplant model of KRAS-driven lung cancer. *Kras*^{G12D}*;p53*$^{fl/fl}$ lung adenocarcinoma cells were infected with retroviral vectors expressing rtTA3, luciferase and a tet-on shKras. Cells were transplanted into recipient mice by tail vein injection. 7 days later, mice were treated with doxycycline diet to induce shKras in tumor cells (D0). (**b**) Time course of tumor regression and relapse after Kras suppression. Data represent mean +/- SD. N=3 off dox and N=10 on dox. (**c**) Suppression of Kras in tumor tissue. Kras mRNA was measured by qRT-PCR in microdissected lung tumors after the indicated days of doxycycline treatment. (**d**) Enrichment of a published YAP1 signature (Taube et al., 2010) after 21 days doxycycline treatment versus untreated cells.

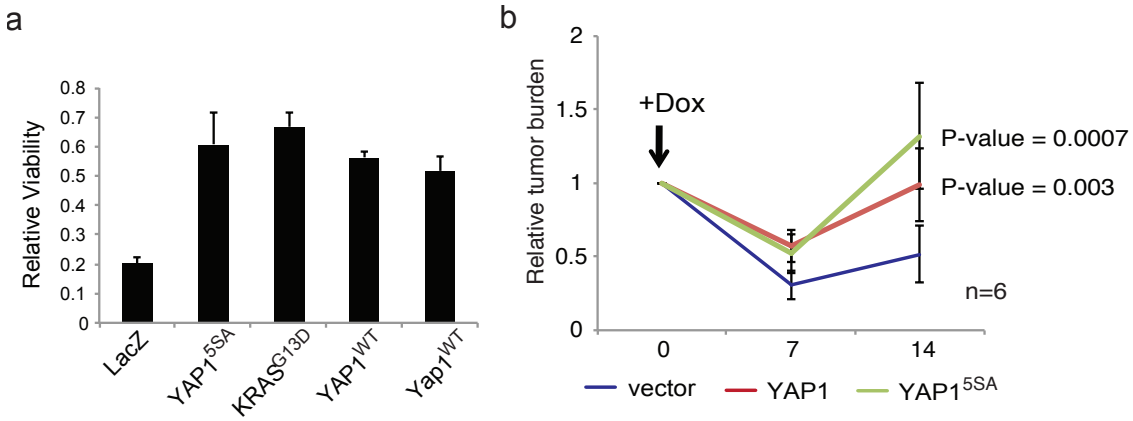

Figure 4-11. Yap1 rescues of Kras suppression in KP-KrasA cells. (a) *In vitro* rescue of Kras suppression by human YAP1 and mouse Yap1. Relative viability is shown after suppression of Kras (dox) compared to without Kras suppression (media). Data represent mean +/- 1SD, n=3. (D) *In vivo* rescue of Kras suppression in KP-KrasA cells in orthotopic transplant model. Tumor burden measured by luciferase imaging.

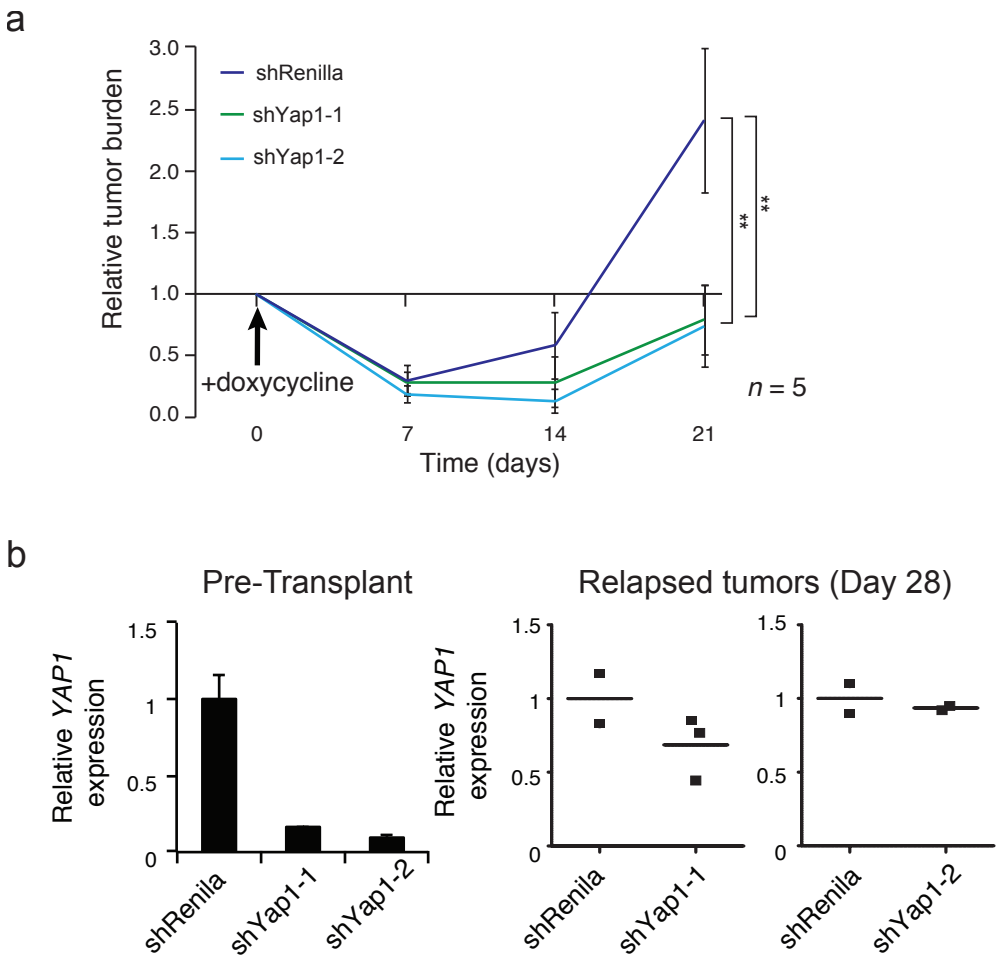

Figure 4-12. Yap1 suppression prevents tumor recurrence after Kras suppression. (**a**) Tumor response to suppression of Kras in combination with Yap1 or control suppression. (**b**) Yap1 expression level in parental cell line (Pre-Transplant) expressing shYap1 compared to *in vivo* tumors that arise (Relapsed tumors). Bars indicate mean +/- 1SD. Points indicate YAP1 expression in individual relapsed micro-dissected tumors corresponding to day 28.

4.4 Mechanism of YAP1 activity

4.4.1 Functions of YAP1 required for the survival of KRAS-dependent cells

YAP1 is regulated by several signaling pathways and regulates the function of different downstream transcription factors. YAP1 is composed of several domains including the TEAD-binding domain, an SH3 binding motif, two WW domains and a transcription activation domain (Fig. 4-13). YAP1 localization to the nucleus is mediated by its C-terminal PDZ domain-binding motif (Oka and Sudol, 2009). To identify regions of YAP1 necessary to promote survival after KRAS suppression, we expressed previously described YAP1 domain-specific mutants in HCTtetK cells to assess their effect on YAP1 function (Fig. 4-13).

First, we explored the role of TEAD transcription factors for the KRAS-related functions of YAP1. Prior work identified YAP1 mutants that disrupt binding to TEAD transcription factors by alterations within the TEAD-binding domain ($YAP1^{S94A}$ and $YAP1^{\Delta 60-89}$) (Zhao et al., 2008b; Cao et al., 2008). We confirmed that expression of these mutants abolished the ability of YAP1 to activate a TEAD-specific reporter (Fig. 4-14a)(Ota and Sasaki, 2008). These YAP1 mutants rescued the proliferation effects of KRAS suppression similar to the level of wild-type YAP1 (Fig. 4-14b). Accordingly, expression of a constitutively active fusion of the VP16 domain to the DNA-binding region of TEAD1 (TEAD1-VP16)(Cao et al., 2008) failed to rescue the effect of KRAS suppression in HCTtetK cells (Fig. 4-14c). Together, these results suggest that the TEAD family is not the primary mediator of the ability of YAP1 to complement loss of KRAS.

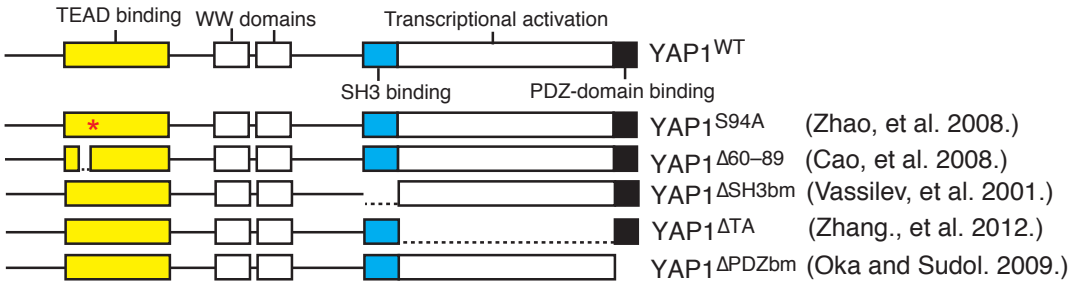

Figure 4-13. Schematic diagram of YAP1 domain structure and YAP1 mutants.

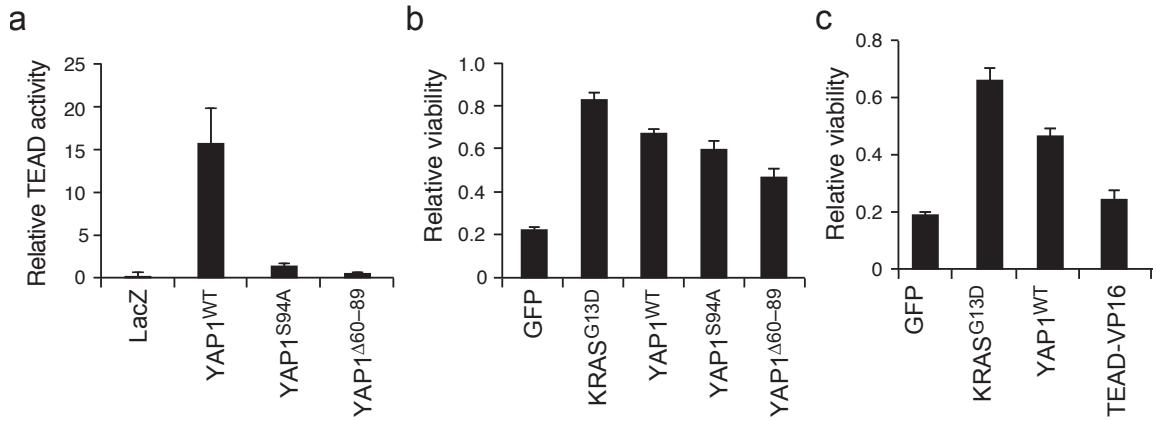

Figure 4-14. TEAD binding is not required for the functional ability of YAP1 to replace KRAS signaling. (**a**) Effects of expressing YAP1 TEAD-defective mutants on the activity of a TEAD reporter in 293T cells. (**b**) Consequences of expressing YAP1 mutants defective in TEAD activation on viability after KRAS suppression in HCTtetK cells. Viability of doxycycline treated (KRAS suppressed) relative to untreated samples displayed. (**c**) Effects of expressing a constitutively active TEAD1-VP16 fusion in HCTtetK cells after KRAS suppression. Viability of doxycycline treated (KRAS suppressed) relative to untreated samples displayed. (**a**-**c**) Data represent mean +/- SD for triplicate experiments.

Since we recently reported a role for YES1 phosphorylation of YAP1 for β-catenin signaling (Rosenbluh et al., 2012), we investigated whether YES1 played a role in KRAS signaling as well. We expressed two YAP1 mutants: $YAP1^{\Delta SH3bm}$, which disrupts the interaction of YES1 with YAP1 (Vassilev et al., 2001; Sudol, 1994), and YAP^{Y357F}, which prevents YES1 phosphorylation of YAP1 (Rosenbluh et al., 2012){Rosenbluh:2012ew}. Expression of either $YAP1^{\Delta SH3bm}$ or $YAP1^{Y357F}$ rescued the loss of cell viability observed after KRAS suppression to the level of wild-type YAP1 (YAP^{WT}; Fig. 4-15a, 4-15b). These observations suggest that YES1 modulation of β-catenin signaling is not functionally important for YAP1 activity in the context of KRAS suppression.

Finally, we tested whether YAP1-induced transcriptional activation was required to rescue KRAS suppression, since the transcriptional activation domain of YAP1 was reported to be dispensable for YAP1-driven transformation of MCF10A cells (Zhang et al., 2012). We found that expression of YAP1 mutants that harbor a deletion of the transcriptional activation domain ($YAP^{\Delta TA}$) disrupted the ability of YAP1 to rescue cells from KRAS suppression in HCTtetK cells (Fig. 4-16a). Deletion of the 5 amino acid PDZ domain-binding motif ($YAP^{\Delta PDZbm}$), which disrupts YAP1 nuclear localization, induced a similar phenotype (Fig. 4-16a). Expression levels of all mutants were comparable to that of wild-type YAP1 (Fig. 4-16b). Indeed, the two mutations that affect the transcriptional activation domain and PDZ domain-binding motif also prevented YAP1-induced transformation (Fig. 4-16c). In aggregate, these observations show that YAP1 mediates survival after suppression of KRAS by modulating transcriptional activity, independent of the regulation of TEAD or β-catenin activity.

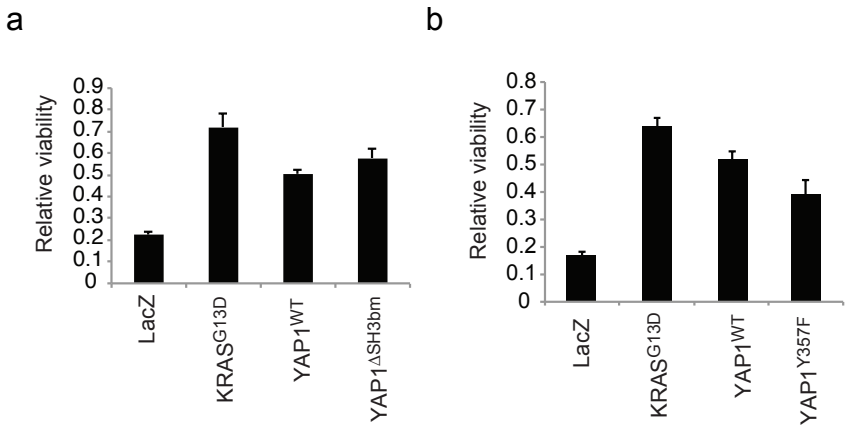

Figure 4-15. YES1 binding is not required for the functional ability of YAP1 to replace KRAS signaling. (**a**) YAP1$^{\Delta SH3bm}$ rescues suppression of KRAS in HCTtetK cells. (**b**) YAP1^{Y357F} rescues suppression of KRAS in HCTtetK cells. (**a**, **b**) Viability of doxycycline treated (KRAS suppressed) relative to untreated samples displayed. Data represent mean +/- SD, n=3.

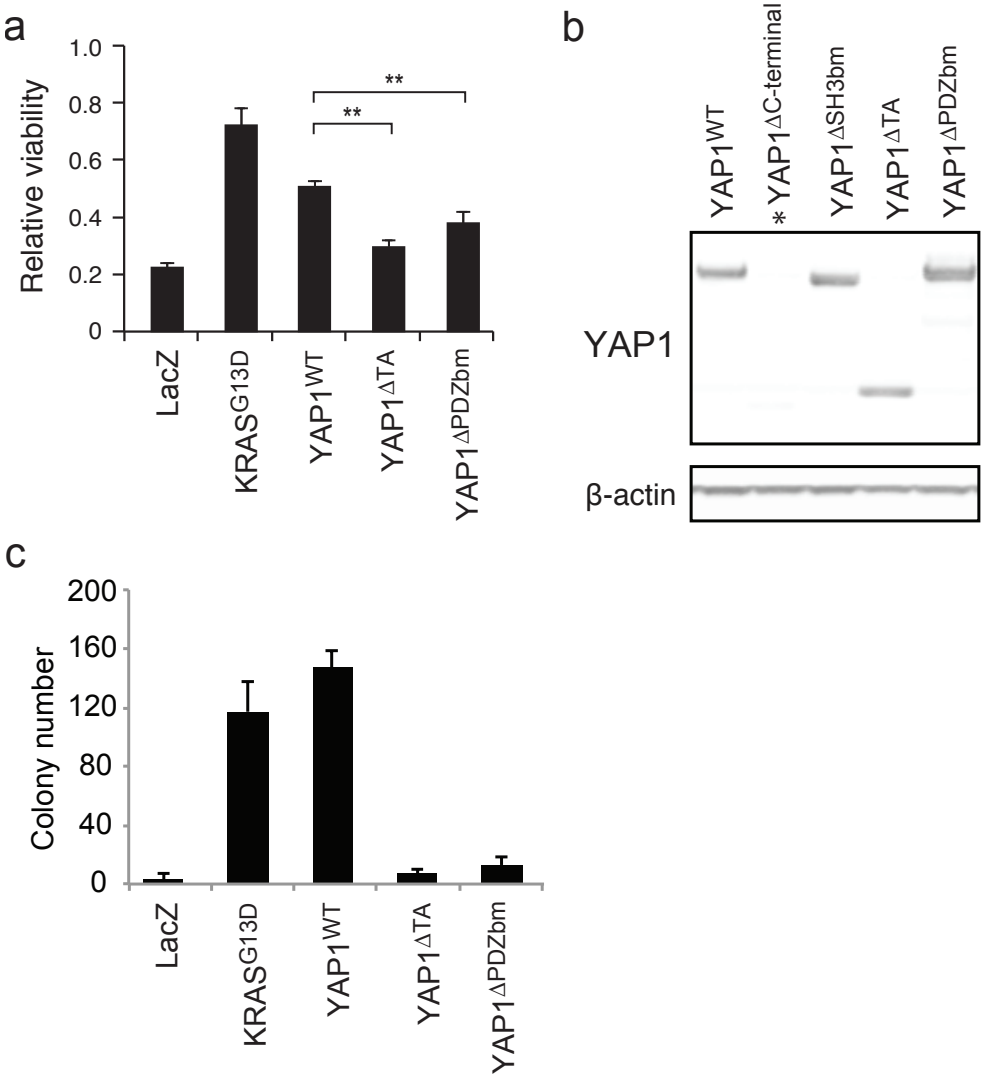

Figure 4-16. YAP1 transcriptional activation domain and PDZ-domain binding motif are required for functional activity. (**a**) Effects of expressing the YAP1 mutants defective in transcriptional activation or nuclear localization in HCTtetK cells after KRAS suppression. **P-value <0.01. Viability of doxycycline treated (KRAS suppressed) relative to untreated samples displayed. (**b**) Immunoblot of expression of YAP1 mutants. *, protein not expressed. (**c**) Ability of YAP1 mutants to replace KRAS in transformation of HA1E as assessed by colony formation assay. (**a**,**c**) Data represent mean +/- SD for triplicate experiments.

4.4.2 YAP1 regulates AP-1 family transcription factors and the epithelial-mesenchymal transition

Having determined that the transcriptional activation functions of YAP1 were required for YAP1 to rescue suppression of KRAS in KRAS-dependent cells, we analyzed transcriptional profiles in cells in which we manipulated KRAS or YAP1 expression. We first identified genes that were downregulated in HCTtetK cells treated with doxycycline (KRAS suppressed) for 30 hours compared to untreated cells. Among these genes, we identified genes whose expression was rescued by the expression of YAP1 or by the expression of KRAS (Fig. 4-17). Using the 1,045 genes that were rescued by both YAP1 and KRAS, we performed analyses to identify both enriched transcription factor motifs and enriched gene sets that may explain the rescue of KRAS oncogenic addiction.

We used TransFind motif analysis (Kielbasa et al., 2010) to identify transcription factor motifs enriched in promoter regions of genes rescued by both YAP1 and KRAS compared to genes rescued by KRAS alone. Enriched motifs (P-values <0.05) revealed a group of transcription factors involved in the immediate early gene response (Fig. 4-18a), previously demonstrated to be a key transcriptional program regulated by both growth factor stimulation and Ras signaling (Healy et al., 2013). Specifically, this category included gene families such as activating transcription factor (ATF), early growth response protein (EGR), and specificity protein (SP). To determine which transcription factors in these gene families play a role in YAP1 function, we systematically suppressed members of these gene families using multiple independent shRNAs to assess the effect of gene suppression on YAP1-induced anchorage independent growth (Fig. 4-18b). We found that suppression of several SP family and ATF family transcription factors repressed YAP1-mediated cell transformation. Genes in

the SP family, in particular Sp1, are required for development and are ubiquitously involved in maintaining cell survival (Suske, 1999). We were particularly interested in the role of the ATF genes since they are members of the AP-1 transcription factor family, previously shown to be regulated by KRAS – MAPK signaling (Hollenhorst et al., 2011; Mechta et al., 1997).

We re-examined our initial genome-scale screening data and noted that *FOS* was the only member of the AP-1 transcription factor family that scored in our initial screen (3.75 SD above controls; Appendix 2). We confirmed that expression of FOS rescued HCTtetK cells upon suppression of KRAS (Fig. 4-19a). To test whether FOS expression was necessary for YAP1 function, we assessed whether suppression of FOS affected YAP1-induced anchorage independent growth in HA1E cells. Suppression of FOS with two specific shRNAs greatly reduced the ability of YAP1 to promote anchorage independent colony formation (Fig. 4-19b). These observations support the role of AP-1 family of transcription factors as effectors of YAP1 in addition to its previously known roles downstream of KRAS.

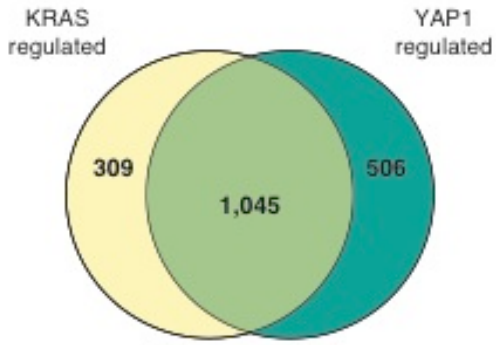

Figure 4-17. Microarray results depicting genes rescued by YAP1, KRAS, or both, in the context of KRAS suppression.

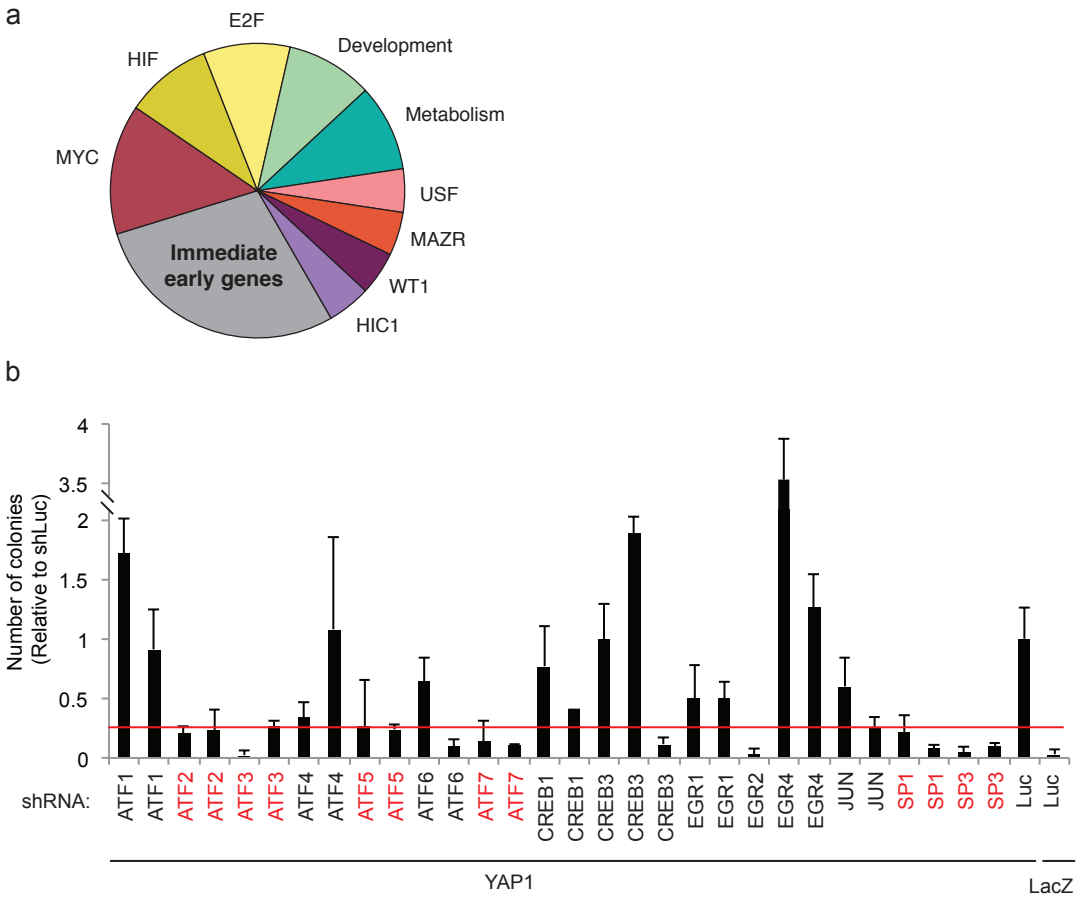

Figure 4-18. YAP1 and KRAS regulated genes share AP-1 motifs. (**a**) Categories of transcription factor motifs enriched amongst genes rescued by both KRAS and YAP1. (**b**) Effect of expressing shRNA targeting gene families ATF, CREB, EGR, and SP on YAP1-induced anchorage-independent colony formation. Data is only shown for shRNAs that suppress the target gene by at least 25% as measured by qRT-PCR. Red color indicates genes for which at least two shRNAs suppress colony formation at least 70% of control shRNA. Data represent mean +/- 1SD, n=3.

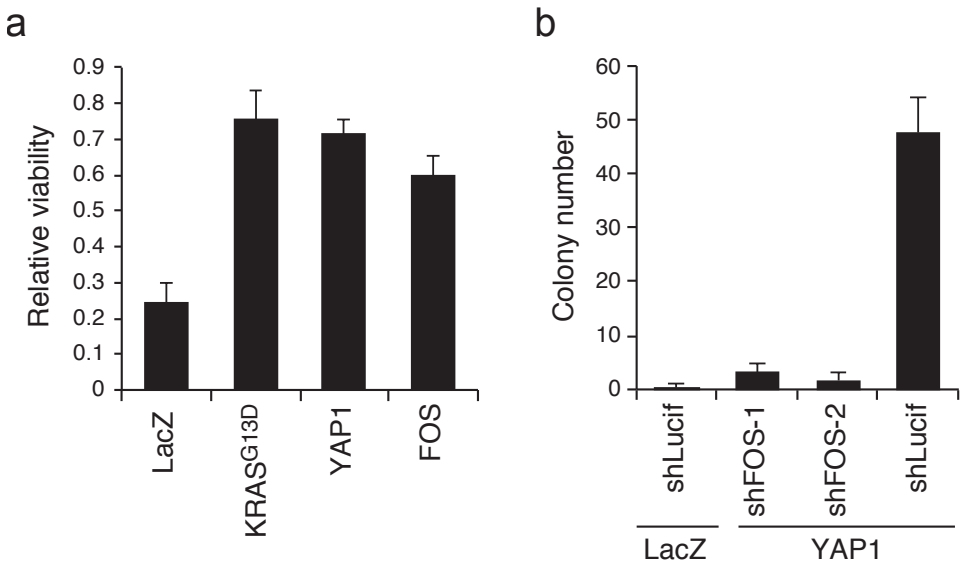

Figure 4-19. FOS replaces KRAS suppression and is required for YAP1-mediated transformation. (**a**) Expression of FOS rescues suppression of KRAS in HCTtetK cells. Viability of doxycycline treated (KRAS suppressed) relative to untreated samples displayed. (**b**) Effects of suppressing FOS on YAP1-mediated cell transformation of HA1E cells. (**a**, **b**) Data represent mean +/- SD for triplicate experiments.

In addition to transcription factor motifs, we also looked for gene sets that are enriched among the genes whose expression was rescued by both KRAS and YAP1 according to our microarray data (Fig. 4-17) using the Molecular Signatures Database (MSigDB)(Subramanian et al., 2005). We noted that enriched transcriptional motifs are not expected to directly correspond to enriched gene sets due to pleotropic activity of many transcription factors. The top gene sets, enriched to P-value < 10^{-3}, included several related to stem cells and development (Table 4-3), suggesting a role of the epithelial-mesenchymal transition (EMT). In many cancers, normal cell differentiation is often perturbed, by interruption of specific differentiation steps or by the aberrant activation of programs such as EMT. Furthermore, EMT has not only been implicated in resistance to therapies targeting receptor tyrosine kinases (Singh and Settleman, 2010; Witta et al., 2006), but FOS is also known to directly regulate EMT (Reichmann et al., 1992). Thus we hypothesized that EMT induced by YAP1 contributed to the survival of cells after suppression of KRAS.

We found that both KRAS and YAP1 expression strongly induced expression of mesenchymal genes such as Vimentin (*VIM*), Fibronectin (*FN1*), Slug (*SNAI2*), and Zinc-finger E-box-binding homeobox 1 (*ZEB1*) and reduced the expression of epithelial genes such as E-cadherin (*CDH1*) and Occludin (*OCLN*) (Fig. 4-20a). Some genes implicated in EMT were not regulated by either KRAS or YAP1 (Fig. 4-20b), suggesting that KRAS and YAP1 regulated EMT markers in a similar manner. We confirmed that EMT markers were also regulated by expression of YAP1 and KRAS in an additional cancer cell line SU86.86 (Fig. 4-20c).

Table 4-3. Enriched gene sets amongst genes regulated by both YAP1 and KRAS. Gene sets from MSigDB database C2 collection of chemical and genetic perturbations were used (http://www.broadinstitute.org/gsea/msigdb). P-values were calculated by hypergeometric test.

MSigDB Gene Set	P-value
BENPORATH_ES_WITH_H3K27ME3	1.02 × 10–10
BENPORATH_SUZ12_TARGETS	3.12 × 10–7
DOUGLAS_BMI1_TARGETS_UP	6.89 × 10–7
BENPORATH_PRC2_TARGETS	9.46 × 10–7
TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_HSC_DN	2.30 × 10–6
DEURIG_T_CELL_PROLYMPHOCYTIC_LEUKEMIA_UP	5.25 × 10–6
SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_UP	7.44 × 10–6
BENPORATH_EED_TARGETS	8.53 × 10–6
BYSTRYKH_HEMATOPOIESIS_STEM_CELL_QTL_TRANS	1.96 × 10–5
CHEOK_RESPONSE_TO_MERCAPTOPURINE_DN	3.72 × 10–5
FUJII_YBX1_TARGETS_DN	4.35 × 10–5
ONDER_CDH1_TARGETS_2_DN	5.10 × 10–5
KAAB_HEART_ATRIUM_VS_VENTRICLE_UP	6.88 × 10–5
DELYS_THYROID_CANCER_UP	1.26 × 10–4
SWEET_LUNG_CANCER_KRAS_DN	1.32 × 10–4
POOLA_INVASIVE_BREAST_CANCER_UP	1.34 × 10–4
NGUYEN_NOTCH1_TARGETS_DN	1.47 × 10–4
FULCHER_INFLAMMATORY_RESPONSE_LECTIN_VS_LPS_UP	1.64 × 10–4
IVANOVA_HEMATOPOIESIS_STEM_CELL_AND_PROGENITOR	1.69 × 10–4
HELLER_SILENCED_BY_METHYLATION_UP	1.77 × 10–4

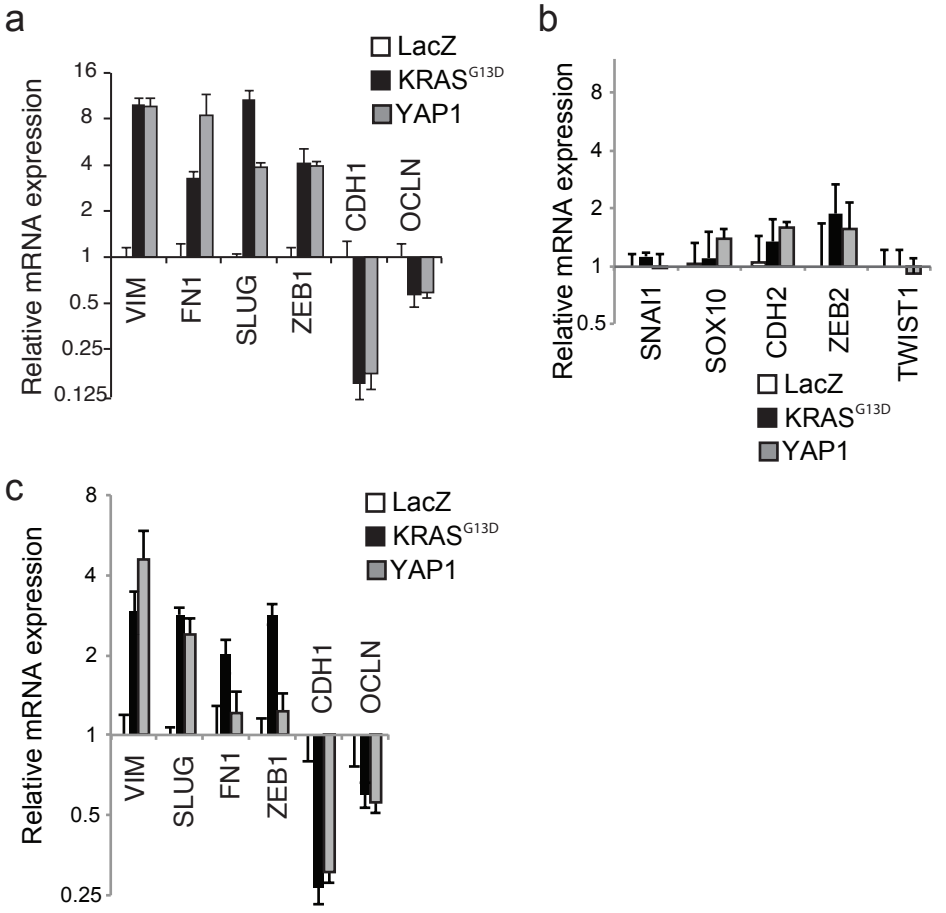

Figure 4-20. YAP1 regulates genes involved in the epithelial-mesenchymal transition (EMT). (**a**) Validation of EMT regulation by KRAS and by YAP1 expression in HCT116 cells. Data represent mean +/- SD relative to LacZ control for triplicate experiments. (**b**) qRT-PCR shows that EMT markers not regulated by KRAS are also not regulated by YAP1. (**c**) qRT-PCR for EMT markers upon KRAS and YAP1 expression in additional pancreatic cancer cell line SU86.86. (**a**-**c**) Data represent mean +/- SD relative to expression in LacZ control, n=3.

We then tested whether key transcriptional regulators of EMT such as Slug and Snail are necessary and sufficient to rescue KRAS dependence. Expression of either Slug or Snail in HCTtetK cells rescued the loss of cell viability induced by suppressing KRAS (Fig. 4-21a). Moreover, we tested whether Slug was required for YAP1 rescue of KRAS suppression. Expression of two Slug-specific shRNAs reduced Slug expression and decreased the ability of YAP1 to rescue suppression of KRAS (Fig. 4-21b, 4-21c). To assess whether these observed effects were specific to KRAS, we also suppressed c-MYC and expressed FOS, Slug, or Snail. Neither expression of EMT transcriptional regulators nor FOS could rescue loss of viability upon MYC suppression (Fig. 4-22), suggesting that induction of EMT did not broadly rescue oncogenic suppression.

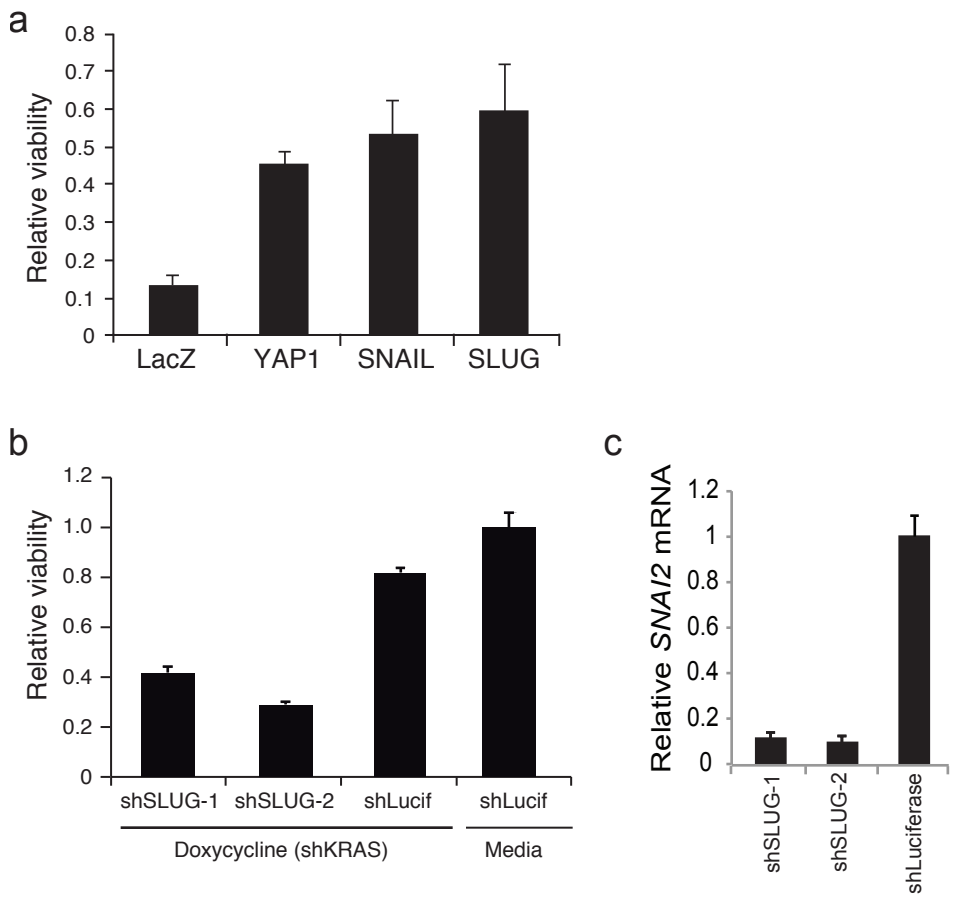

Figure 4-21. Functional role of Slug for KRAS and YAP1 signaling. (**a**) Effects of expressing Slug and Snail after suppression of KRAS in HCTtetK cells. Relative viability of doxycycline treated (KRAS suppressed) normalized to untreated samples displayed. (**b**) Role of Slug in functional ability of YAP1 to replace KRAS. Viability of HCTtetK cells expressing YAP1 together with the indicated shRNA after doxycycline treatment. Data normalized to untreated shLuciferase (shLucif) control. (**c**) Confirmation of Slug suppression by qRT-PCR, corresponding to panel (b). (**a**-**c**) Data represent mean +/- SD, for triplicate experiments.

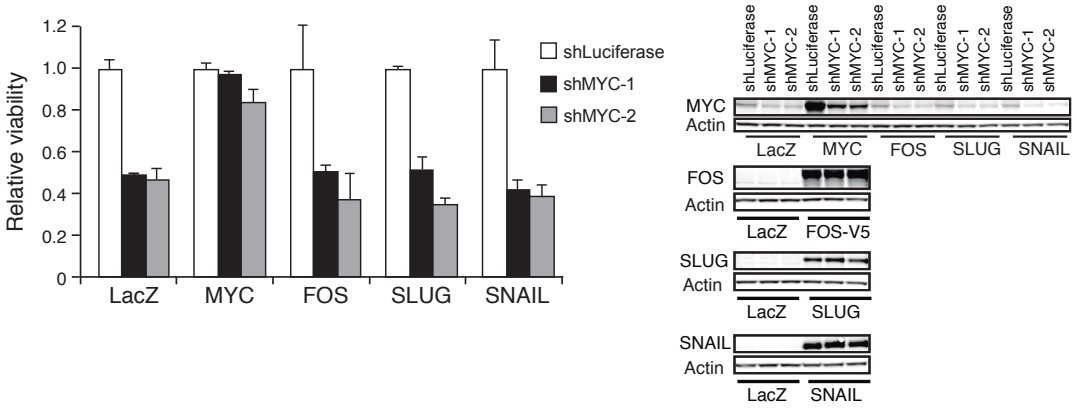

Figure 4-22. Ability of FOS, Slug, and Snail to rescue MYC suppression. Effects of expressing FOS, Slug, and Snail after suppression of MYC in HCT116 cells. Viability after expression of shMYC normalized to shLuciferase control in the presence of each indicated ORF. Corresponding immunoblots are shown.

4.3.3 FOS and YAP1 coordinately regulate downstream targets

Because FOS expression was required for YAP1-induced cell transformation, we examined whether FOS and YAP directly interact. We expressed a V5-epitope tagged FOS in HCT116 cells and isolated YAP1 or control (anti-GFP) immune complexes. In YAP1 complexes but not control immune complexes, we detected FOS (Fig. 4-23a). In addition, when we isolated V5 immune complexes (FOS), we found endogenous YAP1 (Fig. 4-23a). In contrast, when we performed similar co-immunoprecipitation experiments with a V5-epitope tagged version of JUN, a common FOS interaction partner, we failed to find an interaction between YAP1 and JUN (Fig. 4-23b).

We hypothesized that FOS and YAP1 regulate common downstream target genes that are important for KRAS oncogenic addiction. Using transcription factor ChIP-seq data from the Encyclopedia of DNA Elements (ENCODE) (ENCODE Project Consortium et al., 2012) as visualized using the UCSC Genome Browser (http://genome.ucsc.edu), we determined that YAP1-regulated genes such as *VIM* and *Slug* (official gene name *SNAI2*) harbor FOS binding sites in their promoter regions. We confirmed that expression of two independent FOS-specific shRNAs suppressed *FOS* mRNA and also induced suppression of VIM and Slug (Fig. 4-24a). To determine whether FOS and YAP1 bind at the same loci in the genome, we performed chromatin immunoprecipitation using antibodies specific to YAP1, FOS, or control IgG in HCT116 cells. We first confirmed that FOS was enriched at the promoter regions of *VIM* and *Slug* (compared to a region 3' of each gene) as reported by ENCODE (Fig. 4-24b). We then showed that YAP1 binding is also enriched at the same genomic loci as FOS binding, whereas binding of control IgG was not enriched (Fig. 4-25a). We confirmed the specificity of the YAP1 antibody at these loci by showing decreased YAP1 binding at the promoter regions of target genes when YAP1-specific shRNAs were expressed. Notably,

suppression of FOS decreased YAP binding at the *VIM* promoter (Fig. 4-25b), suggesting that YAP likely functions through FOS as a transcriptional co-activator and drives specific genes related to KRAS oncogenic addiction.

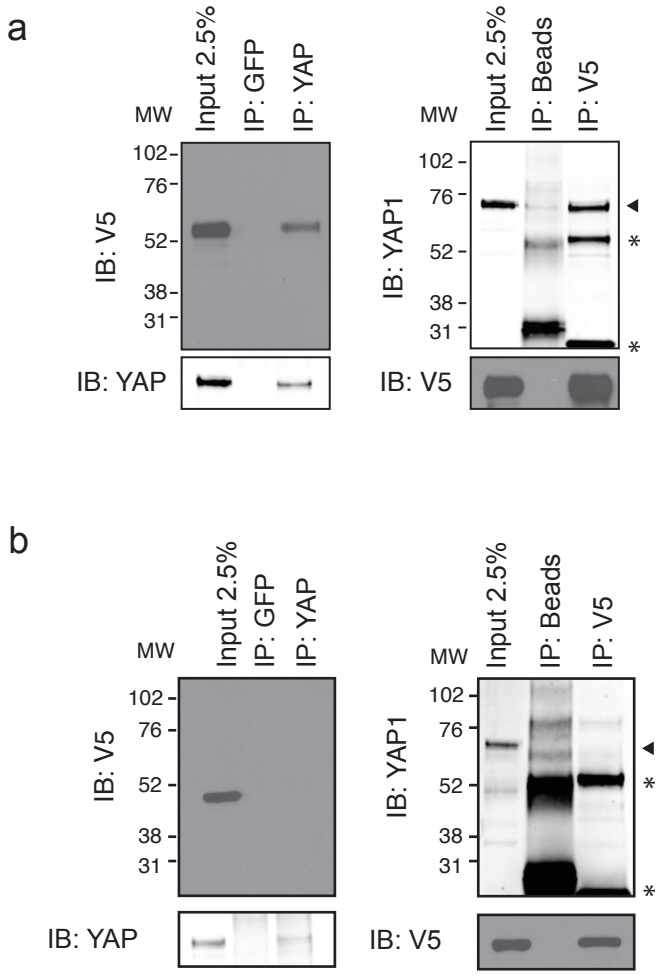

Figure 4-23. YAP1 interacts with FOS in the nucleus. Co-immunoprecipitation was performed using control antibody or target-specific antibody using lysates from HCT116 cells expressing YAP1 and V5-tagged FOS (**a**), or V5-tagged JUN (**b**). Binding of the reciprocal protein was assessed by immunoblotting. YAP1 indicated by arrowhead. *, IgG heavy and light chains. MW, molecular weight in kDa.

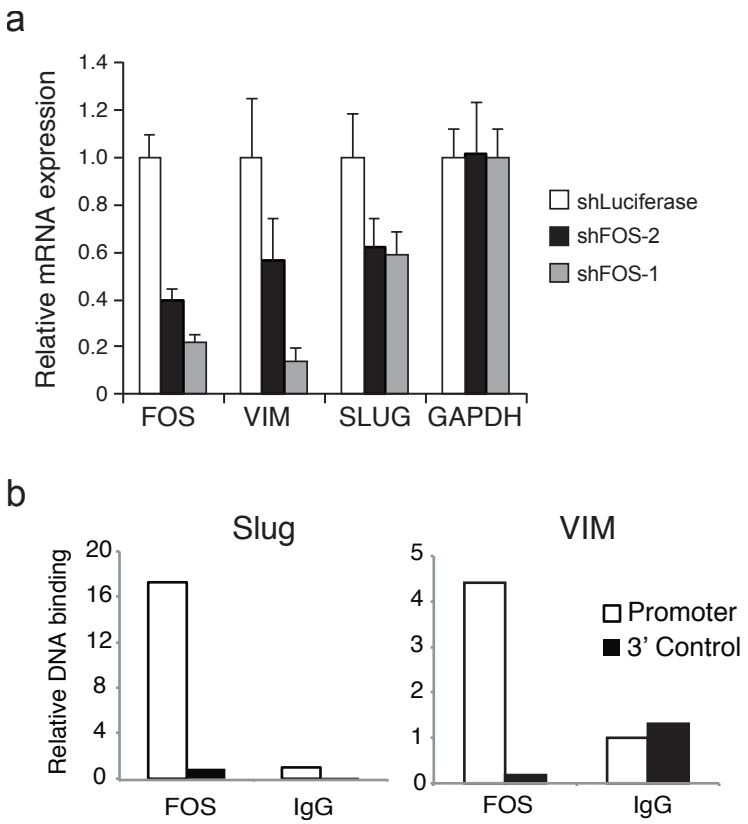

Figure 4-24. FOS binds to promoter regions of *VIM* and *SLUG* and regulates their expression. (B) FOS regulates Vimentin and Slug mRNA levels. Data from HCT116, represents mean +/- SD relative to shLuciferase control in triplicate experiments. (B) Confirmation of FOS binding sites identified by ENCODE data in promoter regions of *VIM* and *SLUG*. Chromatin immunoprecipitation was performed using antibody targeting FOS or control IgG. Binding was quantified by standard DNA curve and bars represent mean of 3 replicates normalized to promoter region binding of IgG.

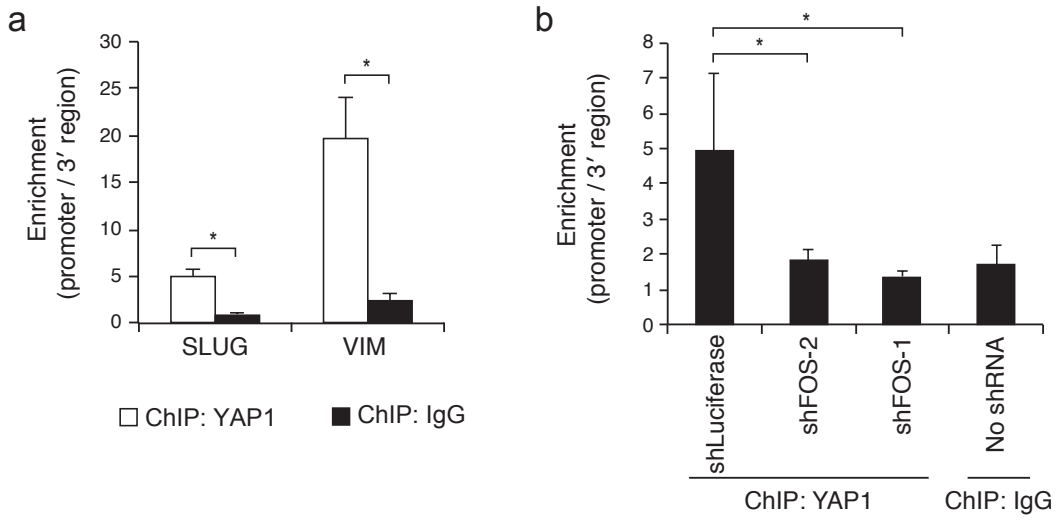

Figure 4-25. YAP1 binds with FOS at promoter regions of *SLUG* and *VIM*. (**a**) Chromatin immunoprecipitation to assess YAP1 DNA binding at target gene promoter regions of *SLUG* (*SNAI2*) and Vimentin (*VIM*). (**b**) Chromatin immunoprecipitation to assess YAP1 binding at target regions of *SLUG* after FOS suppression. (**a**, **b**) Cross-linked cell lysates from HCT116 were used. Bars represent enrichment of target region in the promoter compared to 3' region of each gene. Mean of 3 replicates are shown for each sample. *P-value <0.05.

4.5 Discussion

Using a genetic screening approach, we identified YAP1 as a gene that complemented the suppression of KRAS in KRAS-dependent cancer cells. Using a murine model of Kras-driven lung cancer, we found that tumors that arose in the setting of Kras suppression exhibited increased YAP1 activity. Through transcriptional profiling, we found that KRAS and YAP1 share regulation of many genes harboring an AP-1 transcriptional motif. Furthermore, genes in the AP-1 family, namely FOS, cooperate with YAP1 to regulate genes involved in EMT. These observations identify a *YAP1*-driven transcriptional program that recapitulates the oncogenic signals in *KRAS*-driven cancers. A summary diagram of our findings is provided in Fig. 4-26.

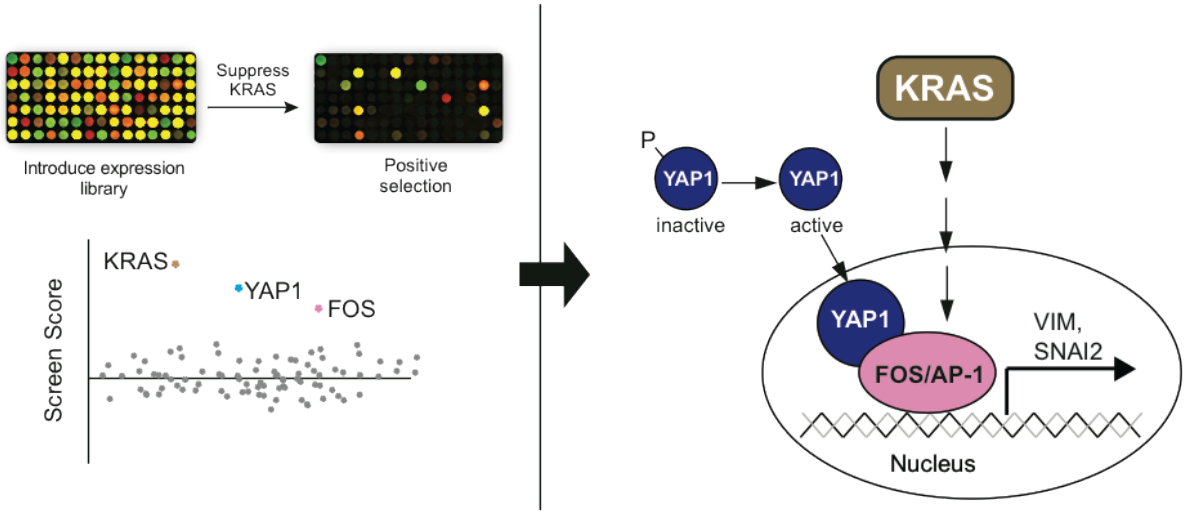

Figure 4-26. KRAS and YAP1 signaling converge at the transcriptional level through activation FOS to drive EMT.

Complementation screening in cancer cells to elucidate KRAS signaling

Using a genome-scale library of expression clones, we identified 147 candidate ORFs that rescued cell viability in the setting of KRAS suppression. We found that a substantial proportion of those genes activate known KRAS downstream MAPK and PI3K pathways. These observations confirmed previous observations that genes whose protein products activate MAPK and PI3K signaling provide a general mechanism to substitute for Ras signaling (Lim and Counter, 2005; Boehm et al., 2007). Indeed we identified several receptor tyrosine kinases that may bypass KRAS to activate downstream pathways or increase the activity of residual KRAS. YAP1 was the highest scoring gene in our screen, and we noted that WWTR1, a YAP1 homologue, also met our scoring criteria. We also identified FOS in this screen, which we showed physically binds with YAP1 at promoter regions of target genes and is necessary for YAP1-mediated cell transformation. Although we focused on YAP1 and FOS herein, these observations suggest that other genes that scored in this screen may represent novel components of KRAS-regulated signaling pathways. More generally, these observations provide proof-of-principle evidence that genome scale complementation screening by sequential genetic manipulations is feasible in mammalian systems.

KRAS and YAP1 convergence at the transcriptional level

RAS signaling has been linked to YAP1 in *Drosophila melanogaster* imaginal wing discs, in which YAP1 was required for EGFR and RAS activity (Reddy and Irvine, 2013). We showed that YAP1 is required for KRAS-driven transformation. However, we failed to find evidence for direct biochemical modulation of YAP1 activity by KRAS signaling. Although downstream effectors of KRAS such as RASSF1 and AKT have been shown to regulate YAP1 phosphorylation in specific contexts (Zhao et al., 2008a), their influence on YAP1 may be context-specific.

Recurrent amplifications of *YAP1* have been observed in liver, breast, prostate, and esophageal cancer (Zhang et al., 2011b; Overholtzer et al., 2006; Muramatsu et al., 2011; Zender et al., 2006). As an oncogene, YAP1 has been shown to drive development of hepatocellular cancer and to induce colonic adenomas in mouse models (Zender et al., 2006; Camargo et al., 2007). YAP1 has been reported to engage with different context-specific transcription factors based on complex inputs, and these signaling mechanisms are still being elucidated. In the context of KRAS suppression, we found that YAP1-regulation of AP-1 family transcription factors and specifically FOS, rather than canonical TEAD transcription factors, was required for rescue. Furthermore, RAS has been previously shown to require FOS for transformation (Saez et al., 1995). These observations provide evidence that YAP1 and KRAS converge on FOS to mediate survival in *KRAS* mutant cancer cells.

Models of resistance to RAS oncogenic addiction

Withdrawal of oncogenic *Ras* results in rapid tumor regression in mouse models (Fisher et al., 2001; Chin et al., 1999; Jechlinger et al., 2009). These genetic systems used tetracycyline-inducible overexpression of oncogenic *Ras* transgenes, and tumor relapse was often associated with mutations in the tetracycline transactivator (Podsypanina et al., 2008), resulting in reactivation of the oncogene itself. Here, we harnessed inducible *in vivo* RNAi in order to model Kras inhibition in mouse lung adenocarcinoma cells driven by $Kras^{G12D}$ expressed from its endogenous promoter. Interestingly, by monitoring long-term tumor response, we observed tumor relapse through *Kras*-independent mechanisms. This experimental model in which relapsed tumors naturally overcome Kras oncogenic addiction provided strong evidence that increased expression of YAP1 is a physiologically relevant mechanism to bypass loss of KRAS signaling in tumors harboring mutant KRAS.

We found that YAP1 replaces oncogenic KRAS signaling at least in part by regulating an EMT-like transcriptional program in mouse and human KRAS-driven cancers. RAS itself activates EMT (Wong et al., 2013; Pylayeva-Gupta et al., 2011), and strong EMT gene expression signatures are found in cell lines with decreased sensitivity to RAS suppression (Singh et al., 2009). Indeed, genes involved in EMT, such as *Slug* are essential in *KRAS* mutant cells (Wang et al., 2010). Furthermore, our assessment of the role of FOS in YAP1 function is consistent with an EMT-regulatory activity of YAP1 since FOS itself drives an EMT phenotype (Kalluri and Weinberg, 2009; Eger et al., 2000). Interestingly, sustaining the established EMT phenotype appears to be an important part of the ability of YAP1 to complement suppression of KRAS.

EMT phenotypes are upregulated in cancer cells resistant to chemotherapy (Fuchs et al., 2002; Kajiyama et al., 2007; Cheng et al., 2007), and modulation of EMT influences cellular dependency on receptor tyrosine kinases (Witta et al., 2006; Singh and Settleman, 2010). It remains unclear whether EMT reflects a refractory cell state or contributes directly to resistance. We found that YAP1-induced regulation of EMT was specific for oncogenic KRAS, since expression of YAP1 or EMT regulators such as Snail and Slug were able to rescue suppression of KRAS but failed to revert the loss of another oncogene, MYC. The specificity of this mechanism to KRAS is intriguing, as it suggests resistance associated with EMT might be attributed to bypass of specific oncogenic pathways.

CHAPTER FIVE

Discussion

5.1 Advances and limitations of gain-of-function and loss-of-function screens

By providing a foundation for interpreting RNAi gene suppression effects as quantifiable values in individual samples, we aim to expand the repertoire of possible analyses. We illustrated ways to apply gene phenotype values to systematically examine relationships amongst different gene dependencies, and to examine relationships between gene dependency and other datasets such as copy number and gene expression profiles. Extensions of the ATARiS analysis of the Project Achilles dataset (Cheung et al., 2011a) have been used by other investigators to identify cell essential genes that are heterozygously deleted in cancer. Such genes are thus more susceptible to suppression in a cancer cell harboring the heterozygous deletion compared to their normal counterparts. Genes that exhibit this property have been termed copy number alterations yielding cancer liabilities owing to partial loss (CYCLOPS) genes (Nijhawan et al., 2012). The ATARiS results have also been applied to identify essential genes in cancer cells with increased Wnt activity. By integrative analysis of gene phenotype values in relation to the activity of a reporter for the Wnt pathway as assessed all the Project Achilles cell lines, cell lines that had high reporter activity were found to require YAP1 expression. This relationship was validated through cellular and biochemical studies, and revealed a novel aspect of Wnt signaling involving beta-catenin interaction with YAP1 (Rosenbluh et al., 2012). Together, these studies support the utility of our algorithm.

We identified HNF1B by examining the correlation between each gene's expression and ATARiS gene solution in a sample-specific manner. On the other hand, when we instead examined the correlation between measurements of each individual shRNA and corresponding gene expression values, we found HNF1B shRNAs spread

throughout the ranked results: one ranked in the top 10 shRNAs, two in the top 500, and two ranked considerably lower. In the same way that using co-regulated sets of genes, rather than individual genes, can increase the signal in transcription profiling data (Subramanian et al., 2005), ATARiS uses multiple shRNAs, enriching for on target effects, to increase the signal from individual RNAi reagents. A limitation of this method includes the inability to determine gene solutions for genes that have only one or fewer reagents that perform well.

ATARiS is able to computationally distinguish the target-specific component of each RNAi reagent's effects. In contrast to standard laborious approaches to validate RNAi performance through biological assessment of gene suppression or rescue experiments with cDNA, our method provides a fast and computational method for identifying on-target reagents. Furthermore, it avoids screening biases attributed to seed sequence similarity (Marine et al., 2012; Sigoillot et al., 2012), a problem that confounds large-scale RNAi data. This bias arises due to a short 7mer sequence that affects endogenous miRNA processing or is associated with a large proportion of complementary sequences in the 3'UTR of mRNA. The likelihood that the ~5 reagents targeting each gene share a common seed sequence is low; thus, ATARiS gene solutions tend to avoid seed effects. Interestingly, the concept of seed sequence similarity forms a novel opportunity. Since some seeds match endogenous microRNA (miRNA) sequences, finding "solutions" for seeds in an analogous manner to gene solutions may yield novel insights about miRNA gene function.

Using a gain-of-function approach, we systematically probed signaling pathways involved in supporting cell viability when the RAS oncogene is suppressed. Because sequential genetic manipulations must occur (i.e. expression of shRNA to suppress RAS, and expression of ORFs), we considered ways to improve robustness of each individual step. By stable introduction of a doxycycline-inducible shRNA, and by clonal

selection for a cell line that robustly and reproducibly induced shRNA expression, we were able to effectively perform this screen at genome-scale. While we only investigate the suppression of RAS, future applications of such a screen may consider parallel investigation of other oncogenes. By comparison of ORFs that rescue RAS suppression compared to suppression of other oncogenes, we can begin to fully dissect convergent and divergent aspects of major signaling pathways in the cell.

Although one might think that genes which score highly in our gain-of-function screen might be general suppressors of cell death, rather than specific mechanisms of rescue, none of the genes that scored in the screen are general inhibitors of apoptosis. The likely explanation is that genes that simply halt cell death will not receive a high score in comparison to genes that additionally engage proliferative signaling pathways. A small gain in proliferative signaling increases measured cell viability exponentially over the course of the assay. However, as with any large-scale screening effort, we cannot confidently distinguish this concept from the possibility that genes involved in regulation of cell survival are not well represented in the screen. Larger genes are more likely to be poorly expressed, and are more likely to harbor mutations (Yang et al., 2011). Also, all genes in the library are epitope-tagged with a V5 sequence at the C-terminus, which may disrupt endogenous function of some genes. These biases may be ameliorated as expression libraries improve.

We anticipate that future libraries involving RNAi reagents will be iteratively improved, as methods such as ATARiS enable the ability to distinguish reagents with on-target performance from those without. Furthermore, as gain-of-function libraries are moved into vectors that allow for screening in pooled format, the cost and scale of gain-of-function screens will improve. The combination of gain-of-function tools and loss-of-function tools will improve our ability to systematically interrogate signaling pathways in cancer and to assess strategies to inhibit tumor growth.

5.2 Relationship between gene expression and dependency

While we showed that gene expression and gene dependency are highly correlated for HNF1B, this trend was also true for a number of other genes (Table 3-2). Examination of the distribution of all correlations suggests that for the majority of genes, high gene expression is *not* associated with increased gene dependency, but rather with decreased gene dependency. This effect is likely due to a buffering effect conferred by high gene expression levels to RNAi suppression. In light of this effect, any positive associations are particularly intriguing. However, careful validation is required since high expression level does not confirm a functional role in cancer. Even non-cancerous cells may temporarily increase gene expression. Thus, we also examined genomic amplification at the HNF1B locus to confirmation of its role in cancer development. In contrast to the buffering effect observed in highly expressed genes, genes that are expressed at low levels tend to be more sensitive to gene suppression. This phenomenon forms the basis of CYCLOPS genes as described Chapter 5.1.

Although we generated a large list of gene dependencies, validation of each gene's role in cancer is not straightforward. Some genes may gain high expression through indirect mechanisms, e.g. epigenetic regulation or increased activity of upstream activators; thus, direct genomic amplifications or mutations may not be observed. Also, functional experiments for transformation are limited by minor variations that affect results. For example, the ability of HNF1B to increase anchorage independent growth was stunted when using a V5-epitope tagged construct compared to untagged HNF1B. New strategies are necessary to functionally validate large numbers of genes. As assays to assess oncogenic function become standardized, putative oncogenes may be able to be interrogated through a combination of different assays.

5.3 Expanding understanding of RAS signaling

Screens for gain-of-function activity to replace RAS have had long-standing interest. Many screens have examined factors that replace RAS tumor initiation – from the initial small-scale screens using specific *RAS* mutations that activate specific downstream effectors (Hamad et al., 2002), to screens that systematically identify genes that replace specific components of signaling in tumor initiation (Boehm et al., 2007; Shrestha et al., 2012). Regarding genes that replace RAS after the establishment of a tumor, some differences may exist. For example, we found that an activated form of MEK, MEK-DD, which is incapable of replacing RAS in tumor initiation (Boehm et al., 2007), very strongly supported RAS signaling in established cancer cells (Fig. 4-3). Activated versions of PI3K have also been reported as sufficient to maintain RAS-driven tumors (Lim and Counter, 2005).

Similar questions arise regarding the relationship of YAP1 and RAS in tumor initiation versus maintenance. We showed that YAP1 expression alone is capable of transformation of immortalized cells, and that YAP1 rescues suppression of RAS in cancer cells; thus YAP1 appears to be sufficient to replace RAS in both initiation and maintenance of signaling. Regarding a requirement for YAP1 expression for RAS signaling, we showed that YAP1 is required for RAS-driven transformation *in vitro*. Mouse models to assess this requirement *in vivo* are in development by other investigators. However, it is unclear whether YAP1 expression is required in established RAS-driven tumors. Our ability to interpret this relationship is confounded by the fact that YAP1 dependency is associated with other factors such as *YAP1* amplification and high Wnt activity (Rosenbluh et al., 2012). Since many RAS-driven tumors also have increased Wnt activity, we have not been able to determine whether cell lines depend on YAP1 directly as a result of RAS signaling.

Our model supports a convergence of YAP1 signaling and mitogenic signaling at the transcriptional level (Fig. 4-26), however additional questions regarding the nature of the endogenous relationship of RAS and YAP1 exist. Although we did not observe direct regulation of YAP1 phosphorylation by RAS, individual components of RAS downstream signaling have been reported to regulate YAP1 activity. AKT has been shown to directly phosphorylate YAP1, leading to YAP1 sequestration in the cytoplasm and attenuation of signaling (Basu et al., 2003). The RAS effector RASSF has been shown to bind to MST1 (the mammalian homologue of Hippo) and inhibit its activity (Polesello et al., 2006), leading to increased YAP1 activity. Thus, AKT and RASSF regulate YAP1 through opposing and likely context-dependent mechanisms, and the overall effect of RAS on YAP1 activity remains elusive.

The role of FOS in YAP1 signaling merits further investigation. We showed that YAP1 and FOS physically interact, and the protein domains of YAP1 that mediate this interaction are currently under investigation. Additional AP-1 transcription factors are likely to be involved, since AP-1 family members are known to form heterodimers in order to activate transcription (Bakiri et al., 2002). YAP1 is likely a transcriptional co-activator rather than a dimerization partner since it lacks bZIP domains that mediate dimerization. Candidate genes for the dimerization partner include ATF-family transcription factors, which we showed are required for YAP1-mediated transformation (Fig. 4-18).

We were initially surprised that YAP1 specifically supported cell viability after RAS suppression but not the suppression of other oncogenes such as MYC. FOS, another gene that scored in the screen, was similarly specific for RAS rescue. Since YAP1 converges with RAS signaling at the transcriptional level, we expect that it may also rescue suppression of other oncogenes in the mitogenic signaling pathway downstream of RAS. Indeed, in screens to identify resistance mechanisms to

pharmacologic inhibitors of MEK and ERK, YAP1 expression decreased cellular sensitivity to inhibition of these pathways (personal communication, Cory Johannessen). These results independently confirm the functional ability of YAP1 to support mitogenic pathway signals.

REFERENCES

Ahearn, I. M., Haigis, K., Bar-Sagi, D., and Philips, M. R. (2011). Regulating the regulator: post-translational modification of RAS. Nature Reviews Molecular Cell Biology *13*, 39–51.

Baca, S. C., Prandi, D., Lawrence, M. S., Mosquera, J. M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., Macdonald, T. Y., Ghandi, M., et al. (2013). Punctuated evolution of prostate cancer genomes. Cell *153*, 666–677.

Baez, M. V., and Boccaccio, G. L. (2005). Mammalian Smaug is a translational repressor that forms cytoplasmic foci similar to stress granules. J. Biol. Chem. *280*, 43131–43140.

Bakiri, L., Matsuo, K., Wisniewska, M., Wagner, E. F., and Yaniv, M. (2002). Promoter specificity and biological activity of tethered AP-1 dimers. Mol. Cell. Biol. *22*, 4952–4964.

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K. K., Carter, S. L., Frederick, A. M., Lawrence, M. S., Sivachenko, A. Y., Sougnez, C., Zou, L., et al. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. Nature *486*, 405–409.

Barbacci, E., Reber, M., Ott, M. O., Breillat, C., Huetz, F., and Cereghini, S. (1999). Variant hepatocyte nuclear factor 1 is required for visceral endoderm specification. Development *126*, 4795–4805.

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., Schinzel, A. C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112.

Bard, F., Casano, L., Mallabiabarrena, A., Wallace, E., Saito, K., Kitayama, H., Guizzunti, G., Hu, Y., Wendler, F., Dasgupta, R., et al. (2006). Functional genomics reveals genes involved in protein secretion and Golgi organization. Nature *439*, 604–607.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

Basu, S., Totty, N. F., Irwin, M. S., Sudol, M., and Downward, J. (2003). Akt Phosphorylates the Yes-Associated Protein, YAP, to Induce Interaction with 14-3-3 and Attenuation of p73-Mediated Apoptosis. Mol Cell *11*, 11–23.

Bell, E. S., and Park, M. (2012). Models of crk adaptor proteins in cancer. Genes Cancer *3*, 341–352.

Bellanne-Chantelot, C., Clauin, S., Chauveau, D., Collin, P., Daumont, M., Douillard, C., Dubois-Laforgue, D., Dusselier, L., Gautier, J. F., Jadoul, M., et al. (2005). Large Genomic Rearrangements in the Hepatocyte Nuclear Factor-1 (TCF2) Gene Are the Most Frequent Cause of Maturity-Onset Diabetes of the Young Type 5. Diabetes *54*, 3126–3132.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B.

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J. C., Huang, J. H., Alexander, S., et al. (2007). Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. PNAS *104*, 20007–20012.

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J. S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.

Birmingham, A., Anderson, E. M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J., et al. (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nat. Methods *3*, 199–204.

Bluteau, O., Jeannot, E., Bioulac-Sage, P., Marqués, J. M., Blanc, J.-F., Bui, H., Beaudoin, J.-C., Franco, D., Balabaud, C., Laurent-Puig, P., et al. (2002). Bi-allelic inactivation of TCF1 in hepatic adenomas. Nature Genetics *32*, 312–315.

Boehm, J. S., Zhao, J. J., Yao, J., Kim, S. Y., Firestein, R., Dunn, I. F., Sjostrom, S. K., Garraway, L. A., Weremowicz, S., Richardson, A. L., et al. (2007). Integrative genomic approaches identify IKBKE as a breast cancer oncogene. Cell *129*, 1065–1079.

Brideau, C., Gunter, B., Pikounis, B., and Liaw, A. (2003). Improved Statistical Methods for Hit Selection in High-Throughput Screening. Journal of Biomolecular Screening *8*, 634–647.

Brough, R., Frankum, J. R., Sims, D., Mackay, A., Mendes-Pereira, A. M., Bajrami, I., Costa-Cabral, S., Rafiq, R., Ahmad, A. S., Cerone, M. A., et al. (2011). Functional Viability Profiles of Breast Cancer. Cancer Discov *1*, 260–273.

Bussey, K. J., Chin, K., Lababidi, S., Reimers, M., Reinhold, W. C., Kuo, W.-L., Gwadry, F., Ajay, Kouros-Mehr, H., Fridlyand, J., et al. (2006). Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. Mol. Cancer Ther. *5*, 853–867.

Camargo, F. D., Gokhale, S., Johnnidis, J. B., Fu, D., Bell, G. W., Jaenisch, R., and Brummelkamp, T. R. (2007). YAP1 increases organ size and expands undifferentiated progenitor cells. Curr. Biol. *17*, 2054–2060.

Cao, X., Pfaff, S. L., and Gage, F. H. (2008). YAP regulates neural progenitor cell number via the TEA domain transcription factor. Genes Dev. *22*, 3320–3334.

Cheng, G. Z., Chan, J., Wang, Q., Zhang, W., Sun, C. D., and Wang, L.-H. (2007). Twist transcriptionally up-regulates AKT2 in breast cancer cells leading to increased migration, invasion, and resistance to paclitaxel. Cancer Res. *67*, 1979–1987.

Cheung, H. W., Cowley, G. S., Weir, B. A., Boehm, J. S., Rusin, S., Scott, J. A., East, A., Ali, L. D., Lizotte, P. H., Wong, T. C., et al. (2011a). Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. Proc. Natl. Acad. Sci. U.S.A. *108*, 12372–12377.

Cheung, H. W., Du, J., Boehm, J. S., He, F., Weir, B. A., Wang, X., Butaney, M., Sequist, L. V., Luo, B., Engelman, J. A., et al. (2011b). Amplification of CRKL induces transformation and epidermal growth factor receptor inhibitor resistance in human non-small cell lung cancers. Cancer Discov *1*, 608–625.

Chien, Y., Kim, S., Bumeister, R., Loo, Y.-M., Kwon, S. W., Johnson, C. L., Balakireva, M. G., Romeo, Y., Kopelovich, L., Gale, M., et al. (2006). RalB GTPase-mediated activation of the IkappaB family kinase TBK1 couples innate immune signaling to tumor cell survival. Cell *127*, 157–170.

Chin, L., Tam, A., Pomerantz, J., Wong, M., Holash, J., Bardeesy, N., Shen, Q., O'Hagan, R., Pantginis, J., Zhou, H., et al. (1999). Essential role for oncogenic Ras in tumour maintenance. Nature *400*, 468–472.

Chung, N., Zhang, X. D., Kreamer, A., Locco, L., Kuan, P.-F., Bartz, S., Linsley, P. S., Ferrer, M., and Strulovici, B. (2008). Median absolute deviation to improve hit selection for genome-scale RNAi screens. Journal of biomolecular screening : the official journal of the Society for Biomolecular Screening *13*, 149–158.

Collinet, C., Stöter, M., Bradshaw, C. R., Samusik, N., Rink, J. C., Kenski, D., Habermann, B., Buchholz, F., Henschel, R., Mueller, M. S., et al. (2010). Systems survey of endocytosis by multiparametric image analysis. Nature *464*, 243–249.

Csiszar, I., and Tusnady, G. (1984). Information geometry and alternating minimization procedures. Statistics and Decisions *Supplement*, 205–237.

Cullen, B. R. (2006). Enhancing and confirming the specificity of RNAi experiments. Nat. Methods *3*, 677–681.

Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., Bottomley, W., et al. (2002). Mutations of the BRAF gene in human cancer. Nature *417*, 949–954.

Dean, M., Fojo, T., and Bates, S. (2005). Tumour stem cells and drug resistance. Nat. Rev. Cancer *5*, 275–284.

Dearth, R. K., Cui, X., Kim, H.-J., Kuiatse, I., Lawrence, N. A., Zhang, X., Divisova, J., Britton, O. L., Mohsin, S., Allred, D. C., et al. (2006). Mammary tumorigenesis and metastasis caused by overexpression of insulin receptor substrate 1 (IRS-1) or IRS-2. Mol. Cell. Biol. *26*, 9302–9314.

Deming, S. L., Nass, S. J., Dickson, R. B., and Trock, B. J. (2000). C-myc amplification in breast cancer: a meta-analysis of its occurrence and prognostic relevance. Br J Cancer *83*, 1688–1695.

Deng, T., and Karin, M. (1994). c-Fos transcriptional activity stimulated by H-Ras-

activated protein kinase distinct from JNK and ERK. Nature *371*, 171–175.

Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J., and Waterfield, M. D. (1984). Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. Nature *307*, 521–527.

DuPage, M., Dooley, A. L., and Jacks, T. (2009). Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. Nat Protoc *4*, 1064–1072.

Dupont, S., Morsut, L., Aragona, M., Enzo, E., Giulitti, S., Cordenonsi, M., Zanconato, F., Le Digabel, J., Forcato, M., Bicciato, S., et al. (2011). Role of YAP/TAZ in mechanotransduction. Nature *474*, 179–183.

Dupuy, A. J., Jenkins, N. A., and Copeland, N. G. (2006). Sleeping beauty: a novel cancer gene discovery tool. Human Molecular Genetics *15 Spec No 1*, R75–9.

Dyson, N. (1998). The regulation of E2F by pRB-family proteins. Genes & development.

Echeverri, C., Beachy, P., Baum, B., and Boutros, M. (2006). Minimizing the risk of reporting false positives in large-scale RNAi screens. Nat. Methods.

Eger, A., Stockinger, A., Schaffhauser, B., Beug, H., and Foisner, R. (2000). Epithelial mesenchymal transition by c-Fos estrogen receptor activation involves nuclear translocation of beta-catenin and upregulation of beta-catenin/lymphoid enhancer binding factor-1 transcriptional activity. J Cell Biol *148*, 173–188.

Ehrenreiter, K., Kern, F., Velamoor, V., Meissl, K., Galabova-Kovacs, G., Sibilia, M., and Baccarini, M. (2009). Raf-1 Addiction in Ras-Induced Skin Carcinogenesis. Cancer Cell *16*, 149–160.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494–498.

ENCODE Project Consortium, Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Engelman, J. A., Chen, L., Tan, X., Crosby, K., Guimaraes, A. R., Upadhyay, R., Maira, M., McNamara, K., Perera, S. A., Song, Y., et al. (2008). Effective use of PI3K and MEK inhibitors to treat mutant Kras G12D and PIK3CA H1047R murine lung cancers. Nature Medicine *14*, 1351–1356.

Fearon, E. R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759–767.

Filmus, J., Robles, A. I., Shi, W., Wong, M. J., Colombo, L. L., and Conti, C. J. (1994). Induction of cyclin D1 overexpression by activated ras. Oncogene *9*, 3627–3633.

Finco, T. S., Westwick, J. K., Norris, J. L., Beg, A. A., Der, C. J., and Baldwin, A. S.

(1997). Oncogenic Ha-Ras-induced signaling activates NF-kappaB transcriptional activity, which is required for cellular transformation. J. Biol. Chem. *272*, 24113–24116.

Firestein, R., Bass, A. J., Kim, S. Y., Dunn, I. F., Silver, S. J., Guney, I., Freed, E., Ligon, A. H., Vena, N., Ogino, S., et al. (2008). CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. Nature *455*, 547–551.

Fisher, G. H., Wellen, S. L., Klimstra, D., Lenczowski, J. M., Tichelaar, J. W., Lizak, M. J., Whitsett, J. A., Koretsky, A., and Varmus, H. E. (2001). Induction and apoptotic regression of lung adenocarcinomas by regulation of a K-Ras transgene in the presence and absence of tumor suppressor genes. Genes Dev. *15*, 3249–3262.

Flier, J. S., Mueckler, M. M., Usher, P., and Lodish, H. F. (1987). Elevated levels of glucose transport and transporter messenger RNA are induced by ras or src oncogenes. Science *235*, 1492–1495.

Forment, J. V., Kaidi, A., and Jackson, S. P. (2012). Chromothripsis and cancer: causes and consequences of chromosome shattering. Nature Publishing Group *12*, 663–670.

Fuchs, B. C., Fujii, T., Dorfman, J. D., Goodwin, J. M., Zhu, A. X., Lanuti, M., and Tanabe, K. K. (2008). Epithelial-to-mesenchymal transition and integrin-linked kinase mediate sensitivity to epidermal growth factor receptor inhibition in human hepatoma cells. Cancer Res. *68*, 2391–2399.

Fuchs, I. B., Lichtenegger, W., Buehler, H., Henrich, W., Stein, H., Kleine-Tebbe, A., and Schaller, G. (2002). The prognostic significance of epithelial-mesenchymal transition in breast cancer. Anticancer Res. *22*, 3415–3419.

Gaglio, D., Metallo, C. M., Gameiro, P. A., Hiller, K., Danna, L. S., Balestrieri, C., Alberghina, L., Stephanopoulos, G., and Chiaradonna, F. (2011). Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth. Mol Syst Biol *7*, 523.

Giehl, K. (2005). Oncogenic Ras in tumour progression and metastasis. Biol. Chem. *386*, 193–205.

González-García, A., Pritchard, C. A., Paterson, H. F., Mavria, G., Stamp, G., and Marshall, C. J. (2005). RalGDS is required for tumor formation in a model of skin carcinogenesis. Cancer Cell *7*, 219–226.

Gschwind, A., Fischer, O. M., and Ullrich, A. (2004). The discovery of receptor tyrosine kinases: targets for cancer therapy. Nat. Rev. Cancer *4*, 361–370.

Gupta, S., Ramjaun, A. R., Haiko, P., Wang, Y., Warne, P. H., Nicke, B., Nye, E., Stamp, G., Alitalo, K., and Downward, J. (2007). Binding of Ras to Phosphoinositide 3-Kinase p110α Is Required for Ras- Driven Tumorigenesis in Mice. Cell *129*, 957–968.

Gutman, A., Wasylyk, C., and Wasylyk, B. (1991). Cell-specific regulation of oncogene-

responsive sequences of the c-fos promoter. Mol. Cell. Biol. *11*, 5381–5387.

Hahn, W. C., Counter, C. M., Lundberg, A. S., Beijersbergen, R. L., Brooks, M. W., and Weinberg, R. A. (1999). Creation of human tumour cells with defined genetic elements : Article : Nature. Nature *400*, 464–468.

Haigis, K. M., Kendall, K. R., Wang, Y., Cheung, A., Haigis, M. C., Glickman, J. N., Niwa-Kawakita, M., Sweet-Cordero, A., Sebolt-Leopold, J., Shannon, K. M., et al. (2008). Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. Nature Genetics *40*, 600–608.

Hamad, N. M., Elconin, J. H., Karnoub, A. E., Bai, W., Rich, J. N., Abraham, R. T., Der, C. J., and Counter, C. M. (2002). Distinct requirements for Ras oncogenesis in human versus mouse cells. Genes Dev. *16*, 2045–2057.

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. Cell, 1–29.

Healy, S., Khan, P., and Davie, J. R. (2013). Immediate early response genes and cell transformation. Pharmacol. Ther. *137*, 64–77.

Helleday, T. (2011). The underlying mechanism for the PARP and BRCA synthetic lethality: clearing up the misunderstandings. Mol Oncol *5*, 387–393.

Hirsch, A. J. (2010). The use of RNAi-based screens to identify host proteins involved in viral replication. Future Microbiol *5*, 303–311.

Hirsch, F. R., Varella-Garcia, M., Bunn, P. A., Di Maria, M. V., Veve, R., Bremmes, R. M., Barón, A. E., Zeng, C., and Franklin, W. A. (2003). Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. J Clin Oncol *21*, 3798–3807.

Hofer, F., Fields, S., Schneider, C., and Martin, G. S. (1994). Activated Ras interacts with the Ral guanine nucleotide dissociation stimulator. Proc. Natl. Acad. Sci. U.S.A. *91*, 11089–11093.

Hollenhorst, P. C., Ferris, M. W., Hull, M. A., Chae, H., Kim, S., and Graves, B. J. (2011). Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. Genes Dev. *25*, 2147–2157.

Horn, T., Sandmann, T., Fischer, B., and Axelsson, E. (2011). Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. Nat. Methods.

Huang, F. W., Hodis, E., Xu, M. J., Kryukov, G. V., Chin, L., and Garraway, L. A. (2013). Highly recurrent TERT promoter mutations in human melanoma. Science *339*, 957–959.

Ihle, N. T., Lemos, R., Wipf, P., Yacoub, A., Mitchell, C., Siwak, D., Mills, G. B., Dent, P., Kirkpatrick, D. L., and Powis, G. (2009). Mutations in the phosphatidylinositol-3-kinase pathway predict for antitumor activity of the inhibitor PX-866 whereas oncogenic Ras is a dominant predictor for resistance. Cancer Res. *69*, 143–150.

Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., and Speed, T. P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res *31*.

Jackson, A. L., Bartz, S. R., Schelter, J., Kobayashi, S. V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P. S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nat. Biotechnol. *21*, 635–637.

Jackson, A. L., Burchard, J., Schelter, J., Chau, B. N., Cleary, M., Lim, L., and Linsley, P. S. (2006). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. RNA *12*, 1179–1187.

Jänne, P. A., Gray, N., and Settleman, J. (2009). Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. Nat Rev Drug Discov *8*, 709–723.

Jechlinger, M., Podsypanina, K., and Varmus, H. (2009). Regulation of transgenes in three-dimensional cultures of primary mouse mammary cells demonstrates oncogene dependence and identifies cells that survive deinduction. Genes Dev. *23*, 1677–1688.

Johannessen, C. M., Boehm, J. S., Kim, S. Y., Thomas, S. R., Wardwell, L., Johnson, L. A., Emery, C. M., Stransky, N., Cogdill, A. P., Barretina, J., et al. (2010). COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. Nature *468*, 968–972.

Kajiyama, H., Shibata, K., Terauchi, M., Yamashita, M., Ino, K., Nawa, A., and Kikkawa, F. (2007). Chemoresistance to paclitaxel induces epithelial-mesenchymal transition and enhances metastatic potential for epithelial ovarian carcinoma cells. Int. J. Oncol. *31*, 277–283.

Kalluri, R., and Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. Journal of Clinical Investigation *119*, 1420–1428.

Kamata, T., and Feramisco, J. R. (1984). Epidermal growth factor stimulates guanine nucleotide binding activity and phosphorylation of ras oncogene proteins. Nature *310*, 147–150.

Kamb, A. (2003). Consequences of nonadaptive alterations in cancer. Mol. Biol. Cell *14*, 2201–2205.

Karin, M. (1995). The regulation of AP-1 activity by mitogen-activated protein kinases. J. Biol. Chem. *270*, 16483–16486.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, A. D. (2002). The Human Genome Browser at UCSC. Genome Res. *12*, 996–1006.

Khan, O., and La Thangue, N. B. (2011). HDAC inhibitors in cancer biology: emerging mechanisms and clinical applications. Immunol Cell Biol *90*, 85–94.

Khokhlatchev, A., Rabizadeh, S., Xavier, R., Nedwidek, M., Chen, T., Zhang, X.-F., Seed, B., and Avruch, J. (2002). Identification of a novel Ras-regulated proapoptotic

pathway. Curr. Biol. *12*, 253–265.

Kielbasa, S. M., Klein, H., Roider, H. G., Vingron, M., and Bluthgen, N. (2010). TransFind--predicting transcriptional regulators for gene sets. Nucleic Acids Research *38*, W275–W280.

Kisselev, A. F., and Goldberg, A. L. (2001). Proteasome inhibitors: from research tools to drug candidates. Chemistry & Biology *8*, 739–758.

Kittler, R., and Pelletier, L. (2008). Systems biology of mammalian cell division. Cell Cycle.

Koh, J. L. Y., Brown, K. R., Moffat, J., 6 (2012). COLT-Cancer: functional genetic screening resource for essential genes in human cancer cell lines. Nucleic Acids Research *40*, D957–D963.

Kolch, W., Heidecker, G., Lloyd, P., and Rapp, U. R. (1991). Raf-1 protein kinase is required for growth of induced NIH/3T3 cells. Nature *349*, 426–428.

König, R., Chiang, C.-Y., Tu, B. P., Yan, S. F., Dejesus, P. D., Romero, A., Bergauer, T., Orth, A., Krueger, U., Zhou, Y., et al. (2007). A probability-based approach for the analysis of large-scale RNAi screens. Nat. Methods *4*, 847–849.

Lamar, J. M., Stern, P., Liu, H., Schindler, J. W., Jiang, Z.-G., and Hynes, R. O. (2012). The Hippo pathway target, YAP, promotes metastasis through its TEAD-interaction domain. Proc. Natl. Acad. Sci. U.S.A. *109*, E2441–50.

Lamesch, P., Li, N., Milstein, S., Fan, C., Hao, T., Szabo, G., Hu, Z., Venkatesan, K., Bethel, G., Martin, P., et al. (2007). hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics *89*, 307–315.

Land, H., Parada, L. F., and Weinberg, R. A. (1983). Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature *304*, 596–602.

Lau, K. S., and Haigis, K. M. (2009). Non-Redundancy within the RAS Oncogene Family: Insights into Mutational Disparities in Cancer. Molecules and Cells *28*, 315–328.

Leone, G., Sears, R., Huang, E., Rempel, R., Nuckolls, F., Park, C. H., Giangrande, P., Wu, L., Saavedra, H. I., Field, S. J., et al. (2001). Myc requires distinct E2F activities to induce S phase and apoptosis. Mol Cell *8*, 105–113.

Li, C. G., Nyman, J. E., Braithwaite, A. W., and Eccles, M. R. (2011). PAX8 promotes tumor cell growth by transcriptionally regulating E2F1 and stabilizing RB protein. Oncogene.

Li, C., and Hung Wong, W. (2001). Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol *2*, RESEARCH0032.

Lim, K.-H., and Counter, C. M. (2005). Reduction in the requirement of oncogenic Ras signaling to activation of PI3K/AKT pathway during tumor maintenance. CCELL *8*, 381–392.

Liu, J. J., Chao, J. R., Jiang, M. C., Ng, S. Y., Yen, J. J., and Yang-Yen, H. F. (1995). Ras transformation results in an elevated level of cyclin D1 and acceleration of G1 progression in NIH 3T3 cells. Mol. Cell. Biol. *15*, 3654–3663.

Liu, P., Cheng, H., Roberts, T. M., and Zhao, J. J. (2009). Targeting the phosphoinositide 3-kinase pathway in cancer. 1–18.

Lock, R., Roy, S., Kenific, C. M., Su, J. S., Salas, E., Ronen, S. M., and Debnath, J. (2011). Autophagy facilitates glycolysis during Ras-mediated oncogenic transformation. Mol. Biol. Cell *22*, 165–178.

Luo, B., Cheung, H. W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X., Hinkle, G., Boehm, J. S., Beroukhim, R., Weir, B. A., et al. (2008). Highly parallel identification of essential genes in cancer cells. Proc. Natl. Acad. Sci. U.S.A. *105*, 20380–20385.

Luo, J., Emanuele, M. J., Li, D., Creighton, C. J., Schlabach, M. R., Westbrook, T. F., Wong, K.-K., and Elledge, S. J. (2009a). A Genome-wide RNAi Screen Identifies Multiple Synthetic Lethal Interactions with the Ras Oncogene. Cell *137*, 835–848.

Luo, J., Solimini, N. L., and Elledge, S. J. (2009b). Principles of cancer therapy: oncogene and non-oncogene addiction. Cell *136*, 823–837.

Maeda, N., Fan, H., and Yoshikai, Y. (2008). Oncogenesis by retroviruses: old and new paradigms. Rev. Med. Virol. *18*, 387–405.

Malumbres, M., and Barbacid, M. (2003). RAS oncogenes: the first 30 years. Nat. Rev. Cancer *3*, 459–465.

Marcotte, R., Brown, K. R., Suarez, F., Sayad, A., Karamboulas, K., Krzyzanowski, P. M., Sircoulomb, F., Medrano, M., Fedyshyn, Y., Koh, J. L. Y., et al. (2012). Essential gene profiles in breast, pancreatic, and ovarian cancer cells. Cancer Discov *2*, 172–189.

Marine, S., Bahl, A., Ferrer, M., and Buehler, E. (2012). Common seed analysis to identify off-target effects in siRNA screens. Journal of Biomolecular Screening *17*, 370–378.

Mason, S. J., and Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. Q. J. R. Meteorol. Soc. *128*, 2145–2166.

Mayo, M. W., and Baldwin, A. S. (2000). The transcription factor NF-kappaB: control of oncogenesis and cancer therapy resistance. Biochim. Biophys. Acta *1470*, M55–62.

Meade-Tollin, L. C., Boukamp, P., Fusenig, N. E., Bowen, C. P., Tsang, T. C., and Bowden, G. T. (1998). Differential expression of matrix metalloproteinases in

activated c-ras-Ha-transfected immortalized human keratinocytes. Br J Cancer *77*, 724–730.

Mechta, F., Lallemand, D., Pfarr, C. M., and Yaniv, M. (1997). Transformation by ras modifies AP1 composition and activity. Oncogene *14*, 837–847.

Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol *12*, R41.

Moffat, J., Grueneberg, D. A., Yang, X., Kim, S. Y., Kloepfer, A. M., Foo, S. Y., Stewart, S. A., Stockwell, B. R., Hacohen, N., Hahn, W. C., et al. (2006). A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. Cell *124*, 1283–1298.

Moodie, S. A., Willumsen, B. M., Weber, M. J., and Wolfman, A. (1993). Complexes of Ras.GTP with Raf-1 and mitogen-activated protein kinase kinase. Science *260*, 1658–1661.

Muramatsu, T., Imoto, I., Matsui, T., Kozaki, K.-I., Haruki, S., Sudol, M., Shimada, Y., Tsuda, H., Kawano, T., and Inazawa, J. (2011). YAP is a candidate oncogene for esophageal squamous cell carcinoma. Carcinogenesis *32*, 389–398.

Musgrove, E. A., Caldon, C. E., Barraclough, J., Stone, A., and Sutherland, R. L. (2011). Cyclin D as a therapeutic target in cancer. Nat. Rev. Cancer *11*, 558–572.

Müller, P., Kuttenkeuler, D., Gesellchen, V., Zeidler, M. P., and Boutros, M. (2005). Identification of JAK/STAT signalling components by genome-wide RNA interference. Nature *436*, 871–875.

Nambiar, M., Kari, V., and Raghavan, S. C. (2008). Chromosomal translocations in cancer. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer *1786*, 139–152.

Nijhawan, D., Zack, T. I., Ren, Y., Strickland, M. R., Lamothe, R., Schumacher, S. E., Tsherniak, A., Besche, H. C., Rosenbluh, J., Shehata, S., et al. (2012). Cancer vulnerabilities unveiled by genomic loss. Cell *150*, 842–854.

Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W. C., Waltham, M., Kouros-Mehr, H., Bussey, K. J., Lee, J. K., Espina, V., et al. (2003). Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. Proc. Natl. Acad. Sci. U.S.A. *100*, 14229–14234.

Oeggerli, M., Schraml, P., Ruiz, C., Bloch, M., Novotny, H., Mirlacher, M., Sauter, G., and Simon, R. (2006). E2F3 is the main target gene of the 6p22 amplicon with high specificity for human bladder cancer. Oncogene *25*, 6538–6543.

Oka, T., and Sudol, M. (2009). Nuclear localization and pro-apoptotic signaling of YAP2 require intact PDZ-binding motif. Genes Cells *14*, 607–615.

Ota, M., and Sasaki, H. (2008). Mammalian Tead proteins regulate cell proliferation and contact inhibition as transcriptional mediators of Hippo signaling. Development *135*,

4059–4069.

Overholtzer, M., Zhang, J., Smolen, G. A., Muir, B., Li, W., Sgroi, D. C., Deng, C.-X., Brugge, J. S., and Haber, D. A. (2006). Transforming properties of YAP, a candidate oncogene on the chromosome 11q22 amplicon. Proc. Natl. Acad. Sci. U.S.A. *103*, 12405–12410.

Paez, J. G., Jänne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F. J., Lindeman, N., Boggon, T. J., et al. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science *304*, 1497–1500.

Pan, D. (2010). The Hippo Signaling Pathway in Development and Cancer. Developmental Cell *19*, 491–505.

Parada, L. F., Tabin, C. J., Shih, C., and Weinberg, R. A. (1982). Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus ras gene. Nature *297*, 474–478.

Paweletz, C. P., Charboneau, L., Bichsel, V. E., Simone, N. L., Chen, T., Gillespie, J. W., Emmert-Buck, M. R., Roth, M. J., Petricoin, E. F., III, and Liotta, L. A. (2001). Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene *20*, 1981–1989.

Peschard, P., and Park, M. (2003). Escape from Cbl-mediated downregulation: a recurrent theme for oncogenic deregulation of receptor tyrosine kinases. CCELL *3*, 519–523.

Podsypanina, K., Politi, K., Beverly, L. J., and Varmus, H. E. (2008). Oncogene cooperation in tumor maintenance and tumor recurrence in mouse mammary tumors induced by Myc and mutant Kras. Proc. Natl. Acad. Sci. U.S.A. *105*, 5242–5247.

Polesello, C., Huelsmann, S., Brown, N. H., and Tapon, N. (2006). The Drosophila RASSF homolog antagonizes the hippo pathway. Curr. Biol. *16*, 2459–2465.

Potenza, N., Vecchione, C., Notte, A., De Rienzo, A., Rosica, A., Bauer, L., Affuso, A., De Felice, M., Russo, T., Poulet, R., et al. (2005). Replacement of K-Ras with H-Ras supports normal embryonic development despite inducing cardiovascular pathology in adult mice. EMBO Rep. *6*, 432–437.

Prindle, M. J., Fox, E. J., and Loeb, L. A. (2010). The mutator phenotype in cancer: molecular mechanisms and targeting strategies. Curr Drug Targets *11*, 1296–1303.

Pylayeva-Gupta, Y., Grabocka, E., and Bar-Sagi, D. (2011). RAS oncogenes: weaving a tumorigenic web. Nature Publishing Group *11*, 761–774.

Ramsay, R. (2008). MYB function in normal and cancer cells. Nat. Rev. Cancer.

Reddy, B. V. V. G., and Irvine, K. D. (2013). Regulation of Hippo Signaling by EGFR-MAPK Signaling through Ajuba Family Proteins. Developmental Cell *24*, 459–471.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J. P. (2006). GenePattern 2.0. Nat Genet *38*, 500–501.

Reichmann, E., Schwarz, H., Deiner, E. M., Leitner, I., Eilers, M., Berger, J., Busslinger, M., and Beug, H. (1992). Activation of an inducible c-FosER fusion protein causes loss of epithelial polarity and triggers epithelial-fibroblastoid cell conversion. Cell *71*, 1103–1116.

Richter, A. M., Pfeifer, G. P., and Dammann, R. H. (2009). The RASSF proteins in cancer; from epigenetic silencing to functional characterization. Biochim. Biophys. Acta *1796*, 114–128.

Rodriguez-Viciana, P., Warne, P. H., Dhand, R., Vanhaesebroeck, B., Gout, I., Fry, M. J., Waterfield, M. D., and Downward, J. (1994). Phosphatidylinositol-3-OH kinase as a direct target of Ras. Nature *370*, 527–532.

Root, D. E., Hacohen, N., Hahn, W. C., Lander, E. S., and Sabatini, D. M. (2006). Genome-scale loss-of-function screening with a lentiviral RNAi library. Nat. Methods *3*, 715–719.

Rosenbluh, J., Nijhawan, D., Cox, A. G., Li, X., Neal, J. T., Schafer, E. J., Zack, T. I., Wang, X., Tsherniak, A., Schinzel, A. C., et al. (2012). β-Catenin-driven cancers require a YAP1 transcriptional complex for survival and tumorigenesis. Cell *151*, 1457–1473.

Ryffel, G. U. (2001). Mutations in the human genes encoding the transcription factors of the hepatocyte nuclear factor (HNF)1 and HNF4 families: functional and pathological consequences. J. Mol. Endocrinol. *27*, 11–29.

Saez, E., Rutberg, S. E., Mueller, E., Oppenheim, H., Smoluk, J., Yuspa, S. H., and Spiegelman, B. M. (1995). c-fos is required for malignant progression of skin tumors. Cell *82*, 721–732.

Schmidt, C. R., Gi, Y. J., Patel, T. A., Coffey, R. J., Beauchamp, R. D., and Pearson, A. S. (2005). E-cadherin is regulated by the transcriptional repressor SLUG during Ras-mediated transformation of intestinal epithelial cells. Surgery *138*, 306–312.

Scholl, C., Fröhling, S., Dunn, I. F., Schinzel, A. C., Barbie, D. A., Kim, S. Y., Silver, S. J., Tamayo, P., Wadlow, R. C., Ramaswamy, S., et al. (2009). Synthetic Lethal Interaction between Oncogenic KRAS Dependency and STK33 Suppression in Human Cancer Cells. Cell *137*, 821–834.

Schumacher, F. R., Berndt, S. I., Siddiq, A., Jacobs, K. B., Wang, Z., Lindstrom, S., Stevens, V. L., Chen, C., Mondul, A. M., Travis, R. C., et al. (2011). Genome-wide association study identifies new prostate cancer susceptibility loci. Human Molecular Genetics *20*, 3867–3875.

Shao, D. D., Tsherniak, A., Gopal, S., Weir, B. A., Tamayo, P., Stransky, N., Schumacher, S. E., Zack, T. I., Beroukhim, R., Garraway, L. A., et al. (2013). ATARiS: Computational quantification of gene suppression phenotypes from multisample RNAi screens. Genome Res.

Sharma, S. V., and Settleman, J. (2007). Oncogene addiction: setting the stage for molecularly targeted cancer therapy. Genes Dev. *21*, 3214–3231.

Sharma, S. V., Gajowniczek, P., Way, I. P., Lee, D. Y., Jiang, J., Yuza, Y., Classon, M., Haber, D. A., and Settleman, J. (2006). A common signaling cascade may underlie "addiction" to the Src, BCR-ABL, and EGF receptor oncogenes. CCELL *10*, 425–435.

Shen, H., Fridley, B. L., Song, H., Lawrenson, K., Cunningham, J. M., Ramus, S. J., Cicek, M. S., Tyrer, J., Stram, D., Larson, M. C., et al. (2013). Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. Nat Commun *4*, 1628.

Sherbenou, D. W., and Druker, B. J. (2007). Applying the discovery of the Philadelphia chromosome. J. Clin. Invest. *117*, 2067–2074.

Shih, C., Padhy, L. C., Murray, M., and Weinberg, R. A. (1981). Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. Nature *290*, 261–264.

Shrestha, Y., Schafer, E. J., Boehm, J. S., Thomas, S. R., He, F., Du, J., Wang, S., Barretina, J., Weir, B. A., Zhao, J. J., et al. (2012). PAK1 is a breast cancer oncogene that coordinately activates MAPK and MET signaling. Oncogene *31*, 3397–3408.

Sigoillot, F. D., Lyman, S., Huckins, J. F., and Adamson, B. (2012). A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. Nat. Methods.

Silva, J. M., Marran, K., Parker, J. S., Silva, J., Golding, M., Schlabach, M. R., Elledge, S. J., Hannon, G. J., and Chang, K. (2008). Profiling essential genes in human mammary cells by multiplex RNAi screening. Science *319*, 617–620.

Singh, A., and Settleman, J. (2010). EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. Oncogene *29*, 4741–4751.

Singh, A., Greninger, P., Rhodes, D., Koopman, L., Violette, S., Bardeesy, N., and Settleman, J. (2009). A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival. Cancer Cell *15*, 489–500.

Smith, M. R., DeGudicibus, S. J., and Stacey, D. W. (1986). Requirement for c-ras proteins during viral oncogene transformation. Nature *320*, 540–543.

Solit, D. B., Garraway, L. A., Pratilas, C. A., Sawai, A., Getz, G., Basso, A., Ye, Q., Lobo, J. M., She, Y., Osman, I., et al. (2006). BRAF mutation predicts sensitivity to MEK inhibition. Nature *439*, 358–362.

Soule, H. D., Maloney, T. M., Wolman, S. R., Peterson, W. D., Brenz, R., McGrath, C. M., Russo, J., Pauley, R. J., Jones, R. F., and Brooks, S. C. (1990). Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Res. *50*, 6075–6086.

Spurdle, A. B., Thompson, D. J., Ahmed, S., Ferguson, K., Healey, C. S., O apos Mara, T., Walker, L. C., Montgomery, S. B., Dermitzakis, E. T., Fahey, P., et al. (2011). Genome-wide association study identifies a common variant associated with risk of endometrial cancer. Nat Genet *43*, 451–454.

Stehelin, D., Fujita, D. J., Padgett, T., Varmus, H. E., and Bishop, J. M. (1977). Detection and enumeration of transformation-defective strains of avian sarcoma virus with molecular hybridization. Virology *76*, 675–684.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., and Williams, R. M. J. (1949). The American Soldier, Vol. 1: Adjustment during Army Life 1st ed. (Princeton U. Press).

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. *102*, 15545–15550.

Sudol, M. (2012). YAP1 oncogene and its eight isoforms. Oncogene.

Sudol, M. (1994). Yes-associated protein (YAP65) is a proline-rich phosphoprotein that binds to the SH3 domain of the Yes proto-oncogene product. Oncogene *9*, 2145–2152.

Suske, G. (1999). The Sp-family of transcription factors. Gene *238*, 291–300.

Taube, J. H., Herschkowitz, J. I., Komurov, K., Zhou, A. Y., Gupta, S., Yang, J., Hartwell, K., Onder, T. T., Gupta, P. B., Evans, K. W., et al. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. PNAS *107*, 15449–15454.

Terasawa, K., Toyota, M., Sagae, S., Ogi, K., Suzuki, H., Sonoda, T., Akino, K., Maruyama, R., Nishikawa, N., Imai, K., et al. (2006). Epigenetic inactivation of TCF2 in ovarian cancer and various cancer cell lines. Br J Cancer *94*, 914–921.

To, M. D., Wong, C. E., Karnezis, A. N., Del Rosario, R., Di Lauro, R., and Balmain, A. (2008). Kras regulatory elements and exon 4A determine mutation specificity in lung cancer. Nature Genetics *40*, 1240–1244.

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tsuchiya, A., Sakamoto, M., Yasuda, J., Chuma, M., Ohta, T., Ohki, M., Yasugi, T., Taketani, Y., and Hirohashi, S. (2003). Expression profiling in ovarian clear cell carcinoma: identification of hepatocyte nuclear factor-1β as a molecular marker and a possible molecular target for therapy of ovarian clear cell carcinoma. The

American journal of pathology *163*, 2503–2512.

Tuveson, D. A., Shaw, A. T., Willis, N. A., Silver, D. P., Jackson, E. L., Chang, S., Mercer, K. L., Grochow, R., Hock, H., Crowley, D., et al. (2004). Endogenous oncogenic K-ras(G12D) stimulates proliferation and widespread neoplastic and developmental defects. CCELL *5*, 375–387.

Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Libermann, T. A., and Schlessinger, J. (1984). Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature *309*, 418–425.

Vassilev, A., Kaneko, K. J., Shu, H., Zhao, Y., and DePamphilis, M. L. (2001). TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. Genes Dev. *15*, 1229–1241.

Verdeguer, F., Le Corre, S., Fischer, E., Callens, C., Garbay, S., Doyen, A., Igarashi, P., Terzi, F., and Pontoglio, M. (2009). A mitotic transcriptional switch in polycystic kidney disease. Nat Med *16*, 106–110.

Vicent, S., Chen, R., Sayles, L. C., Lin, C., Walker, R. G., Gillespie, A. K., Subramanian, A., Hinkle, G., Yang, X., Saif, S., et al. (2010). Wilms tumor 1 (WT1) regulates KRAS-driven oncogenesis and senescence in mouse and human models. J. Clin. Invest. *120*, 3940–3952.

Wang, K., Degerny, C., Xu, M., and Yang, X.-J. (2009). YAP, TAZ, and Yorkie: a conserved family of signal-responsive transcriptional coregulators in animal development and human disease. Biochem. Cell Biol. *87*, 77–91.

Wang, Y., Ngo, V. N., Marani, M., Yang, Y., Wright, G., Staudt, L. M., and Downward, J. (2010). Critical role for transcriptional repressor Snail2 in transformation by oncogenic RAS in colorectal carcinoma cells. Oncogene *29*, 4658–4670.

Warne, P. H., Viciana, P. R., and Downward, J. (1993). Direct interaction of Ras and the amino-terminal region of Raf-1 in vitro. Nature *364*, 352–355.

Weinstein, B., and Joe, A. (2008). Oncogene Addiction. Cancer Res. *69*, 1–4.

Weinstein, I. B. (2000). Disorders in cell circuitry during multistage carcinogenesis: the role of homeostasis. Carcinogenesis *21*, 857–864.

Wheeler, D. L., Dunn, E. F., and Harari, P. M. (2010). Understanding resistance to EGFR inhibitors—impact on future treatment strategies. Nat Rev Clin Oncol *7*, 493–507.

Whitlock, M. C. (2005). Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. Journal of Evolutionary Biology *18*, 1368–1373.

Witta, S. E., Gemmill, R. M., Hirsch, F. R., Coldren, C. D., Hedman, K., Ravdel, L., Helfrich, B., Dziadziuszko, R., Chan, D. C., Sugita, M., et al. (2006). Restoring E-

cadherin expression increases sensitivity to epidermal growth factor receptor inhibitors in lung cancer cell lines. Cancer Res. *66*, 944–950.

Wong, C. E., Yu, J. S., Quigley, D. A., To, M. D., Jen, K. Y., Huang, P. Y., Del Rosario, R., and Balmain, A. (2013). Inflammation and Hras signaling control epithelial-mesenchymal transition during skin tumor progression. Genes Dev. *27*, 670–682.

Xue, W., Meylan, E., Oliver, T. G., Feldser, D. M., Winslow, M. M., Bronson, R., and Jacks, T. (2011). Response and Resistance to NF-kappaB Inhibitors in Mouse Models of Lung Adenocarcinoma. Cancer Discov *1*, 236–247.

Yang, X., Boehm, J. S., Yang, X., Salehi-Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas, S. R., Alkan, O., Bhimdi, T., et al. (2011). A public genome-scale lentiviral expression library of human ORFs. Nat. Methods *8*, 659–661.

Yu, F.-X., Zhao, B., Panupinthu, N., Jewell, J. L., Lian, I., Wang, L. H., Zhao, J., Yuan, H., Tumaneng, K., Li, H., et al. (2012). Regulation of the Hippo-YAP Pathway by G-Protein-Coupled Receptor Signaling. Cell *150*, 780–791.

Yu, Q., Geng, Y., and Sicinski, P. (2001). Specific protection against breast cancers by cyclin D1 ablation. Nature *411*, 1017–1021.

Zender, L., Spector, M. S., Xue, W., Flemming, P., Cordon-Cardo, C., Silke, J., Fan, S.-T., Luk, J. M., Wigler, M., Hannon, G. J., et al. (2006). Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. Cell *125*, 1253–1267.

Zhang, X. D., Ferrer, M., Espeseth, A. S., Marine, S. D., Stec, E. M., Crackower, M. A., Holder, D. J., Heyse, J. F., and Strulovici, B. (2007). The use of strictly standardized mean difference for hit selection in primary RNA interference high-throughput screening experiments. Journal of biomolecular screening : the official journal of the Society for Biomolecular Screening *12*, 497–509.

Zhang, X. D., Santini, F., Lacson, R., Marine, S. D., Wu, Q., Benetti, L., Yang, R., McCampbell, A., Berger, J. P., Toolan, D. M., et al. (2011a). cSSMD: Assessing Collective Activity for Addressing Off-Target Effects in Genome-Scale RNA Interference Screens. Bioinformatics.

Zhang, X., George, J., Deb, S., Degoutin, J. L., Takano, E. A., Fox, S. B., Bowtell, D. D. L., and Harvey, K. F. (2011b). The Hippo pathway transcriptional co-activator, YAP, is an ovarian cancer oncogene. Oncogene *30*, 2810–2822.

Zhang, X., Grusche, F. A., and Harvey, K. F. (2012). Control of Tissue Growth and Cell Transformation by the Salvador/Warts/Hippo Pathway. PLoS ONE *7*, e31994.

Zhao, B., Lei, Q.-Y., and Guan, K.-L. (2008a). The Hippo-YAP pathway: new connections between regulation of organ size and cancer. Current Opinion in Cell Biology *20*, 638.

Zhao, B., Wei, X., Li, W., Udan, R. S., Yang, Q., Kim, J., Xie, J., Ikenoue, T., Yu, J., Li, L., et al. (2007). Inactivation of YAP oncoprotein by the Hippo pathway is involved in

cell contact inhibition and tissue growth control. Genes Dev. *21*, 2747–2761.

Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., Yu, J., Lin, J. D., Wang, C.-Y., Chinnaiyan, A. M., et al. (2008b). TEAD mediates YAP-dependent gene induction and growth control. Genes Dev. *22*, 1962–1971.

Zuber, J., McJunkin, K., Fellmann, C., Dow, L. E., Taylor, M. J., Hannon, G. J., and Lowe, S. W. (2011). Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. Nat. Biotechnol. *29*, 79–83.

APPENDIX ONE

ATARiS modeling algorithm

Statistical Modeling

Given the measurements of phenotypic effects produced by a set of RNAi reagents designed to target the same gene *G*, ATARiS generates a consensus profile that represents the effect of suppressing *G* in each screened sample relative to the other samples. Let *n* denote the number of screened samples and *p* denote the number of reagents targeting *G* for which the measurements are given. Let **X** denote a $p \times n$ matrix with each element $x_{i,j}$ representing the observed phenotypic effect produced by reagent *i* in sample *j*. As we are only interested in finding the *relative* effects of gene suppression, we median-center each row of **X** to obtain $\mathbf{X}^* = \mathbf{X} - \mu\mathbf{1}_n^T$ where μ is a vector of length *p* such that $\mu_i = \text{median}(x_{i,*})$ and $\mathbf{1}_n$ is a vector of 1's of length *n*.

Let **c** denote a vector of length *n* representing the consensus profile for $\mathbf{X}^*$ and let **e** denote a vector of length *p* consisting of a relative effect size for each RNAi reagent. ATARiS models each measurement $x^*_{i,j}$ as a product of its corresponding (unknown) relative effect size e_i and phenotypic effect c_j, such that an approximation for $\mathbf{X}^*$ is given by $\hat{\mathbf{X}}^* = \mathbf{e}\mathbf{c}^T$, and we set $\max(\mathbf{e}) = 1$ for identifiability. We can then formulate the problem of finding the values for **e** and **c** as the following optimization problem:

$$\text{minimize}_{\mathbf{e},\mathbf{c}} \left\| \mathbf{X}^* - \mathbf{e}\mathbf{c}^T \right\|_1 \text{ subject to } \max(\mathbf{e}) = 1, \text{ where } \|A\|_1 = \sum_i \sum_j |a_{i,j}| .$$

This criterion, which can also be seen as a rank-1 matrix factorization problem, although not convex, is bilinear in **c** and **e** (i.e., with **c** fixed, it is linear in **e** and vice versa). To optimize it we use the following iterative algorithm of alternating minimizations(Csiszar and Tusnady, 1984):

We begin by initializing ***c*** with the mean values of $\mathbf{X}^*$ in each sample:

$$c_j \leftarrow \frac{1}{p}\sum_i x^*_{i,j} \text{ for } j \in 1, \ldots, n .$$

We then update **e** and **c** repeatedly until convergence:

$$\mathbf{e} \leftarrow \arg\min_{\mathbf{e}} \left\| \mathbf{X}^* - \mathbf{e}\mathbf{c}^T \right\|_1$$

$$\mathbf{c} \leftarrow \arg\min_{\mathbf{c}} \left\| \mathbf{X}^* - \mathbf{e}\mathbf{c}^T \right\|_1$$

The elements of **e** and **c** are updated in an element-wise manner, i.e.,

$$e_i \leftarrow \arg\min_{\hat{e}} \sum_j \left| x^*_{i,j} - \hat{e}c_j \right| \text{ for } i \in 1, \ldots, p$$

and similarly

$$c_j \leftarrow \arg\min_{\hat{c}} \sum_i \left| x^*_{i,j} - e_i\hat{c} \right| \text{ for } j \in 1, \ldots, n .$$

Each such assignment can be viewed as a problem of finding a weighted median, which can be solved efficiently.

We cease iterating when a decrease of less than 1% in $\left\| \mathbf{X}^* - \mathbf{e}\mathbf{c}^T \right\|_1$ is observed. For the Achilles and Marcotte et al. datasets, we found that convergence almost always occurs after fewer than 20 iterations. As this optimization problem is not convex, we are not guaranteed to find a global minimum. To test the performance of the optimization in practice, we ran it multiple times with random initialization values and found that the variations in the parameters estimated are minimal. Finally, to identify the solution we set $\mathbf{e} \leftarrow \frac{1}{\max(\mathbf{e})} \cdot \mathbf{e}$ and $\mathbf{c} \leftarrow \max(\mathbf{e}) \cdot \mathbf{c}$.

Refinement of RNAi reagent subset

For each gene, ATARiS tries to identify subsets of its RNAi reagents that produce similar effects across the screened samples. Given a set of reagents R_G targeting gene

G, we iteratively refine R_G until we identify a subset $\hat{R}_G \subseteq R_G$ that consists of reagents whose profiles (i.e., effects across the samples) are all similar to the consensus profile computed for $\hat{R}_G$. We then consider $\hat{R}_G$ to be a consistent set and use its consensus profile as a gene solution, as described in the main text. We begin by computing a consensus profile for the reagent set $R_G^* = R_G$. We then evaluate the following criteria to determine whether R_G^* is a consistent set of reagents:

1. For each reagent $r \in R_G^*$, the Spearman correlation coefficient ρ_r between the reagent profile $(x_{r,1}, x_{r,2}, \ldots, x_{r,n})$ and the consensus profile **c** must be greater than the 85th percentile of the corresponding Spearman correlation coefficients similarly generated from data of random reagent sets of size $\left|R_G^*\right|$. (Note: The threshold can be adjusted depending on the user's desired confidence and properties of the data. See also Supplementary Fig. 2 for an analysis of the influence of this threshold on the number of solutions found.)
2. All the reagents in R_G^* must have a relative effect size e_r of at least 0.3, i.e., $e_r \geq 0.3, \forall r \in R_G^*$. We therefore favor reagents whose effects have comparable magnitudes, avoiding the inclusion of reagents whose effects are mainly due to noise (assuming that noise magnitudes are similar across reagents).

If either criterion is not fulfilled we remove one reagent from the set R_G^* as follows:

1. If any reagent $r \in R_G^*$ does not satisfy criterion (2), we discard the one with the lowest effect magnitude e_r.
2. Otherwise, we discard the reagent $r \in R_G^*$ with the lowest Spearman correlation coefficient between its profile and the consensus profile **c**.

The refinement process is repeated until R_G^* is consistent or until it consists of only one reagent, in which case we conclude that there is no solution to the set R_G of reagents. Our refinement algorithm is greedy so that it is scalable and can be used to analyze RNAi screens performed using reagent libraries that have a large number of reagents per gene.

Assignment of consistency scores

We determine a consistency score for each RNAi reagent of a given gene *G* based on its similarity to other reagents targeting *G*. For a reagent *r* that is part of a consistent set $\hat{R}_G$ (and hence was used to generated a gene solution), we determine the similarity of its profile $(x_{r,1}, x_{r,2}, \ldots, x_{r,n})$ to the profiles of all other reagents in $\hat{R}_G$ by computing the corresponding Spearman correlation coefficients. We estimate a *p*-value for each correlation coefficient based on an empirical null distribution of Spearman correlation coefficients of random pairs of reagent profiles. We combine the *p*-values associated with *r* into a single significance estimate, $p\text{-value}_r^*$, using Stouffer's method(Stouffer et al., 1949; Whitlock, 2005). The consistency score of *r* is defined as $-log_{10}(p\text{-value}_r^*)$.

For reagents that were not used to generate a gene solution we proceed as above to estimate a consistency score and a *p*-value but use the set of *all* reagents that target gene *G*.

APPENDIX TWO

Open reading frames that rescue

KRAS suppression

Gene Symbol	Clone ID	NCBI Gene ID	Adjusted Score
YAP1	ccsbBroad304_07601	10413	16.73
FGF3	ccsbBroad304_00556	2248	10.44
SAMD11	ccsbBroad304_14402	148398	9.67
FGF6	ccsbBroad304_06207	2251	9.60
SP8	ccsbBroad304_14442	221833	8.76
FGR	ccsbBroad304_00564	2268	8.57
VPS52	ccsbBroad304_04101	6293	8.34
GPR139	ccsbBroad304_04780	124274	7.67
FGF10	ccsbBroad304_00557	2255	7.64
PNO1	ccsbBroad304_03752	56902	7.53
NR4A1	ccsbBroad304_00760	3164	7.44
SAMD4B	ccsbBroad304_03521	55095	7.36
SAMD4A	ccsbBroad304_07831	23034	7.33
LPAR1	ccsbBroad304_00477	1902	7.33
WWTR1	ccsbBroad304_02889	25937	7.09
FEZF2	ccsbBroad304_15088	NA	7.06
FOXJ1	ccsbBroad304_00575	2302	7.01
FAM120A	ccsbBroad304_07851	23196	6.40
TNFAIP1	ccsbBroad304_01685	7126	6.40
OSM	ccsbBroad304_01126	5008	6.10
HGF	ccsbBroad304_06362	3082	6.08
F2RL1	ccsbBroad304_06187	2150	5.96
PLEKHG5	ccsbBroad304_12347	57449	5.92
FRS3	ccsbBroad304_07686	10817	5.89
RELB	ccsbBroad304_01388	5971	5.81
SNAPC3	ccsbBroad304_01562	6619	5.80
PURG	ccsbBroad304_03114	29942	5.78
APOBEC3D	ccsbBroad304_09588	140564	5.65
LHX9	ccsbBroad304_03761	56956	5.62
VPS28	ccsbBroad304_03232	51160	5.56
HEY2	ccsbBroad304_02781	23493	5.25
RBM47	ccsbBroad304_08387	54502	5.16
MAPK8IP2	ccsbBroad304_11744	23542	5.15
DBP	ccsbBroad304_06084	1628	5.14
CXorf67	ccsbBroad304_10617	340602	5.08
BATF3	ccsbBroad304_03599	55509	5.08
KCTD17	ccsbBroad304_12606	79734	5.01
TFEB	ccsbBroad304_01847	7942	4.90
FOXP2	ccsbBroad304_13001	93986	4.88
GLOD4	ccsbBroad304_08199	51031	4.88

HOXD4	ccsbBroad304_00779	3233	4.79
ROD1	ccsbBroad304_11442	9991	4.77
ZNF444	ccsbBroad304_14197	55311	4.76
LPAR1	ccsbBroad304_10795	1902	4.72
POU2AF1	ccsbBroad304_01247	5450	4.69
C1orf63	ccsbBroad304_08688	57035	4.68
POLR3K	ccsbBroad304_08340	51728	4.62
STK35	ccsbBroad304_13210	140901	4.62
RGL3	ccsbBroad304_08706	57139	4.57
MLYCD	ccsbBroad304_11734	23417	4.56
HTR2C	ccsbBroad304_00805	3358	4.55
IL4	ccsbBroad304_00849	3565	4.55
POU2F2	ccsbBroad304_11047	5452	4.52
LMNA	ccsbBroad304_00945	4000	4.42
IGF2BP2	ccsbBroad304_10208	10644	4.41
POU5F1	ccsbBroad304_06753	5460	4.40
PLA2G5	ccsbBroad304_01214	5322	4.40
C6orf154	ccsbBroad304_05257	221424	4.36
NACC1	ccsbBroad304_04636	112939	4.33
NOG	ccsbBroad304_02118	9241	4.31
CARD9	ccsbBroad304_03933	64170	4.30
GOLGA6L9	ccsbBroad304_13690	440295	4.30
PURB	ccsbBroad304_06823	5814	4.25
SIX4	ccsbBroad304_15076	51804	4.23
DPP3	ccsbBroad304_07545	10072	4.22
NR0B1	ccsbBroad304_00041	190	4.21
TAF5L	ccsbBroad304_08053	27097	4.16
CELF6	ccsbBroad304_15129	NA	4.13
IRF5	ccsbBroad304_00881	3663	4.05
KLHL3	ccsbBroad304_11822	26249	4.04
PLEKHG6	ccsbBroad304_03548	55200	4.01
PIM1	ccsbBroad304_01203	5292	3.99
TMEM44	ccsbBroad304_16069	93109	3.99
SIX2	ccsbBroad304_07671	10736	3.98
RASGRP3	ccsbBroad304_07939	25780	3.95
YBX2	ccsbBroad304_03200	51087	3.92
CCDC9	ccsbBroad304_02914	26093	3.90
THRB	ccsbBroad304_01669	7068	3.88
CRB3	ccsbBroad304_04577	92359	3.87
GATAD2A	ccsbBroad304_12088	54815	3.87
DNAJC17	ccsbBroad304_03546	55192	3.84
TGM5	ccsbBroad304_02141	9333	3.83
BRAF	ccsbBroad304_00174	673	3.82

TRIM15	ccsbBroad304_09278	89870	3.82
CYP46A1	ccsbBroad304_07694	10858	3.81
FOS	ccsbBroad304_06220	2353	3.75
NR4A2	ccsbBroad304_06662	4929	3.74
SOCS6	ccsbBroad304_02132	9306	3.73
KLHL34	ccsbBroad304_09923	257240	3.72
RHOBTB1	ccsbBroad304_07503	9886	3.71
PCDHB14	ccsbBroad304_08623	56122	3.71
CYP1B1	ccsbBroad304_13839	1545	3.68
HNRNPC	ccsbBroad304_06387	3183	3.68
FAM120A	ccsbBroad304_02734	23196	3.63
ICAM3	ccsbBroad304_06420	3385	3.58
HOXD9	ccsbBroad304_13874	3235	3.57
HNF4A	ccsbBroad304_00763	3172	3.53
CHTOP	ccsbBroad304_11810	26097	3.53
S1PR3	ccsbBroad304_00478	1903	3.50
MAP3K9	ccsbBroad304_10970	4293	3.50
MITF	ccsbBroad304_01017	4286	3.49
SALL4	ccsbBroad304_08715	57167	3.48
TBX22	ccsbBroad304_03162	50945	3.47
SPINT2	ccsbBroad304_15723	10653	3.44
MOBKL2C	ccsbBroad304_05021	148932	3.42
RGL1	ccsbBroad304_02728	23179	3.41
ALX3	ccsbBroad304_00059	257	3.41
MEOX1	ccsbBroad304_01000	4222	3.39
GAB2	ccsbBroad304_02260	9846	3.39
PRPF39	ccsbBroad304_12136	55015	3.39
TYR	ccsbBroad304_13974	7299	3.38
THEG	ccsbBroad304_03273	51298	3.38
MPPED1	ccsbBroad304_00197	758	3.36
MAFF	ccsbBroad304_02841	23764	3.36
HEY1	ccsbBroad304_02771	23462	3.36
POU3F2	ccsbBroad304_14774	NA	3.36
IGF2	ccsbBroad304_00836	3481	3.35
TERF2IP	ccsbBroad304_08370	54386	3.34
GET4	ccsbBroad304_03344	51608	3.32
FOXP3	ccsbBroad304_03161	50943	3.30
HOXC4	ccsbBroad304_06393	3221	3.30
FBXL12	ccsbBroad304_08418	54850	3.29
VCX2	ccsbBroad304_11990	51480	3.28
C3orf20	ccsbBroad304_04325	84077	3.27
ILKAP	ccsbBroad304_04228	80895	3.27
PRPS2	ccsbBroad304_01298	5634	3.26

FGF22	ccsbBroad304_08040	27006	3.25
PPP1R3B	ccsbBroad304_08943	79660	3.24
C18orf1	ccsbBroad304_10703	753	3.23
SRM	ccsbBroad304_06996	6723	3.19
RBP4	ccsbBroad304_06852	5950	3.19
HMGB1	ccsbBroad304_06381	3146	3.15
C12orf34	ccsbBroad304_09239	84915	3.14
PSPC1	ccsbBroad304_10243	55269	3.14
IRF2BP1	ccsbBroad304_07994	26145	3.14
MAF1	ccsbBroad304_09159	84232	3.14
CH25H	ccsbBroad304_07339	9023	3.14
PABPC5	ccsbBroad304_04956	140886	3.13
GTF2A1	ccsbBroad304_00705	2957	3.12
KLF6	ccsbBroad304_06021	1316	3.12
BATF	ccsbBroad304_07631	10538	3.12
HOXC8	ccsbBroad304_00777	3224	3.07
GLRA1	ccsbBroad304_06284	2741	3.05
ARHGEF9	ccsbBroad304_02740	23229	3.03
IQSEC1	ccsbBroad304_14033	9922	3.03
CDK5R1	ccsbBroad304_15646	8851	3.02
OR5M3	ccsbBroad304_09839	219482	3.02
FAM122B	ccsbBroad304_05100	159090	3.00
PMM2	ccsbBroad304_06743	5373	3.00