Published in final edited form as:

Cancer Discov. 2018 January ; 8(1): 108–125. doi:10.1158/2159-8290.CD-17-0532.

Somatic super-enhancer duplications and hotspot mutations lead to oncogenic activation of the KLF5 transcription factor

Xiaoyang Zhang^{1,2}, Peter S. Choi^{1,2}, Joshua M. Francis^{1,2,5}, Galen F. Gao², Joshua D. Campbell^{1,2,6}, Aruna Ramachandran^{1,2}, Yoichiro Mitsuishi^{1,2}, Gavin Ha^{1,2}, Juliann Shih², Francisca Vazquez², Aviad Tsherniak², Alison M. Taylor^{1,2}, Jin Zhou¹, Zhong Wu¹, Ashton C. Berger², Marios Giannakis^{1,2}, William C. Hahn^{1,2,4}, Andrew D. Cherniack^{1,2}, and Matthew Meyerson^{1,2,3,4}

¹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

²Cancer Program, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

³Department of Pathology, Harvard Medical School, Boston, MA 02115, USA

⁴Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA 02215, USA

Abstract

The Krüppel-like family of transcription factors (KLF) plays critical roles in human development and is associated with cancer pathogenesis. KLF5 has been shown to promote cancer cell proliferation and tumorigenesis, and to be genomically amplified in cancer cells. We recently reported that the *KLF5* gene is also subject to other types of somatic coding and noncoding genomic alterations in diverse cancer types. Here we show that these alterations activate KLF5 by three distinct mechanisms. 1) Focal amplification of super-enhancers activates *KLF5* expression in squamous cell carcinomas. 2) Missense mutations disrupt KLF5-FBXW7 interactions to increase KLF5 protein stability in colorectal cancer. 3) Cancer type-specific hotspot mutations within a zinc-finger DNA binding domain of KLF5 change its DNA binding specificity and reshape cellular transcription. Utilizing data from CRISPR/Cas9 gene knockout screening, we reveal that cancer cells with *KLF5* overexpression are dependent on KLF5 for their proliferation, suggesting KLF5 as a putative therapeutic target.

INTRODUCTION

Genomic alterations during tumorigenesis can lead to the activation of oncogenic transcription factors resulting in aberrant gene regulation throughout the genome. For example, somatic structural variations such as copy number amplifications increase gene dosage of *MYC*, *MYCN*, *AR*, *MITF* and *SOX2* and upregulate their expression (1–6); chromosomal translocations can place regulatory elements such as enhancers or super-

Corresponding author: Matthew Meyerson (matthew_meyerson@dfci.harvard.edu).

⁵Current address: Gritstone Oncology, Cambridge, MA 02142, USA

⁶Current address: School of Medicine, Boston University, Boston, MA 02118, USA

Conflict of interest: Galen F. Gao, Ashton C. Berger, Andrew D. Cherniack, and Matthew Meyerson receive research support from Bayer Pharmaceuticals. Matthew Meyerson is a consultant for and equity holder in OrigiMed. William C. Hahn receives research support from Novartis.

enhancers adjacent to oncogenes and activate their expression, as observed with *MYC*, *MYB* and *ERG* (7–12); while amplification of noncoding super-enhancers are known to activate *MYC* (13–15). In addition, somatic single nucleotide variants (SNV) can activate oncogenic transcription factors: for example, missense mutations in the degron domains of *NFE2L2* stabilize the protein by preventing its binding to the E3 ubiquitin ligase, KEAP1 (16,17). In noncoding regions, somatic mutations are known to increase the activity of distal enhancers or super-enhancers to activate *ESR1* and *TAL1* expression (18,19).

We and others have recently obtained genomic evidence that the Krüppel-like factor 5 gene, *KLF5*, could act as an oncogene. Previous studies have reported copy number amplification of broad regions on chromosome 13q harboring the *KLF5* gene in gastric and salivary gland tumors (20,21). We identified noncoding super-enhancers that are focally amplified ~300 kb 3' to the *KLF5* gene in head and neck squamous cell carcinomas (HNSC), which correlates with *KLF5* overexpression (15). In addition, we have identified recurrent missense mutations in a zinc-finger DNA binding domain of *KLF5* in lung adenocarcinomas and lung squamous cell carcinomas, and in a phospho-degron domain of *KLF5* in colorectal carcinomas (22,23).

Krüppel-like transcription factors (KLFs) play important roles in development and disease. KLF4 is one of the four key transcription factors required for maintaining the pluripotency of embryonic stem cells (24). In epithelial cells, KLF4 inhibits cell cycle progression and is highly expressed in terminally differentiated cells (25). In contrast, KLF5 promotes cell proliferation and is highly expressed in actively dividing cells (26). Previous studies have suggested that KLF5 has oncogenic properties. In addition to its role of as a positive regulator of cancer cell proliferation (27,28), overexpression of KLF5 has been reported to promote tumorigenesis of multiple cancer types including intestinal, bladder and gastric cancers (29–31). KLF5 has also been linked to intestinal tumorigenesis at the stem-cell level (32,33). Furthermore, KLF5 overexpression is also a prognostic marker for worse survival of breast cancer patients (34).

In light of this previous literature and our recent genomic data, we decided to systematically investigate noncoding and coding genomic alterations related to the *KLF5* gene and their transcriptional and phenotypic consequences. We performed functional analysis of each of these genomic alterations to understand how they contribute to oncogenic activation of KLF5 and their effects on KLF5 gene expression, protein stability and protein function. Our results highlight a variety of somatic genome alterations that converge to enhance the levels and activity of KLF5, and thereby to reshape cellular transcriptional programs and promote cancer cell proliferation.

RESULTS

Focal amplification of noncoding super-enhancers activates *KLF5* expression

To define the prevalence of *KLF5* super-enhancer amplification across cancers, we examined SNP-array-based copy number data targeting the ~600 kb intergenic region between *KLF5* and *KLF12* on chromosome segment 13q22.1 across 10,844 samples from 33 cancer types included in The Cancer Genome Atlas (TCGA). We discovered recurrent amplifications of this noncoding region in six other cancer types beyond head and neck squamous cell

carcinoma (15/522), including esophageal carcinomas (ESCA, 7/184), cervical squamous cell carcinomas (CESC, 14/295), lung squamous cell carcinomas (LUSC, 14/501), bladder carcinomas (BLCA, 12/408), stomach adenocarcinomas (STAD, 7/441), and colorectal adenocarcinomas (CRC, 5/615) (Figure 1A). Consistent with these observations, an analysis of SNP-array-based copy number data from 1043 cancer cell lines within the Broad Institute's Cancer Cell Line Encyclopedia (CCLE) project (35) identified focal amplification of this noncoding region in 12 cell lines, from the seven cancer types reported above (Figure S1A). Examination of the copy number profile from TCGA normal tissues (n=11,813) found no evidence of amplifications of the *KLF5* noncoding region.

To investigate the molecular basis for amplification of the *KLF5/KLF12* intervening region, we analyzed whole genome sequencing (WGS) data (36) for six head and neck squamous cell carcinoma samples bearing this amplification. DNA rearrangement analysis of the WGS data, using the structural variant calling program LUMPY (37), validated the focal amplification events in five of the six samples, and revealed that they occur in a tandem duplication pattern (Figure 1B). Correspondingly, DNA rearrangement analysis of WGS data from three cancer cell lines with this amplification, from disparate cancer types, revealed tandem duplication of the noncoding region (Figure S1B). Chromatin immunoprecipitation-sequencing (ChIP-seq) in cell lines representing eight head and neck squamous cell carcinomas, three esophageal carcinomas and three stomach adenocarcinomas showed similar profiles of histone H3 lysine 27 acetylation (H3K27ac), a marker of enhancer elements (38), at the *KLF5* noncoding locus (Figure 1C and S2A). Distinct from typical enhancers, super-enhancers are large clusters of enhancers that are associated with the activation of cell identity genes and cancer-related genes (39–41). We analyzed the H3K27ac ChIP-seq data using the ROSE pipeline (39–41) and identified several super-enhancers in the amplified region (Figure 1C, rectangle), one in ESCA cells and two in HNSC and STAD cells (Figure 1C, indicated by bars within the rectangle). Taken together, these data suggest that focal amplification of noncoding super-enhancers near the *KLF5* gene is a recurrent event in multiple cancer types, particularly squamous cell carcinomas.

The amplified super-enhancers are located in a ~600 kb noncoding region flanked by *KLF5* on the centromeric side and *KLF12* on the telomeric side (Figure 1A). Enhancers regulate gene expression through physical interaction with gene promoters (42–44) and these interactions are restricted by topologically associating domains (TAD), chromatin “neighborhoods” that are highly conserved across tissue types (45–51). Utilizing publicly available Hi-C data from IMR90 lung fibroblast cells that measures physical interactions between chromatin regions and defines TADs in the genome (49), we found that the amplified super-enhancers lie within the same TAD (small TAD, chr13:73,570,000–74,290,000; large TAD, chr13:73,350,000–74,290,000) as the promoter region and gene body of *KLF5*, but not the promoter or complete gene body of *KLF12*, suggesting that *KLF5* is the candidate target gene (Figure 2A and 2B). Indeed, a recent study in a stomach adenocarcinoma cell line identified significant chromatin interaction between the super-enhancer region and the *KLF5* (but not the *KLF12*) promoter, using circularized chromosome conformation capture assays (4C) (52). We performed chromosome conformation capture (3C) assays and validated the physical interaction between the super-

enhancer region and the *KLF5* promoter in cells with (BICR31) or without (BICR6) the super-enhancer duplication (Figure 2C, upper panel). Because most of the 13q22.1 super-enhancer amplifications were observed in squamous cell carcinomas (Figure 1), we analyzed RNA sequencing data from TCGA squamous cell carcinoma samples, including head and neck, cervical, lung and esophageal squamous carcinoma samples (16,36,53–55). We found *KLF5* expression to be greater than *KLF12* expression across all of these cancers. In addition, we observed a mean of 39.7% statistically significant elevation in expression (t test: $P < 0.0001$) of *KLF5* in cancers harboring the super-enhancer amplifications, compared to tumors without the amplifications, but no significant increase of *KLF12* expression (Figure S2B).

A combination of three individual enhancers within the amplified super-enhancers drives *KLF5* overexpression

We selected the head and neck squamous cell carcinoma cell line BICR31, in which the *KLF5* super-enhancers are focally amplified (Figure S1A), as a model system for detailed functional studies. ChIP-seq assays of p300, a marker for active enhancers (56), identified four strong individual enhancer elements in the super-enhancers in BICR31 cells (Figure 2B). We applied the CRISPR-mediated repression system (15,57,58), which uses a short guide RNA (sgRNA) to recruit inactivated Cas9 (dCas9) fused to the Krüppel-associated box (KRAB) transcriptional repressor domain (KRAB-dCas9), to repress the e1–e4 enhancers, individually. Repression of each of the individual enhancers e1, e3, or e4 alone (but not e2) resulted in a modest yet significant reduction (20~34%) in *KLF5* expression (Figure 2D). In agreement, 3C assays detected stronger physical interaction between the *KLF5* promoter and the e1, e3 and e4 enhancers, compared to the e2 enhancer (Figure 2C, lower panel). Repression of the individual enhancers did not affect expression of *KLF12*, present outside the TAD domain, or of *PIBF1* or *DIS3* within the same large TAD domain (Figure S3).

We next sought to interrogate the combinatorial effects of these enhancers. Transfection of a luciferase reporter construct containing all three enhancers gave rise to significantly higher luciferase expression than reporter constructs carrying an individual enhancer, suggesting a joint effect of the three enhancers in activating gene expression (Figure 2E). Furthermore, multiplexed repression of the e1, e3 and e4 enhancers by KRAB-dCas9 resulted in a marked decrease in overall enhancer activity, as observed by a loss of H3K27ac enrichment at the targeted regions (Figure 2F), along with a strong reduction (~51%) in *KLF5* expression and a modest reduction (~25%) in *PIBF1* and *DIS3* expression (Figure 2D and S3). These data reveal that the e1, e3 and e4 enhancers exert a combinatorial effect on gene activation, with *KLF5* as the primary gene target.

***KLF5* activates cell identity genes and cancer-related genes in squamous cell carcinomas**

To assess the gene regulatory functions of *KLF5* in squamous cell carcinomas, we performed ChIP-seq assays using an antibody against endogenous *KLF5* in the head and neck squamous carcinoma cell line BICR31. We observed that 20.7% of *KLF5* binding sites occurred at promoter regions (promoter enrichment: Fisher exact test, $P = 10^{-322}$), with 73.3% distributed across intergenic or intronic regions (Figure S4A). Motif analysis of the

KLF5 binding sites, using the SeqPos tool (59), revealed that KLF5 recognizes the same DNA binding motif (GGGG T/C GGGGC) as other Krüppel-like factors (KLF) and Specificity proteins (Sp) (Figure 3A) (60,61). We also identified DNA binding motifs for other transcription factors including ETS1, ERG, AP1 and TP63, suggesting their involvement in the oncogenic role of KLF5 (Figure S4B). Further analysis revealed that the KLF5 binding sites are enriched for p300 binding and H3K27ac modifications, indicating that KLF5 binding is associated with active regulatory elements (Figure 3B). Our results are consistent with previous reports that the transactivation function of KLF5 depends on its interaction with the CBP/p300 co-activator complex (62). Annotating KLF5 binding sites in more detail, we observed that KLF5 binding sites are more prevalent in super-enhancers than in typical enhancers. Indeed, individual active enhancers (as defined by p300 binding in BICR31 cells), are more likely to be bound by KLF5 (~33%) when present in super-enhancers rather than in typical enhancers (~22%) (Figure 3C).

To investigate the transcriptional impact of KLF5 expression, we conducted RNA-sequencing (RNA-seq) assays in BICR31 cells with and without siRNA-mediated silencing of *KLF5* (Figure S5A). We integrated the RNA-seq and KLF5 ChIP-seq results with the Binding and Expression Target Analysis (BETA) pipeline (63), which first assigns each gene in the genome a KLF5 regulatory potential score based on two criteria: (1) the number of KLF5 binding sites within ± 50 kb of the transcription start site for each queried gene, and (2) the distance between these KLF5 binding sites and the transcription start site. We then used BETA to interrogate the impact of perturbing KLF5 abundance on expression of each of these genes. This analysis revealed that KLF5 activates the expression of genes with higher KLF5 regulatory potential scores more often than it represses the expression of such genes (Figure 3D; Kolmogorov-Smirnov test, $P=2.3 \times 10^{-8}$), suggesting that KLF5 mainly acts as a transcriptional activator. We also detected a modest yet significant reduction (~20% in average, t test: $P < 0.001$) in the H3K27ac level surrounding KLF5 binding sites that are nearest to KLF5-activated genes (Figure S6). We observe that KLF5 activates squamous cell identity genes such as *KRT5*, *KRT8*, *KRT6A*, *KRT13*, *LAMA3*, *LAMB3*, and *LAMC2*, and cancer-related genes such as *ID1*, *CCND1*, *TP63*, *DEK*, *WNT10A*, *PDGFA*, and *PDGFB* (Figure 3E and S5B). To validate this observation, we targeted the KLF5 binding sites surrounding *ID1* by the KRAB-dCas9 repressor complex and found a significant decrease in *ID1* expression, demonstrating a direct role of KLF5 binding in activating *ID1* (Figure 3F).

Hotspot mutations in a phospho-degron domain increase KLF5 protein stability

In addition to focal amplifications of noncoding super-enhancers, mutations within the *KLF5* gene are also frequently found in cancer (22,23). We examined the mutation profile of KLF5 in > 11,000 tumor samples from the TCGA project (64–69) as well as 619 colorectal cancer samples from two prospective cohort studies (23). This analysis confirmed the presence of two mutation hotspots in *KLF5*: one within a phospho-degron domain and the other within a DNA binding domain (Figure 4A and S7; see Methods section for additional details on mutation hotspot identification). Three FBXW7 (also known as CDC4) phospho-degron domains (CPDs) have been identified in KLF5 (70), which have been shown, when phosphorylated, to bind the E3 ubiquitin ligase FBXW7, leading to the ubiquitination and degradation of KLF5 (70). Our analysis revealed that the second CPD (amino acids 301–

307: PPSPSS) is a target of missense mutations, seen mainly in colorectal adenocarcinomas (7/619; $P = 5.65 \times 10^{-30}$; data from (23)) (Figures 4A and S7A). A previous study has shown that the P301S mutation inhibits the interaction between FBXW7 and KLF5 and increases the protein stability of KLF5 (71). To assess if this is a common mechanism for the hotspot mutations, we included two other mutations, S303P and P304A, and performed a cycloheximide (CHX) chase assay in the colorectal cancer cell line HCT116 to measure their effects on KLF5 protein stability. We found that the three tested mutations significantly reduced degradation of KLF5 to a similar extent, compared to wild-type (WT) KLF5 (Figure 4B). Co-immunoprecipitation assays confirmed that the mutations impaired the interaction of KLF5 with FBXW7 (Figure 4C).

Notably, the *FBXW7* gene is also significantly mutated in colorectal cancers (~13%), with recurrent mutations enriched in the WD40 repeat domains required for interaction with its substrates (72,73) (Figure S8A). None of the colorectal cancer samples harboring KLF5 hotspot mutations had mutations in FBXW7 (Figure S8B). We tested three of the most recurrent FBXW7 missense mutations, R465C, R465H and R505C, and found that they indeed impaired the interaction of FBXW7 with KLF5 (Figure 4D). While overexpression of wild-type FBXW7 in HCT116 cells decreased the protein level of KLF5, the FBXW7 mutants showed an opposite effect (Figure 4E), consistent with previous findings that FBXW7 mutations have dominant-negative effects (74,75). Taken together, we found here that hotspot mutations within either the KLF5 CPD domain or the FBXW7 WD40 repeat domains act to stabilize KLF5 levels by preventing its binding to FBXW7.

Hotspot mutations in a DNA binding domain of KLF5 alter its DNA binding specificity

An additional hotspot mutation is found in KLF5 ($P = 4.26 \times 10^{-63}$; TCGA pan-cancer dataset) (64) within the second of three DNA-binding zinc-finger (ZNF) domains that are highly conserved within KLF family members (61), with significant recurrent mutations at the codons for D418 and E419 in lung adenocarcinomas (2/502) and lung squamous cell carcinomas (7/464) (22) (Figure 5A and S7B). Pan-cancer analysis identified additional hotspot mutations at these positions in cervical squamous cell carcinomas (6/272), bladder carcinomas (5/398), and stomach adenocarcinomas (1/383) (Figure 5A). Interestingly, these mutations are cancer-type specific. For example, the E419K mutation occurs predominantly in cervical squamous cell carcinomas while the E419Q mutation is observed only in lung cancers (Figure 5A).

To assess the function of these mutations, we generated N-terminal V5-tagged versions of wild-type KLF5 and three of the most recurrent mutants, D418N, E419K and E419Q, and infected them into HEK293T cells (Figure S9A). ChIP-seq analysis of these cells revealed that the mutations in the DNA binding domain alter the DNA binding specificity of KLF5 in a mutation-specific manner (Figure 5B). Changes in the cognate DNA binding motifs of KLF5 appeared to be predominantly restricted to nucleotides at the 5th and 6th position of the DNA motifs (Figure 5B), consistent with a report that the second zinc-finger domain of KLF transcription factors recognizes the 4–6th position of the DNA motif (76). The D418N mutant, seen in lung squamous cell carcinomas and bladder carcinomas, preferentially binds to thymidine (T) at the 6th nucleotide in the DNA motif, compared to guanine (G) for wild-

type KLF5 (Figure 5B). In addition, the E419K mutant, seen mainly in cervical squamous cell carcinomas, binds preferentially to guanine (G) at the 5th nucleotide of the DNA motif, while the E419Q mutant, specific to lung cancers, binds preferentially to adenine (A) at the same nucleotide position, compared to cytosine (C) or thymidine (T) for wild-type KLF5 (Figure 5B). Accordingly, KLF5 WT and mutant proteins bind to different regions of the genome (Figure 5B). When KLF5 binding sites are ranked by variability among HEK293T cells overexpressing different wild-type or mutant constructs, ~44%, 26% and 15% of the top 10% variable sites are preferentially bound by KLF5 D418N, E419K and E419Q, respectively (Figure 5B right panel for overview; Figure S9B for examples).

The KLF5 E419Q mutant gains novel binding sites, creates new super-enhancers and activates cancer-related genes such as *FOXE1* and *NAMPT*

To study the function of mutations in the KLF5 DNA binding domain in a more physiologically relevant context, we analyzed the lung cancer-specific E419Q mutation in the lung squamous cancer cell line HCC95, which is wild-type for the *KLF5* gene based on RNA sequencing results from the Cancer Cell Line Encyclopedia project (35). Ectopic expression and ChIP-seq analysis of V5-tagged KLF5 WT and KLF5 E419Q in HCC95 (Figure S10A–B) revealed that both wild-type and mutant KLF5 share 5,511 binding sites. Relative to KLF5 WT, however, KLF5 E419Q lost 483 binding sites and gained 5,611 new binding sites (Figure 5C). Electrophoretic mobility shift assays (EMSAs) using fluorescently-labeled DNA probes containing the KLF5 DNA motifs with the C, T and A variants at the 5th nucleotide revealed that KLF5 E419Q had a stronger binding affinity for the A variant, compared to wild-type KLF5 (Figure S10C), consistent with our results in HCC95 cells, above (Figure 5C). However, like wild-type KLF5, KLF5 E419Q also binds to DNA motifs with the C and T variants (Figure S10C), which explains the observation that KLF5 E419Q gains more binding sites across the genome, compared to KLF5 WT. The regions that are specifically bound by KLF5 E419Q are more enriched in intronic regions (Fisher exact test: $P = 1.1 \times 10^{-42}$) and intergenic regions ($P = 1.3 \times 10^{-6}$) but less enriched in promoter regions ($P = 7.2 \times 10^{-86}$), compared to regions that are shared by KLF5 WT and KLF5 E419Q (Figure S10D), suggesting a shift from promoters to distal enhancers for the novel binding sites of KLF5 E419Q. We then investigated the effect of KLF5 E419Q binding on enhancer activity. In HCC95 cells overexpressing KLF5 E419Q, the gained binding sites show enrichment of H3K27ac, compared to cells overexpressing KLF5 WT (Figure 5D).

We next performed gene expression analysis of HCC95 cells overexpressing untagged KLF5 WT or E419Q. Ectopic expression of either KLF5 WT or E419Q had little effect on the expression level of the endogenous *KLF5* gene, as measured by the PCR primers targeting the 3' UTR of *KLF5* (Figure S10B). By integrating the results of RNA sequencing and ChIP-seq using the BETA pipeline, we found that the binding sites gained by KLF5 E419Q are significantly associated with activation of the target genes (Figure 5E), suggesting a gene activation role for this mutant. Furthermore, the gained binding sites also form novel super-enhancers, as defined by H3K27ac enrichment, that are associated with activation of genes such as *FOXE1*, *NAMPT*, *EPHB3* and *GAS6* (Figure 5F and S10E). For instance, KLF5 E419Q binding occurring ~35 kb 3' to the *FOXE1* gene leads to a marked increase in

enhancer activity, as measured by the H3K27ac ChIP-seq profile, and the formation of a novel super-enhancer that upregulates *FOXE1* expression as measured by RNA-seq (Figure 5F). The *FOXE1* gene, encoding the Forkhead box protein E1, has been linked to thyroid cancer susceptibility (77) and inherited loss-of-function mutations of *FOXE1* cause cleft palate and hypothyroidism (78). Combined expression of *FOXE1* and *SOX2* has been shown to promote anchorage-independent growth of normal lung epithelial cell lines, suggesting an oncogenic role (1). Similarly, the binding of KLF5 E419Q ~95 kb upstream of *NAMPT* created a novel super-enhancer and activated *NAMPT* expression (Figure 5F). The *NAMPT* gene encodes nicotinamide phosphoribosyltransferase, a rate-limiting enzyme in the biosynthesis of the metabolite Nicotinamide Adenine Dinucleotide (NAD) (79). *NAMPT* is overexpressed in many cancer types including colorectal, breast, gastric and prostate cancers (80) and inhibition of *NAMPT* has been shown to impair tumor growth (81), suggesting its oncogenic function. In summary, our results indicate that the KLF5 E419Q mutant gains novel binding sites, creates new super-enhancers, and activates genes implicated in tumorigenesis.

Cancer cells with activated KLF5 are dependent on KLF5 for their proliferation

We next sought to investigate the phenotypic consequences of KLF5 activation in cancer cells. Silencing of *KLF5* using small interfering RNAs (siRNA) in the head and neck squamous cell carcinomas cell line BICR31, in which *KLF5* overexpression is driven by the 13q22.1 super-enhancer amplification (Figure 2), resulted in a marked reduction of cell proliferation (Figure 6A). In addition, multiplexed repression using KRAB-dCas9 and sgRNAs directed against the three enhancers e1, e3 and e4, that are amplified in head and neck squamous carcinomas (Figure 2), also resulted in a significant reduction in proliferation of the BICR31 cell line (Figure 6A). The proliferation-inhibitory effect of silencing *KLF5* can be partially rescued by ectopic expression of *ID1* (Figure 6B), a target gene of KLF5 in head and neck squamous carcinoma cells (Figure 3E–F). We then investigated the phenotypic outcomes of mutations in *KLF5*. Overexpression of the KLF5 E419Q mutant identified in lung squamous carcinomas significantly increased proliferation of the lung squamous cell carcinoma cell line, HCC95, compared to KLF5 WT, in low serum media (Figure 6C), suggesting an oncogenic role for the KLF5 E419Q mutant.

We next asked whether activation of KLF5 correlates with a dependency of cancer cells on the *KLF5* gene. We queried the publicly available genome-wide CRISPR/Cas9 gene knockout screening (GeCKO) dataset, including 32 cancer cell lines originating from diverse tissue types such as bone, skin, colon and pancreas (82). Gene dependency scores were calculated based on the abundance of each sgRNA before and after cell proliferation for 3~4 weeks following infection of the library; gene expression was measured by RNA-seq (82). We found that cancer cells with higher *KLF5* expression were more dependent on *KLF5* (i.e. exhibited a lower gene dependency score), suggesting that increased expression of *KLF5* confers a dependency on the *KLF5* gene for cell viability (Figure 6D). Since none of the 32 cell lines used in this analysis bear *KLF5* coding mutations, we could not investigate the dependency of KLF5 mutants.

DISCUSSION

Here we describe the functional analysis of the altered *KLF5* gene, with findings that support the concept of an oncogenic role for *KLF5*. These discoveries are based on the identification of somatic cancer genome alterations in or near the *KLF5* gene (15,22). Our pan-cancer analysis showed that *KLF5* is activated by multiple somatic genomic alterations including noncoding super-enhancer amplifications and coding mutations in a phospho-degron domain or a DNA binding domain (Figure 6E). The frequency of individual types of *KLF5* genomic alterations is modest. However, the combination of the three types of alterations markedly enhances the significance of *KLF5* as a candidate oncogene. This work extends and provides a mechanistic basis for previous observations that overexpression of wild-type *KLF5* promotes oncogenic phenotypes such as cellular proliferation, invasion and transformation *in vitro* and *in vivo* (27–31).

We have identified focal amplifications of *KLF5* noncoding super-enhancers in many squamous cell carcinomas and some adenocarcinomas. In contrast to the *MYC* locus, in which cancer type-specific super-enhancers are amplified (15), the same noncoding region ~300 kb 3' to *KLF5* is amplified in multiple anatomical and histological forms of cancer. This may occur because, in contrast to the *MYC* locus, the enhancer profile of the *KLF5* locus is shared across different cancer types. We and others have identified single individual enhancers within super-enhancers that drive the activity of the entire super-enhancer region (15,83). In contrast, the activity of the *KLF5* super-enhancer region is dependent on a combination of three individual enhancers, representing another type of enhancer structure within super-enhancers.

In addition to transcriptional regulation, we show that *KLF5* is activated at the protein level by missense mutations. *KLF5* contains three phospho-degron (CPD) domains which, upon phosphorylation, are recognized by the E3 ubiquitin ligase FBXW7, that promotes ubiquitination and degradation of its substrates (70). Studies have shown that several oncogenic proteins including CCNE1, MYC and NOTCH1 are substrates of FBXW7 (84–86) and are stabilized by mutations in the FBXW7 WD40 repeat domains required for substrate recognition (74,75,87). Our studies show that *KLF5* is a substrate of FBXW7 in colorectal cancers, and that mutations either in a phospho-degron domain of *KLF5* or in the WD40-repeat domains of FBXW7 stabilize *KLF5* protein levels by preventing the interaction of *KLF5* and FBXW7. The observation that no colorectal cancer samples have both *KLF5* CPD mutations and coding *FBXW7* mutations further supports their functional convergence. This mirrors the mutation pattern of the E3 ligase gene, *KEAP1*, and the oncogenic transcription factor gene *NFE2L2*, which encodes a substrate of KEAP1, in lung cancer (16,88). We expect that detailed characterization of protein domain interactions combined with mutual exclusivity analysis of genomic alterations will identify more such relationships between cancer-related genes.

Another mutation hotspot in the *KLF5* gene was identified in a zinc-finger DNA binding domain. We find that these mutations promote a change-of-function role, by altering *KLF5* DNA binding specificity. Our observations are consistent with recent findings reporting recurrent mutations in *KLF4*, another KLF family member gene, in meningiomas (89).

Unlike *KLF5* that is mutated in the second zinc-finger domain that recognizes the 4–6th nucleotide position of the KLF DNA motif, *KLF4* is mutated in the first zinc-finger domain that binds to the 7–10th nucleotides (76,89). Accordingly, the DNA motifs recognized by *KLF5* and *KLF4* mutants are different from the canonical KLF motif at the 5–6th and 9th nucleotide position, respectively (89). This suggests distinct oncogenic roles for these two KLF family members in their respective cancer types. Interestingly, although *KLF5* change-of-function mutations occur within a single zinc-finger domain, each mutation is highly cancer-type specific. Moreover, different mutations guide *KLF5* to recognize different DNA sequences, suggesting that individual *KLF5* mutants direct unique gene expression programs to drive tumorigenesis via distinct mechanisms in the relevant tumor types.

We showed that the lung cancer-specific *KLF5* E419Q mutant gains novel binding sites in the genome relative to wild-type *KLF5* while also maintaining the binding sites of the wild-type protein. This result contrasts with the finding of change-of-function mutations in *TP53* that lead to a switch in the DNA binding specificity of TP53 toward novel binding sites while eliminating binding sites recognized by wild-type TP53 (90,91). This difference may be because, unlike the tumor suppressor TP53, wild-type *KLF5* itself is an oncogenic transcription factor and thus losing wild-type *KLF5* binding sites may be disadvantageous to cancer cells. The gained binding sites of *KLF5* E419Q are associated with gene activation, as evident by the increased enhancer activity at these binding sites. Importantly, the newly acquired *KLF5* E419Q binding sites also create novel super-enhancers that drive expression of cancer-associated genes such as *FOXE1* and *NAMPT*, revealing new therapeutic targets. In addition to *KLF5*, somatic hotspot mutations have been identified in the DNA binding domains of other transcription factors such as FOXA1 and MAX (92,93). Furthermore, many germ-line genetic variants in genes encoding transcription factors have been predicted to alter DNA binding activity and specificity (94). Future studies focused on deeper functional characterization of these somatic and germ-line variants will likely uncover the specific mechanisms underlying their pathogenic features.

In addition to somatic genetic alterations, noncoding germ-line genetic variants near the *KLF5* gene have been associated with the development of prostate, pancreatic and endometrial cancers (95–100). This is reminiscent of the noncoding region near the *MYC* oncogene, where genetic variants have been associated with predisposition to multiple cancers (101). It is known that cancer risk-associated variants often target regulatory elements, modulate transcription factor binding and regulate expression of cancer-related genes (102–106). Interestingly, some of the cancer-risk variants near *KLF5*, such as rs9573163 and rs9543325 that are associated with pancreatic cancer risk (97,99,100), are within the super-enhancer regions that we found to be amplified in squamous carcinomas. The functional relevance of the 13q22.1 genetic risk-variants in regulating *KLF5* and cancer development needs further investigation.

In summary, we demonstrate that a single oncogenic transcription factor, *KLF5*, can be activated by multiple somatic genomic alterations including by the creation of noncoding structural genome variations and by hotspot missense mutations within the *KLF5* coding region. Importantly, we show that overexpression of *KLF5* is associated with a strong dependency on *KLF5* across 32 cancer cell lines. In addition, targeting *KLF5* *in vivo* has

been reported as an efficient anti-tumor strategy for breast, bladder and gastric cancers (28,30,107,108). All the evidence indicates the importance of KLF5 activation in cancer cells and its significance as an emerging target for the development of novel cancer therapeutics.

METHODS

Pan-Cancer copy number alteration analysis

GISTIC (Genomic Identification of Significant Targets in Cancer) analyses were performed in 10,844 samples from 33 tumor types, using copy number data from version 3.0 of the SNP pipeline on April 2nd, 2015 from the TCGA copy number portal (2,109). Arm-level amplifications or deletions were removed for GISTIC peak calling.

Cell lines

Cell lines were obtained from the Cancer Cell Line Encyclopedia (CCLE) project (35) in 2015 and 2016. Cells were tested negative for mycoplasma and maintained in RPMI-1640 medium supplemented with 10% heat inactivated FBS and 1% penicillin-streptomycin. Cell line identities were verified by SNP fingerprinting using an Affymetrix SNP array as previously described in the CCLE project (35). Cell lines were used for functional experiments, after less than 3 months of passages post receipt.

ChIP-seq analysis

Chromatin immunoprecipitation-sequencing (ChIP-seq) assays were performed as previously described (102,105). Briefly, cells were cross-linked with 1% Formaldehyde and lysed. The chromatin extract was sonicated by a Diagenode bioruptor and immunoprecipitated with antibodies that were co-incubated with mixed Dynabeads A and G (Thermo Scientific). Antibodies that were used include: H3K27ac (Abcam, ab4729), KLF5 (Abcam, ab137676), p300 (Bethyl Lab, A300-358), V5 (Thermo Fisher, R960-25). The sequencing libraries were prepared using the NEB ChIP-seq library prep kit (NEB, E6200L) and sequenced on the Illumina MiSeq instrument (50-bp single read reads). Sequencing reads were aligned to the hg19 human genome reference by Burrows-Wheeler Aligner (BWA) (110,111) and ChIP-seq binding sites were identified by MACS2 (111). Motif search was performed by using the SeqPos motif tool in the Cistrome pipeline (59).

For investigating the effect of KLF5 silencing on the H3K27ac profile (Figure S6), we used the “DNaseI Hypersensitive Site Master List” file generated by the ENCODE consortium (112) to identify open chromatin regions that are conserved across cell types and used them as “negative controls”. We selected the regions that are enriched with DNase I-hypersensitivity signal in more than half of the 125 ENCODE cell types included in the list and removed the ones that overlap with KLF5 bindings.

For clustering binding sites of KLF5 WT and mutants in HEK293T cells, we first concatenated and merged all of their binding sites identified by MACS2 and then mapped the sequencing reads to each of the merged binding sites by Bedtools (113). The number of reads at these binding sites were normalized by edgeR pipeline (114,115) and then log2

transformed. We performed k-means clustering for the top 10% most variable binding sites. To present the heatmap, the normalized binding signal was scaled by rows. For identifying binding sites that are specific to KLF5 WT or E419Q in the lung squamous cell carcinoma cell line HCC95, we used MACS2 and compared the ChIP-seq signal of V5-tagged KLF5 WT and E419Q by using each other as ‘treatment’ and ‘control’ for MACS2 input. For comparing the H3K27ac ChIP-seq signal between KLF5 E419Q and WT binding sites, because the difference of total sequencing reads between the ChIP-seq experiments are over 10%, we randomly subsampled the larger sample by Samtools (116) to normalize the signal. ChIP-seq data was uploaded to the Gene Expression Omnibus (GSE88976).

Super-enhancer identification

For each cancer type, H3K27ac ChIP-seq data from multiple cell lines was merged into one dataset. Based on the merged ChIP-seq results including the aligned reads and MACS2 binding peaks, we identified super-enhancers for each cancer type using the ROSE pipeline (39–41). To identify super-enhancers that are gained by KLF5 E419Q bindings, we used Bedtools (113) to compare the super-enhancers called from H3K27ac ChIP-seq signal in HCC95 cells overexpressing KLF5 WT and E419Q. We identified the super-enhancers that have > 75% region unique to cells overexpressing KLF5 E419Q and also overlap with KLF5 E419Q-specific binding sites. Genomic coordinates of the KLF5 E419Q-gained super-enhancers and the nearest genes are listed in Table S1.

Chromosome conformation capture (3C) assay

3C-qPCR assays were performed in BICR31 and BICR6 cells, as previously described (42,105). The restriction enzyme BglII was used to fragment DNA. BAC libraries (RP11-689G3, RP11-179I20, RP11-259I24, RP11-343F2, RP11-315L12, RP11-347N11 and RP11-46L3) of DNA fragments covering the tested regions were used as template controls for the normalization of digestion, ligation and primer efficiency. In order to normalize the DNA copy number in BICR31 cells, we doubled the input concentration of the BAC construct RP11-343F2 that covers the super-enhancer region. 3C ligation products were quantified by SYBR Green-based PCR and the primer sequences are listed in Table S2.

CRISPR/Cas9-mediated enhancer repression

CRISPR/Cas9 sgRNAs were identified using the sgRNA designer tool from the Broad Institute (117) and control, non-targeting sgRNAs were selected from the GeCKOv2 library (118). The enhancer repression vector lenti-KRAB-dCas9-blast was generated previously (15) and sgRNAs were cloned into lentiGuide-Puro (Addgene, 52963). BICR31 cells were first infected with lenti-KRAB-dCas9-blast and selected with 6 µg/ml blasticidin, and then subsequently infected with lentiGuide-sgRNAs and selected with 2 µg/ml puromycin. For multiplexed repression of the e1, e3, and e4 enhancers, lentivirus containing each sgRNA was mixed equally and then used for cell infection. sgRNA sequences were listed in Table S2.

Luciferase reporter assays

Luciferase reporter assays were performed as previously described (15). Individual enhancer regions were cloned upstream of the pGL3 minimal promoter vector using MluI and XhoI restriction enzyme sites. For cloning the three enhancers e1, e3, and e4 together into the vector, we used the Gibson assembly cloning method (NEB E2611S) that ligate multiple fragments by their overlaps. The reporter constructs were cotransfected with a control *Renilla* luciferase construct into cells using FuGENE 6 (Promega). The luciferase signal was normalized to the *Renilla* luciferase signal. Primers used for cloning are listed in Table S2.

siRNA-directed gene silencing

BICR31 cells were transfected with negative control, non-targeting siRNA (siNC) or siKLF5 using Lipofectamine RNAiMAX (Thermo Scientific). RNA was extracted 2 days after transfection using the Qiagen RNeasy kit with on-column DNase I treatment. Preverified Silencer Select siRNAs (Thermo Scientific, Negative Control No.1 and No.2 for siNC; s2115 and s2116 for siKLF5) were used. siNC #1 and siKLF5 #1 were used for RNA-seq assays with three biological replicates in BICR31 cells, and all the siRNAs were used for gene expression validation. To assess the effect of siRNAs, immunoblot analysis was performed using antibodies against KLF5 (Abcam, ab137676) and β -actin (Santa Cruz sc-47778).

Identification of *KLF5* mutation hotspots

To estimate the significance of mutation frequency within hotspots in the *KLF5* gene, we computed p-values for a sliding fixed-width window over the primary structure of *KLF5*. We implemented a binomial null distribution with n as the total number of *KLF5* mutations and p as the fraction of the primary structure of *KLF5* represented by our window. The p-value was then computed as the survival function of the binomial distribution where $k+1$ is the number of mutations actually present in the window. Windows of 3 amino acids and 5 amino acids were used to analyze the TCGA Pan-Cancer data set (64) and the colorectal cancer data set (23) respectively.

Ectopic expression of *KLF5* and *FBXW7*

Wild-type (WT) *KLF5* and *FBXW7* cDNA were first cloned into pJET1.2 (Thermo Scientific). Quik-change mutagenesis was then performed to generate cDNA of *KLF5* mutants (P301S, S303P, P304A, D418N, E419K, and E419Q) and *FBXW7* mutants (R465C, R465H, and R505C). The *KLF5* and *FBXW7* (WT and mutants) cDNA were then subcloned into the overexpression vector pLenti-EF1a-PGK-puro and pLenti-EF1a-PGK-blasti, respectively, with or without the V5 tag fused to the N-terminus. Infected HEK293T and HCC95 cells were selected by 2 μ g/ml puromycin or 10 μ g/ml blasticidin. Overexpression was validated by RT-PCR and immunoblot analysis. Primers used for cloning are listed in Table S2.

Cell proliferation assays

For siKLF5 experiments, BICR31 cells were transfected with siNC1 (#1 and #2) or siKLF5 (#1 and #2), and were then maintained in regular media for 6 days before cell counting

(Beckman Coulter Counter). For CRISPR-mediated enhancer repression experiments, BICR31 cells infected with sg-Control (#1 and #2) or combined sg-e1, e3 and e4 were selected by 2 ug/ml puromycin for 5 days. Cells were then seeded at the same cell number and maintained in regular media for 7 days before cell counting. For KLF5 WT versus E419Q overexpression experiments, HCC95 cells infected with KLF5 WT and E419Q overexpression constructs (with or without V5-tagged) were maintained in low serum condition (RPMI-1640 media supplemented with 1% FBS) for 7 days before cell counting.

RNA-seq analysis

For siKLF5 experiments, BICR31 cells transfected with siNC #1 and siKLF5 #1 (three biological replicates each condition) were maintained in regular media for 2 days before RNA extraction. For KLF5 WT vs. E419Q overexpression experiments, HCC95 cells infected with KLF5 WT and E419Q overexpression constructs (no-tagged, two biological replicates) were maintained in low serum condition (RPMI-1640 media supplemented with 1% FBS), which is consistent with the condition of cell proliferation assays of KLF5 E419Q overexpression, for 2 days before RNA extraction. RNA was extracted using Qiagen RNeasy kit and treated with on-column DNase I. RNA-sequencing (RNA-seq) libraries were prepared using the NEBNext Ultra Directional RNA library prep kit (NEB, E7420S) and sequenced on the Illumina MiSeq instrument (75-bp paired end reads). Sequencing reads were aligned using STAR (119) and expression level for each gene was quantified by RSEM (120). The differential expression analysis was performed using the edgeR and limma pipelines (115,121). The RNA-seq results were uploaded to the Gene Expression Omnibus (GSE88977).

BETA analysis to combine ChIP-seq and RNA-seq results

Binding and Expression Target Analysis (BETA) was performed to predict whether KLF5 has activating or repressive function by combining ChIP-seq and RNA-seq results. The analysis pipeline was described as previously described (63). Briefly, BETA estimates KLF5's regulatory potential score for each gene based on the distance between KLF5 binding sites and transcription start sites (TSS) of each gene, and also based on the number of KLF5 binding sites $\pm$ 50 kb centered at TSS of each gene. BETA then uses a nonparametric statistical test (Kolmogorov-Smirnov test) to compare regulatory potential scores for genes that are up-regulated, down-regulated, or not-regulated on the basis of RNA-seq results with and without siRNA-mediated silencing of *KLF5*. Similarly, we performed BETA analysis for analyzing KLF5 E419Q-unique binding sites and genes that are regulated by KLF5 E419Q overexpression (compared to KLF5 WT) in HCC95 cells.

Quantitative PCR

Quantitative PCR (qPCR) was performed using TaqMan Universal PCR Mastermix or Power SYBR green PCR Mastermix (Thermo Fisher) on an Bio-Rad C1000-Touch Real-time PCR instrument. For TaqMan PCR, the following premade 5' nuclease probes were ordered from Integrated DNA technologies: KLF5 (Hs.PT.56a.40282397), KLF12 (Hs.PT.58.28103949), PIBF1 (Hs.PT.58.21509866), DIS3 (Hs.PT.58.39902044), ID1 (Hs.PT.58.18791272.g), and internal references HPRT1 (Hs.PT.58v.45621572; for qPCR signal

normalization) and GAPDH (Hs.PT.58.589810.g). For SYPR green PCR, the primers used are listed in Table S2.

CHX chase assays

HEK293T cells infected with *KLF5* WT, P301S, S303P, P304A were treated with 100 ug/ml cycloheximide (CHX) for 0, 1, 2, and 3 hours before protein extraction and immunoblot analysis. The protein level of *KLF5* WT and mutants was quantified by using the LI-COR Image Studio software.

Co-immunoprecipitation assays

Antibodies were first incubated with mixed Dynabeads A and G (Thermo Fisher) for 5 hours at 4°C. Cells were lysed by cell lysis buffer (1% NP40, 150mM NaCl, 50mM Tris-HCl pH 8.0) supplemented with protease and phosphatase inhibitor. Antibodies that were used include V5 (Thermo Fisher, R960-25) and HA (Abcam, ab9110). Cell lysate were then incubated with the beads-antibody complex. Enriched protein was eluted and denatured at 65°C by LDS sample buffer (Thermo Fisher) supplemented with 20 mM DTT before immunoblot analysis.

EMSA assays

KLF5 WT and E419Q proteins were translated by using the TNT Quick Coupled Transcription/Translation System (Promega L1170). The translated protein was verified by immunoblot analysis using the *KLF5* antibody (Abcam, ab137676). The fluorescent DNA probes containing *KLF5* motifs were made from Integrated DNA Technologies and their sequences were listed in Table S2. For the electrophoretic mobility shift assays (EMSA), the translated *KLF5* proteins and the DNA probes were mixed, and incubated with binding reaction buffer (Final concentration: 10 mM Tris-HCl pH 7.5, 50 mM KCl, 2.5 mM DTT, 0.05 mM EDTA, 0.05 ug/ul Poly-dIdC, 0.25% Tween20) for 30 mins at room temperature. The reaction mix was added with orange loading dye and loaded on a Tris/Borate/EDTA (TBE) gel. Images were taken on a LI-COR instrument.

Accession codes

The newly generated ChIP-seq and RNA-seq data have been deposited to the Gene Expression Omnibus (GEO) public dataset under the series GSE88976 and GSE88977, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Meyerson laboratory for discussions. We thank Craig Strathdee, Hugh Gannon and Lior Golomb for reagents.

Financial support: We acknowledge support from the National Cancer Institute to M. Meyerson (1R35CA197568), National Cancer Institute Pathway to Independence awards to X. Zhang. (1K99CA215244) and to P.S. Choi

(1K99CA208028), and from the Norman R. Seaman Endowment fund to M. Meyerson. M. Meyerson is an American Cancer Society Research Professor.

References

1. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet.* 2009; 41:1238–42. [PubMed: 19801978]
2. Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. *Nature.* 2010; 463:899–905. [PubMed: 20164920]
3. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature.* 2005; 436:117–22. [PubMed: 16001072]
4. Visakorpi T, Hyytinen E, Koivisto P, Tanner M, Keinänen R, Palmberg C, et al. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet.* 1995; 9:401–6. [PubMed: 7795646]
5. Little CD, Nau MM, Carney DN, Gazdar AF, Minna JD. Amplification and expression of the c-myc oncogene in human lung cancer cell lines. *Nature.* 1983; 306:194–6. [PubMed: 6646201]
6. Schwab M, Alitalo K, Klempnauer K-H, Varmus HE, Bishop JM, Gilbert F, et al. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature.* 1983; 305:245–8. [PubMed: 6888561]
7. Bandopadhyay P, Ramkissoon LA, Jain P, Berghthold G, Wala J, Zeid R, et al. MYB-QKI rearrangements in angiogenic glioma drive tumorigenicity through a tripartite mechanism. *Nat Genet.* 2016; 48:273–82. [PubMed: 26829751]
8. Battey J, Moulding C, Taub R, Murphy W, Stewart T, Potter H, et al. The human c-myc oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell.* 1983; 34:779–87. [PubMed: 6414718]
9. Beroukhi R, Zhang X, Meyerson M. Copy number alterations unmasked as enhancer hijackers. *Nat Genet.* 2016; 49:5–6. [PubMed: 28029156]
10. Drier Y, Cotton MJ, Williamson KE, Gillespie SM, Ryan RJH, Kluk MJ, et al. An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. *Nat Genet.* 2016; 48:265–72. [PubMed: 26829750]
11. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun X-W, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science.* 2005; 310:644–8. [PubMed: 16254181]
12. Kron KJ, Murison A, Zhou S, Huang V, Yamaguchi TN, Shiah Y-J, et al. TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. *Nat Genet.* 2017; 49:1336–45. [PubMed: 28783165]
13. Herranz D, Ambesi-Impombato A, Palomero T, Schnell SA, Belver L, Wendorff AA, et al. A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat Med.* 2014; 20:1130–7. [PubMed: 25194570]
14. Shi J, Whyte WA, Zepeda-Mendoza CJ, Milazzo JP, Shen C, Roe J-S, et al. Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev.* 2013; 27:2648–62. [PubMed: 24285714]
15. Zhang X, Choi PS, Francis JM, Imielinski M, Watanabe H, Cherniack AD, et al. Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat Genet.* 2016; 48:176–82. [PubMed: 26656844]
16. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012; 489:519–25. [PubMed: 22960745]
17. Shibata T, Ohta T, Tong KI, Kokubu A, Odogawa R, Tsuta K, et al. Cancer related mutations in NRF2 impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc Natl Acad Sci.* 2008; 105:13568–73. [PubMed: 18757741]

18. Bailey SD, Desai K, Kron KJ, Mazrooei P, Sinnott-Armstrong NA, Treloar AE, et al. Noncoding somatic and inherited single-nucleotide variants converge to promote ESR1 expression in breast cancer. *Nat Genet.* 2016; 48:1260–6. [PubMed: 27571262]
19. Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science.* 2014; 346:1373–7. [PubMed: 25394790]
20. Deng N, Goh LK, Wang H, Das K, Tao J, Tan IB, et al. A comprehensive survey of genomic alterations in gastric cancer reveals systematic patterns of molecular exclusivity and co-occurrence among distinct therapeutic targets. *Gut.* 2012; 61:673–84. [PubMed: 22315472]
21. Giefing M, Wierzbicka M, Rydzanicz M, Cegla R, Kujawski M, Szyfter K. Chromosomal gains and losses indicate oncogene and tumor suppressor gene candidates in salivary gland tumors. *Neoplasma.* 2008; 55:55–60. [PubMed: 18190242]
22. Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet.* 2016; 48:607–16. [PubMed: 27158780]
23. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep.* 2016; 15:857–65.
24. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell.* 2006; 126:663–76. [PubMed: 16904174]
25. Shields JM, Christy RJ, Yang VW. Identification and Characterization of a Gene Encoding a Gut-enriched Krüppel-like Factor Expressed during Growth Arrest. *J Biol Chem.* 1996; 271:20009–17. [PubMed: 8702718]
26. Sun R, Chen X, Yang VW. Intestinal-enriched Krüppel-like factor (Krüppel-like factor 5) is a positive regulator of cellular proliferation. *J Biol Chem.* 2001; 276:6897–900. [PubMed: 11152667]
27. Jia L, Zhou Z, Liang H, Wu J, Shi P, Li F, et al. KLF5 promotes breast cancer proliferation, migration and invasion in part by upregulating the transcription of TNFAIP2. *Oncogene.* 2016; 35:2040–51. [PubMed: 26189798]
28. Qin J, Zhou Z, Chen W, Wang C, Zhang H, Ge G, et al. BAP1 promotes breast cancer cell proliferation and metastasis by deubiquitinating KLF5. *Nat Commun.* 2015; 6:8471. [PubMed: 26419610]
29. Chen C, Benjamin MS, Sun X, Otto KB, Guo P, Dong X-Y, et al. KLF5 promotes cell proliferation and tumorigenesis through gene regulation and the TSU-Pr1 human bladder cancer cell line. *Int J Cancer.* 2006; 118:1346–55. [PubMed: 16184550]
30. Chia N-Y, Deng N, Das K, Huang D, Hu L, Zhu Y, et al. Regulatory crosstalk between lineage-survival oncogenes KLF5, GATA4 and GATA6 cooperatively promotes gastric cancer development. *Gut.* 2015; 64:707–19. [PubMed: 25053715]
31. Nandan MO, McConnell BB, Ghaleb AM, Bialkowska AB, Sheng H, Shao J, et al. Krüppel-like factor 5 mediates cellular transformation during oncogenic KRAS-induced intestinal tumorigenesis. *Gastroenterology.* 2008; 134:120–30. [PubMed: 18054006]
32. Nakaya T, Ogawa S, Manabe I, Tanaka M, Sanada M, Sato T, et al. KLF5 regulates the integrity and oncogenicity of intestinal stem cells. *Cancer Res.* 2014; 74:2882–91. [PubMed: 24626089]
33. Fujii Y, Yoshihashi K, Suzuki H, Tsutsumi S, Mutoh H, Maeda S, et al. CDX1 confers intestinal phenotype on gastric epithelial cells via induction of stemness-associated reprogramming factors SALL4 and KLF5. *Proc Natl Acad Sci U S A.* 2012; 109:20584–9. [PubMed: 23112162]
34. Tong D, Czerwenka K, Heinze G, Ryffel M, Schuster E, Witt A, et al. Expression of KLF5 is a prognostic factor for disease-free survival and overall survival in patients with breast cancer. *Clin Cancer Res Off J Am Assoc Cancer Res.* 2006; 12:2442–8.
35. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature.* 2012; 483:603–7. [PubMed: 22460905]
36. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature.* 2015; 517:576–82. [PubMed: 25631445]

37. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* 2014; 15:R84. [PubMed: 24970577]
38. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci.* 2010; 107:21931–6. [PubMed: 21106759]
39. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell.* 2013; 155:934–47. [PubMed: 24119843]
40. Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell.* 2013; 153:320–34. [PubMed: 23582323]
41. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell.* 2013; 153:307–19. [PubMed: 23582322]
42. Bailey SD, Zhang X, Desai K, Aid M, Corradin O, Cowper-Sal Lari R, et al. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat Commun.* 2015; 2:6186. [PubMed: 25645053]
43. Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, et al. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell.* 2012; 149:1233–44. [PubMed: 22682246]
44. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature.* 2010; 467:430–5. [PubMed: 20720539]
45. Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. *Cell.* 2016; 164:1110–21. [PubMed: 26967279]
46. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature.* 2012; 485:376–80. [PubMed: 22495300]
47. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin Architecture Reorganization during Stem Cell Differentiation. *Nature.* 2015; 518:331–6. [PubMed: 25693564]
48. Downen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell.* 2014; 159:374–87. [PubMed: 25303531]
49. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 2014; 159:1665–80. [PubMed: 25497547]
50. Hnisz D, Day DS, Young RA. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell.* 2016; 167:1188–200. [PubMed: 27863240]
51. Hnisz D, Weintraub AS, Day DS, Valton A-L, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science.* 2016; 351:1454–8. [PubMed: 26940867]
52. Ooi WF, Xing M, Xu C, Yao X, Ramlee MK, Lim MC, et al. Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nat Commun.* 2016; 7:12983. [PubMed: 27677335]
53. Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, Pugh TJ, et al. Landscape of genomic alterations in cervical carcinomas. *Nature.* 2014; 506:371–5. [PubMed: 24390348]
54. The Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. *Nature.* 2017; 541:169–75. [PubMed: 28052061]
55. The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. *Nature.* 2017; 543:378–84. [PubMed: 28112728]
56. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature.* 2009; 457:854–8. [PubMed: 19212405]
57. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell.* 2014; 159:647–61. [PubMed: 25307932]

58. Thakore PI, D'Ippolito AM, Song L, Safi A, Shivakumar NK, Kabadi AM, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods*. 2015; 12:1143–9. [PubMed: 26501517]
59. Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol*. 2011; 12:R83. [PubMed: 21859476]
60. Philipsen S, Suske G. A tale of three fingers: the family of mammalian Sp/XKLF transcription factors. *Nucleic Acids Res*. 1999; 27:2991–3000. [PubMed: 10454592]
61. Turner J, Crossley M. Mammalian Krüppel-like transcription factors: more than just a pretty finger. *Trends Biochem Sci*. 1999; 24:236–40. [PubMed: 10366853]
62. Zhang Z, Teng CT. Phosphorylation of Kruppel-like factor 5 (KLF5/IKLF) at the CBP interaction region enhances its transactivation function. *Nucleic Acids Res*. 2003; 31:2196–208. [PubMed: 12682370]
63. Wang S, Sun H, Ma J, Zang C, Wang C, Wang J, et al. Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nat Protoc*. 2013; 8:2502–15. [PubMed: 24263090]
64. The Cancer Genome Atlas Research Network. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013; 45:1113–20. [PubMed: 24071849]
65. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet*. 2013; 45:1127–33. [PubMed: 24071851]
66. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013; 499:214–8. [PubMed: 23770567]
67. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet*. 2013; 45:1134–40. [PubMed: 24071852]
68. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013; 6:pl1. [PubMed: 23550210]
69. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov*. 2012; 2:401–4. [PubMed: 22588877]
70. Liu N, Li H, Li S, Shen M, Xiao N, Chen Y, et al. The Fbw7/human CDC4 tumor suppressor targets proliferative factor KLF5 for ubiquitination and degradation through multiple phosphodegron motifs. *J Biol Chem*. 2010; 285:18858–67. [PubMed: 20388706]
71. Bialkowska AB, Liu Y, Nandan MO, Yang VW. A colon cancer-derived mutant of Krüppel-like factor 5 (KLF5) is resistant to degradation by glycogen synthase kinase 3 β (GSK3 β) and the E3 ubiquitin ligase F-box and WD repeat domain-containing 7 α (FBW7 α). *J Biol Chem*. 2014; 289:5997–6005. [PubMed: 24398687]
72. Hao B, Oehlmann S, Sowa ME, Harper JW, Pavletich NP. Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol Cell*. 2007; 26:131–43. [PubMed: 17434132]
73. Orlicky S, Tang X, Willems A, Tyers M, Sicheri F. Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell*. 2003; 112:243–56. [PubMed: 12553912]
74. O'Neil J, Grim J, Strack P, Rao S, Tibbitts D, Winter C, et al. FBW7 mutations in leukemic cells mediate NOTCH pathway activation and resistance to gamma-secretase inhibitors. *J Exp Med*. 2007; 204:1813–24. [PubMed: 17646409]
75. Thompson BJ, Buonamici S, Sulis ML, Palomero T, Vilimas T, Basso G, et al. The SCFFBW7 ubiquitin ligase complex as a tumor suppressor in T cell leukemia. *J Exp Med*. 2007; 204:1825–35. [PubMed: 17646408]
76. Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, et al. Structural basis for Klf4 recognition of methylated DNA. *Nucleic Acids Res*. 2014; 42:4859–67. [PubMed: 24520114]

77. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, et al. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet.* 2009; 41:460–4. [PubMed: 19198613]
78. Clifton-Bligh RJ, Wentworth JM, Heinz P, Crisp MS, John R, Lazarus JH, et al. Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. *Nat Genet.* 1998; 19:399–401. [PubMed: 9697705]
79. Revollo JR, Grimm AA, Imai S. The NAD biosynthesis pathway mediated by nicotinamide phosphoribosyltransferase regulates Sir2 activity in mammalian cells. *J Biol Chem.* 2004; 279:50754–63. [PubMed: 15381699]
80. Shackelford RE, Mayhall K, Maxwell NM, Kandil E, Coppola D. Nicotinamide phosphoribosyltransferase in malignancy: a review. *Genes Cancer.* 2013; 4:447–56. [PubMed: 24386506]
81. Wang B, Hasan MK, Alvarado E, Yuan H, Wu H, Chen WY. NAMPT overexpression in prostate cancer and its contribution to tumor cell survival and stress response. *Oncogene.* 2011; 30:907–21. [PubMed: 20956937]
82. Aguirre AJ, Meyers RM, Weir BA, Vazquez F, Zhang C-Z, Ben-David U, et al. Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov.* 2016; 6:914–29. [PubMed: 27260156]
83. Shin HY, Willi M, Yoo KH, Zeng X, Wang C, Metser G, et al. Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat Genet.* 2016; 48:904–11. [PubMed: 27376239]
84. Koepf DM, Schaefer LK, Ye X, Keyomarsi K, Chu C, Harper JW, et al. Phosphorylation-dependent ubiquitination of cyclin E by the SCFFbw7 ubiquitin ligase. *Science.* 2001; 294:173–7. [PubMed: 11533444]
85. Oberg C, Li J, Pauley A, Wolf E, Gurney M, Lendahl U. The Notch intracellular domain is ubiquitinated and negatively regulated by the mammalian Sel-10 homolog. *J Biol Chem.* 2001; 276:35847–53. [PubMed: 11461910]
86. Welcker M, Orian A, Jin J, Grim JE, Grim JA, Harper JW, et al. The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation. *Proc Natl Acad Sci U S A.* 2004; 101:9085–90. [PubMed: 15150404]
87. King B, Trimarchi T, Reavie L, Xu L, Mullenders J, Ntziachristos P, et al. The Ubiquitin Ligase FBXW7 Modulates Leukemia-Initiating Cell Activity by Regulating MYC Stability. *Cell.* 2013; 153:1552–66. [PubMed: 23791182]
88. Jaramillo MC, Zhang DD. The emerging role of the Nrf2–Keap1 signaling pathway in cancer. *Genes Dev.* 2013; 27:2179–91. [PubMed: 24142871]
89. Clark VE, Harmanci AS, Bai H, Youngblood MW, Lee TI, Baranoski JF, et al. Recurrent somatic mutations in POLR2A define a distinct subset of meningiomas. *Nat Genet.* 2016; 48:1253–9. [PubMed: 27548314]
90. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 2014; 505:495–501. [PubMed: 24390350]
91. Zhu J, Sammons MA, Donahue G, Dou Z, Vedadi M, Getlik M, et al. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature.* 2015; 525:206–11. [PubMed: 26331536]
92. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat J-P, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet.* 2012; 44:685–9. [PubMed: 22610119]
93. Kamburov A, Lawrence MS, Polak P, Leshchiner I, Lage K, Golub TR, et al. Comprehensive assessment of cancer missense mutation clustering in protein structures. *Proc Natl Acad Sci U S A.* 2015; 112:E5486–5495. [PubMed: 26392535]
94. Barrera LA, Vedenko A, Kurland JV, Rogers JM, Gisselbrecht SS, Rossin EJ, et al. Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science.* 2016; 351:1450–4. [PubMed: 27013732]

95. Cheng THT, Thompson DJ, O'Mara TA, Painter JN, Glubb DM, Flach S, et al. Five endometrial cancer risk loci identified through genome-wide association analysis. *Nat Genet.* 2016; 48:667–74. [PubMed: 27135401]
96. Childs EJ, Mocci E, Campa D, Bracci PM, Gallinger S, Goggins M, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat Genet.* 2015; 47:911–6. [PubMed: 26098869]
97. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet.* 2010; 42:224–8. [PubMed: 20101243]
98. Takata R, Akamatsu S, Kubo M, Takahashi A, Hosono N, Kawaguchi T, et al. Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet.* 2010; 42:751–4. [PubMed: 20676098]
99. Wolpin BM, Rizzato C, Kraft P, Kooperberg C, Petersen GM, Wang Z, et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat Genet.* 2014; 46:994–1000. [PubMed: 25086665]
100. Wu C, Miao X, Huang L, Che X, Jiang G, Yu D, et al. Genome-wide association study identifies five loci associated with susceptibility to pancreatic cancer in Chinese populations. *Nat Genet.* 2011; 44:62–6. [PubMed: 22158540]
101. Grisanzio C, Freedman ML. Chromosome 8q24–Associated Cancers and MYC. *Genes Cancer.* 2010; 1:555–9. [PubMed: 21779458]
102. Cowper-Sal-lari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoutte J, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet.* 2012; 44:1191–8. [PubMed: 23001124]
103. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet.* 2009; 41:882–4. [PubMed: 19561607]
104. Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in cis through a Large Chromatin Loop Linked to a Cancer Risk-Associated Single-Nucleotide Polymorphism in Colorectal Cancer Cells. *Mol Cell Biol.* 2010; 30:1411–20. [PubMed: 20065031]
105. Zhang X, Cowper-Sal lari R, Bailey SD, Moore JH, Lupien M. Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res.* 2012; 22:1437–46. [PubMed: 22665440]
106. Zhang X, Bailey SD, Lupien M. Laying a solid foundation for Manhattan 'Setting the functional basis for the post-GWAS era'. *Trends Genet TIG.* 2014; 30:140–9. [PubMed: 24661571]
107. Zhao D, Zhi X, Zhou Z, Chen C. TAZ antagonizes the WWP1-mediated KLF5 degradation and promotes breast cell proliferation and tumorigenesis. *Carcinogenesis.* 2012; 33:59–67. [PubMed: 22045023]
108. Gao Y, Wu K, Chen Y, Zhou J, Du C, Shi Q, et al. Beyond proliferation: KLF5 promotes angiogenesis of bladder cancer through directly regulating VEGFA transcription. *Oncotarget.* 2015; 6:43791–805. [PubMed: 26544730]
109. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011; 12:R41. [PubMed: 21527027]
110. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl.* 2009; 25:1754–60.
111. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* 2008; 9:R137. [PubMed: 18798982]
112. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature.* 2012; 489:75–82. [PubMed: 22955617]
113. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma Oxf Engl.* 2010; 26:841–2.
114. Nikolayeva O, Robinson MD. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. *Methods Mol Biol Clifton NJ.* 2014; 1150:45–79.

115. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma Oxf Engl*. 2010; 26:139–40.
116. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl*. 2009; 25:2078–9.
117. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*. 2016; 34:184–91. [PubMed: 26780180]
118. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods*. 2014; 11:783–4. [PubMed: 25075903]
119. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinforma Oxf Engl*. 2013; 29:15–21.
120. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011; 12:323. [PubMed: 21816040]
121. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43:e47. [PubMed: 25605792]
122. Shao DD, Tsherniak A, Gopal S, Weir BA, Tamayo P, Stransky N, et al. ATARiS: Computational quantification of gene suppression phenotypes from multisample RNAi screens. *Genome Res*. 2013; 23:665–78. [PubMed: 23269662]

SIGNIFICANCE

Our observations, together with previous studies that identified oncogenic properties of KLF5, establish the importance of KLF5 activation in human cancers, delineate the varied genomic mechanisms underlying this occurrence, and nominate KLF5 as a putative target for therapeutic intervention in cancer.

Figure 1. Super-enhancers near the *KLF5* gene are focally amplified in diverse cancer types
 A. Copy number profile of the 13q22.1 noncoding region from head and neck squamous cell carcinomas (HNSC), cervical squamous cell carcinomas (CESC), lung squamous cell carcinomas (LUSC), esophageal carcinomas (ESCA), bladder carcinomas (BLCA), stomach adenocarcinomas (STAD), and colorectal carcinomas (CRC). The copy number peak, defined by statistical analysis with GISTIC (2,109), in HNSC is highlighted. Color code is based on lineage types: squamous cell carcinomas, blue; urothelial carcinomas, green; adenocarcinomas, orange.

B. DNA rearrangement analysis of the amplified noncoding region, using whole-genome sequencing data of head and neck squamous carcinoma samples from The Cancer Genome Atlas (TCGA) and the LUMPY program (37), demonstrates tandem duplications, as indicated by the curves.

C. The merged ChIP-seq signal of the enhancer marker H3K27ac from cell lines representing HNSC, ESCA, and STAD. Super-enhancers, indicated by thin bars, are called by the ROSE pipeline (39–41) based on the H3K27ac signal enrichment.

Figure 2. The focally amplified super-enhancers activate *KLF5* expression

A. Chromatin interaction, as measured by Hi-C in the lung fibroblast cell line IMR90, is presented in the *KLF5* locus. The topologically associated domains (TAD) are indicated as grey bars.

B. Four individual enhancers, e1-e4, within the super-enhancers are defined by p300 ChIP-seq signal from the HNSC cell line BICR31.

C. Chromatin interaction, as measured by 3C-qPCR, between the e1 enhancer and promoters of surrounding genes including *BORA*, *DIS3*, *PIBF1*, *KLF5*, and *KLF12* (upper panel), and

between the *KLF5* promoter and the four individual enhancers e1-e4 of the super-enhancer region (lower panel) in BICR31 and BICR6 cells (n=2). The interaction frequency between the *KLF5* promoter and the e1 enhancer in each panel is represented by the same data. 3C viewpoints are indicated as grey lines. Error bars, standard deviation (s.d.).

D. The expression level of *KLF5* after KRAB-dCas9-mediated repression of the individual enhancers e1-e4 in BICR31 cells (n=2). sg-Ctrl #1 and #2: control sgRNAs that are predicted to not recognize any genomic regions. Two separate sgRNAs are applied for each enhancer. Multiplexed repression of the e1, e3, and e4 enhancers (sg-e1 #1, sg-e3 #1, and sg-e4 #1) is highlighted in red. The expression level is normalized to the control (sg-Ctrl #1). Error bars, s.d. *P* values were derived from *t* tests: *P 0.05, **P 0.01.

E. Luciferase reporter assays (n=3) measuring the activity of the individual enhancers e1, e3 and e4, and the combinatorial activity of the three enhancers in driving the luciferase expression in BICR31 cells. Luciferase signal is normalized to the empty luciferase reporter construct. Error bars, s.d. *P* values were derived from *t* tests: **P 0.01, ***P 0.001.

F. ChIP-seq profile of H3K27ac in BICR31 cells with and without KRAB-dCas9-mediated multiplexed repression of the three enhancers e1, e3 and e4. The targeted regions are highlighted as grey boxes.

Figure 3. KLF5 activates cell identity genes and cancer-related genes in head and neck squamous cell carcinoma cells

A. Predicted DNA binding motif of KLF5 derived from the DNA binding pattern of endogenous KLF5, detected by ChIP-seq.

B. Enrichment of p300 binding and H3K27ac marks centered around KLF5 binding sites (n=10,562) in BICR31 cells.

C. Percentage of individual enhancers, as defined by p300 binding, bound by KLF5 in typical and super-enhancers called from H3K27ac ChIP-seq signals merged from eight HNSC cell lines. P value was derived from a fisher exact test.

D. Binding and Expression Target Analysis (BETA) predicting the activating and repressive function of KLF5. The KLF5 ChIP-seq binding sites are integrated with the expression data from the RNA-seq profile in BICR31 cells with and without siRNA-mediated *KLF5* silencing (n=3). More details are described in the Methods section. The red, grey and black lines represent genes activated, repressed or unaffected by KLF5, respectively. Percentage of genes is cumulated by the rank of genes based on their regulatory potential scores. *P* values were derived from Kolmogorov-Smirnov tests.

E. Examples of super-enhancer-driven cell identity genes (left) and cancer-related genes (right) activated by KLF5. ChIP-seq profile of KLF5 binding (in BICR31) and H3K27ac marks (merged from eight HNSC cell lines), and distribution of the identified super-enhancers (SE). Fold change in the expression level of KLF5-target genes in BICR31 cells with and without siRNA-mediated *KLF5* silencing, as measured by RNA-seq (n=3), is indicated underneath.

F. KRAB-dCas9 mediated repression of the e1 and e2 enhancers adjacent to *ID1* (indicated in Figure 3E; n=3) reduced *ID1* expression. Expression levels are normalized to the control (sg-Ctrl #1). Error bars, s.d. *P* values were derived from *t* tests: *P 0.05, **P 0.01, ***P 0.001.

Figure 4. Functional characterization of *KLF5* hotspot mutations in a phospho-degron domain

A. Two mutation hotspots were identified in the *KLF5* gene in a phospho-degron domain, and in a zinc-finger DNA binding domain. Mutations in different cancer types are color-coded.

B. The colorectal carcinoma HCT116 cell line expressing V5-tagged wild-type KLF5, or KLF5 P301S, S303P or P304A mutants, was treated with 100 ug/ml cycloheximide (CHX) for 0, 1, 2, and 3 hours, followed by immunoblotting for V5 and actin. Protein levels of WT

and mutant KLF5 after CHX treatment was quantified (n=2) and normalized to 0 hour (no treatment). Error bars, s.d. *P* values were derived from student *t* tests: *P 0.05.

C. Co-immunoprecipitation assays using antibodies against V5 (tagged to KLF5) and HA (tagged to FBXW7) in HCT116 cells overexpressing V5-tagged KLF5 WT and mutants.

D. Co-immunoprecipitation assays using antibodies against V5 (tagged to KLF5) and HA (tagged to FBXW7) in HCT116 cells overexpressing HA-tagged FBXW7 WT and mutants.

E. Immunoblots show the protein level of V5-tagged KLF5 in HCT116 cells overexpressing HA-tagged FBXW7 WT and mutants. The protein level of V5-tagged KLF5 was quantified (n=3) and normalized to HCT116 cells transfected with an empty vector. Error bars, s.d. *P* values were derived from *t* tests: *P 0.05; ***P 0.001.

Figure 5. Functional characterization of *KLF5* hotspot mutations in a DNA binding domain

A. A *KLF5* mutation hotspot was identified in a zinc-finger DNA binding domain.

Mutations in different cancer types are color-coded.

B. Left: ChIP-seq assays in HEK293T cells revealed the DNA binding motifs recognized by *KLF5* WT and mutants. The nucleotide differences in the DNA binding motifs are highlighted by green boxes. Right: the binding profile of *KLF5* WT and mutants in the top 10% most variable *KLF5* binding sites ($n=1,165$). Normalization of the binding signal is described in the Methods section.

C. Comparison of binding sites of KLF5 WT and E419Q in the lung squamous carcinoma cell line HCC95. DNA binding motifs are identified in the binding sites shared or unique to KLF5 WT and E419Q.

D. Averaged ChIP-seq signal of V5-KLF5 (left) and H3K27ac (right), centered at the gained, shared, or lost binding sites of KLF5 E419Q, in HCC95 cells.

E. Binding and Expression Target Analysis (BETA) predicting the activating and repressive function of KLF5 E419Q. The 5,611 KLF5 E419Q-unique binding sites were used for the analysis. The gene expression data were derived from RNA sequencing in the lung squamous carcinoma HCC95 cell line with KLF5 WT or E419Q overexpressed (n=2).

F. Examples of KLF5 E419Q target genes. ChIP-seq profile of V5 (KLF5) and H3K27ac in HCC95 cells overexpressing KLF5 WT or E419Q. The novel super-enhancers associated with KLF5 E419Q are indicated. The fold change of the target genes *FOXE1* and *NAMPT* between HCC95 cells overexpressing KLF5 WT and E419Q, as measured by RNA-seq (n=2), is indicated on the bottom.

Figure 6. *KLF5* activation confers a dependency of cancer cells on *KLF5*

A. Left: Cell proliferation assay of the head and neck squamous carcinoma cell line, BICR31, with and without siRNA-mediated *KLF5* silencing (n=3, cells were counted 6 days post transfection). Right: Cell proliferation assay of BICR31 with and without KRAB-dCas9-mediated multiplexed repression of the e1, e3 and e4 enhancers of *KLF5* (n=3, cells were counted 7 days after seeded). Cell number is normalized to the controls (siNC #1 or sg-Ctrl #1). Error bars, s.d. *P* values were derived from *t* tests: **P < 0.01.

B. Left: immunoblots showing the ectopic expression of ID1 protein in BICR31 cells. Right: overexpression of *ID1* rescued the proliferation-inhibitory effect of silencing *KLF5* in BICR31 cells (n=3). Cell number is normalized to the controls (siNC #1-Empty or siNC #1-ID1). Error bars, s.d. *P* values were derived from *t* tests: **P < 0.01.

C. Cell proliferation assay in the lung squamous carcinoma cell line HCC95 overexpressing *KLF5* WT or E419Q (with or without V5 tag), in low serum (1% FBS) media (n=3). Cell number is normalized to an empty vector control. Error bars, s.d. *P* values were derived from *t* tests: **P < 0.01.

D. The relationship between the gene expression level of *KLF5* (log2 transformed RNA-seq TPM, transcripts per million reads values) and CRISPR gene dependency ATARiS score (122) of *KLF5* across 32 cancer cell lines that were included in the Broad Institute GeCKO gene knockout screening (82).

E. Schematic diagram: *KLF5* can be activated on the transcriptional level by noncoding super-enhancer amplifications, on the protein level by missense mutations in a CPD phospho-degron domain of *KLF5* or in the WD40-repeat protein interaction domains of FBXW7, and on the activity level by missense mutations in a zinc-finger DNA binding domain of *KLF5*.