Somatic Superenhancer Duplications and Hotspot Mutations Lead to Oncogenic Activation of the KLF5 Transcription Factor

Xiaoyang Zhang1,2, Peter S. Choi1,2, Joshua M. Francis1,2, Galen F. Gao2, Joshua D. Campbell1,2, Aruna Ramachandran1,2, Yoichiro Mitsuishi1,2, Gavin Ha1,2, Juliann Shih2, Francisca Vazquez2, Aviad Tsherniak2, Alison M. Taylor1,2, Jin Zhou1, Zhong Wu1, Ashton C. Berger2, Marios Giannakis1,2, William C. Hahn1,2,3, Andrew D. Cherniack1,2, and Matthew Meyerson1,2,3,4
ABSTRACT

The Krüppel-like family of transcription factors plays critical roles in human development and is associated with cancer pathogenesis. Krüppel-like factor 5 gene (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: (i) Focal amplification of superenhancers activates KLF5 expression in squamous cell carcinomas; (ii) Missense mutations disrupt KLF5–FBXW7 interactions to increase KLF5 protein stability in colorectal cancer; (iii) 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.

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. Cancer Discov; 8(1); 1–18. © 2017 AACR.

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 superenhancers adjacent to oncogenes and activate their expression, as observed with MYC, MYB, and ERG (7–12); whereas amplification of noncoding superenhancers is 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 superenhancers to activate ESRI 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 superenhancers 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 (ZNF) DNA binding domain of KLF5 in lung adenocarcinomas, lung squamous cell carcinomas (LUSC), and in a phospho-degron domain of KLF5 in colorectal carcinomas (22, 23).

Krüppel-like transcription factors (KLF) 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 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 patients with breast cancer (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 Superenhancers Activates KLF5 Expression

To define the prevalence of KLF5 superenhancer 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 HNSC (15/522), including esophageal carcinomas (ESCA; 7/184), cervical squamous cell carcinomas (CESC; 14/295), LUSC (14/501), bladder carcinomas (BLCA; 12/408), stomach adenocarcinomas (STAD; 7/441), and colorectal adenocarcinomas (5/615; Fig. 1A). Consistent with these observations, an analysis of SNP array–based copy-number data from 1,043 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 7 cancer types reported above (Supplementary Fig. 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 HNSC 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 (Fig. 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 (Supplementary Fig. S1B). Chromatin immunoprecipitation sequencing (ChIP-seq) in cell lines representing eight HNSC, three ESCA, and three STAD showed similar profiles of histone H3 lysine 27 acetylation (H3K27ac) signal enrichment, indicating that superenhancers, distinguished by thin bars, are called by the ROSE pipeline (39–41) based on the H3K27ac signal enrichment. BLCA, bladder carcinoma; CESC, cervical squamous cell carcinoma; CRC, colorectal adenocarcinoma; ESCA, esophageal carcinoma; GISTIC, genomic identification of significant targets in cancer; STAD, stomach adenocarcinoma; TCGA, The Cancer Genome Atlas.
Oncogenic Activation of KLF5 by Diverse Somatic Alterations

KLF12

The amplified superenhancers are located in a ∼600 kb noncoding region flanked by KLF5 on the centromeric side and KLF12 on the telomeric side (Fig. 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 measure physical interactions between chromatin regions and define TADs in the genome (49), we found that the amplified superenhancers 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 but not the promoter or complete gene body of KLF12 (Fig. 2A and B). Indeed, a recent study in a stomach adenocarcinoma cell line identified significant chromatin 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 superenhancers in the amplified region (Fig. 1C, rectangle), one in ESCA cells and two in HNSC and STAD cells (Fig. 1C, indicated by bars within the rectangle). Taken together, these data suggest that focal amplification of noncoding superenhancers near the KLF5 gene is a recurrent event in multiple cancer types, particularly squamous cell carcinomas.

The amplified superenhancers are focally amplified (Supplementary Fig. 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 superenhancers in BICR31 cells (Fig. 2B). We applied the CRISPR-mediated repression system (15, 57, 58), which uses a short guide RNA (sgRNA) to recruit inactivated Cas9 to repress the 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 (Supplementary Fig. S2B).

A Combination of Three Individual Enhancers within the Amplified Superenhancers Drives KLF5 Overexpression

We selected the HNSC cell line BICR31, in which the KLF5 superenhancers are focally amplified (Supplementary Fig. 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 superenhancers in BICR31 cells (Fig. 2B). We applied the CRISPR-mediated repression system (15, 57, 58), which uses a short guide RNA (sgRNA) to recruit inactivated Cas9 to repress the 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 (Fig. 2D). In agreement, 3C assays detected stronger physical interaction between the KLF5 promoter and the e1, e3, and e4 enhancers, compared with the e2 enhancer (Fig. 2C, lower panel). Repression of the individual enhancers did not affect expression of KLF12, present outside the TAD domain, or of

Figure 2. The focally amplified superenhancers 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 TADs are indicated as grey bars. B, Four individual enhancers, e1–e4, within the superenhancers are defined by p300 ChIP-seq signal from the HNSC cell line BICR31. (continued on next page)
PIBF1 or DIS3 within the same large TAD domain (Supplementary Fig. 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 (Fig. 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 (Fig. 2F), along with a strong reduction (~51%) in KLF5 expression and a modest reduction (~25%) in PIBM1 and DIS3 expression (Fig. 2D; Supplementary Fig. 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, \( P = 10^{-32} \)), with 73.3% distributed across intergenic or intronic regions (Supplementary Fig. 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 KLFs and specificity proteins (Sp; Fig. 3A; refs. 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 (Supplementary Fig. 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 (Fig. 3B). Our results are consistent with previous reports that the transcriptional function of KLF5 depends on its interaction with the CBP/p300 coactivator complex (62). Annotating KLF5 binding sites in more detail, we observed that KLF5 binding sites are more prevalent in superenhancers 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 superenhancers rather than in typical enhancers (∼22%; Fig. 3C).

To investigate the transcriptional impact of KLF5 expression, we conducted RNA-seq assays in BICR31 cells with and without siRNA-mediated silencing of KLF5 (Supplementary Fig. 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: (i) the number of KLF5 binding sites within ±50 kb of the transcription start site (TSS) for each queried gene, and (ii) the distance between these KLF5 binding sites and the TSS. 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 (Fig. 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 (Supplementary Fig. 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 (Fig. 3E; Supplementary Fig. 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 (Fig. 3F).

**Hotspot Mutations in a Phospho-Degron Domain Increase KLF5 Protein Stability**

In addition to focal amplifications of noncoding superenhancers, 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 2 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 (Fig. 4A; Supplementary Fig. S7; see Methods section for additional details on mutation hotspot identification). Three FBXW7 (also known as CDC4) phospho-degron domains (CPD) 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 (71). Our analysis revealed that the second CPD (amino acids 301–307; PPSPPS) is a target of missense mutations, seen mainly in colorectal cancer (7/619; \( P = 5.65 \times 10^{-30} \); data from ref. 23; Fig. 4A; Supplementary Fig. 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 with wild-type (WT) KLF5 (Fig. 4B). Immunoprecipitation assays confirmed that the mutations impaired the interaction of KLF5 with FBXW7 (Fig. 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 (refs. 72, 73; Supplementary Fig. S8A). None of the colorectal cancer samples harboring KLF5 hotspot mutations had mutations in FBXW7 (Supplementary Fig. 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 (Fig. 4D). Although overexpression of WT FBXW7 in HCT116 cells decreased the protein level of KLF5, the FBXW7 mutants showed an opposite effect (Fig. 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; ref. 64) within the second of three DNA-binding 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 LUSC (7/464; ref. 22; Fig. 5A; Supplementary Fig. S7B). Pan-cancer analysis identified additional hotspot mutations at these positions in CESC (6/272), BLCA (5/398), and STAD (1/383; Fig. 5A). Interestingly, these mutations are cancer type–specific. For example, the E419K mutation occurs predominantly in CESC whereas the E419Q mutation is observed only in lung cancers (Fig. 5A).
**Figure 3.** KLF5 activates cell identity genes and cancer-related genes in HNSC 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 superenhancers called from H3K27ac ChIP-seq signals merged from eight HNSC cell lines. P value was derived from a Fisher exact test. D, 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, gray, 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 superenhancer-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 superenhancers (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 E; n = 3) reduced ID1 expression. Expression levels are normalized to the control (sg-Ctrl #1). Error bars, SD. 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 ZNF DNA binding domain. Mutations in different cancer types are color coded. 

B. The colorectal cancer HCT116 cell line expressing V5-tagged WT KLF5, or KLF5 P301S, S303P or P304A mutants, was treated with 100 ug/ml 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, SD. P values were derived from student t tests: *, P ≤ 0.05. 

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

D. Coimmunoprecipitation 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, SD. 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 ZNF DNA binding domain. Mutations in different cancer types are color coded. Right, the binding profile of KLF5 WT and mutants. The nucleotide differences in the DNA binding motifs are highlighted by green boxes, and their positions are indicated by red asterisks. B, 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. C, ChIP-seq of V5-KLF5 in HCC95 cells. E419Q unique (5,611 sites) and WT unique (483 sites) binding sites were compared. 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, 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-seq 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 superenhancers 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.
To assess the function of these mutations, we generated N-terminal V5-tagged versions of WT KLF5 and three of the most recurrent mutants, D418N, E419K, and E419Q, and infected them into HEK293T cells (Supplementary Fig. 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 (Fig. 5B). Changes in the cognate DNA binding motifs of KLF5 appeared to be predominantly restricted to nucleotides at the 5th and 6th positions of the DNA motifs (Fig. 5B), consistent with a report that the second ZNF domain of KLF transcription factors recognizes the 4–6th position of the DNA motif (76). The D418N mutant, seen in LUSC and BLCA, preferentially binds to thyridine (T) at the 6th nucleotide in the DNA motif, compared with guanine (G) for WT KLF5 (Fig. 5B). In addition, the E419K mutant, seen mainly in CESC, binds preferentially to G at the 5th nucleotide of the DNA motif, whereas the E419Q mutant, specific to lung cancers, binds preferentially to adenine (A) at the same nucleotide position, compared with cytosine (C) or T for WT KLF5 (Fig. 5B). Accordingly, KLF5 WT and mutant proteins bind to different regions of the genome (Fig. 5B). When KLF5 binding sites are ranked by variability among HEK293T cells overexpressing different WT or mutant constructs, ~44%, 26%, and 15% of the top 10% variable sites are preferentially bound by KLF5 D418N, E419K, and E419Q, respectively (Fig. 5B right panel for overview; Supplementary Fig. S9B for examples).

The KLF5 E419Q Mutant Gains Novel Binding Sites, Creates New Superenhancers, 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 WT for the KLF5 gene based on RNA-seq results from the CCLE project (35). Ectopic expression and ChIP-seq analysis of V5-tagged KLF5 WT and KLF5 E419Q in HCC95 (Supplementary Fig. S10A-B) revealed that both WT 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 (Fig. 5C). Electrophoretic mobility shift assays (EMSA) 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 with WT KLF5 (Supplementary Fig. S10C), consistent with our results in HCC95 cells, above (Fig. 5C). However, like WT KLF5, KLF5 E419Q also binds to DNA motifs with the C and T variants (Supplementary Fig. S10C), which explains the observation that KLF5 E419Q gains more binding sites across the genome, compared with 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 (\( P = 1.3 \times 10^{-49} \)) but less enriched in promoter regions (\( P = 7.2 \times 10^{-48} \)), compared with regions that are shared by KLF5 WT and KLF5 E419Q (Supplementary Fig. 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 with cells overexpressing KLF5 WT (Fig. 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 (Supplementary Fig. S10B). By integrating the results of RNA-seq 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 (Fig. 5E), suggesting a gene activation role for this mutant. Furthermore, the gained binding sites also form novel superenhancers, as defined by H3K27ac enrichment, that are associated with activation of genes such as FOXE1, NAMPT, EPHP3, and GAS6 (Fig. 5F; Supplementary Fig. 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 superenhancer that upregulates FOXE1 expression as measured by RNA-seq (Fig. 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 superenhancer and activated NAMPT expression (Fig. 5F).

The NAMPT gene encodes nicotinamide phosphoribosyltransferase, a rate-limiting enzyme in the biosynthesis of the metabolite nicotinamide adenine dinucleotide (NAD; ref. 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 superenhancers, 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 siRNAs in the HNSC cell line BICR31, in which KLF5 overexpression is driven by the 13q22.1 superenhancer amplification (Fig. 2), resulted in a marked reduction of cell proliferation (Fig. 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 (Fig. 2), also resulted in a significant reduction in proliferation of the BICR31 cell line (Fig. 6A). The proliferation-inhibitory effect of silencing KLF5 can be partially rescued by ectopic expression of ID1 (Fig. 6B), a target gene of KLF5 in head and neck squamous carcinoma cells (Fig. 3E and F). We then investigated the phenotypic outcomes of mutations in KLF5. Overexpression of the KLF5 E419Q mutant identified in lung squamous carcinomas
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 after 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 seeding). Cell number is normalized to the controls (siNC #1 or sg-Ctrl #1). Error bars, SD. P values were derived from t tests: **, P ≤ 0.01. B, Left, immuno-blot 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, SD. 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, SD. 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 A TARiS 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 superenhancer 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 ZNF DNA binding domain of KLF5.
significantly increased proliferation of the LUSC cell line HCC95, compared with KLF5 WT, in low-serum media (Fig. 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 (Fig. 6D). Because 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 superenhancer amplifications and coding mutations in a phospho-degron domain or a DNA binding domain (Fig. 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 WT 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 superenhancers in many squamous cell carcinomas and some adenocarcinomas. In contrast to the MYC locus, in which cancer type–specific superenhancers are amplified (15), the same noncoding region ~300 kb 3′ to KLF5 is amplified in multiple anatomic and histologic 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 superenhancers that drive the activity of the entire superenhancer region (15, 83). In contrast, the activity of the KLF5 superenhancer region is dependent on a combination of three individual enhancers, representing another type of enhancer structure within superenhancers.

In addition to transcriptional regulation, we show that KLF5 is activated at the protein level by missense mutations. KLF5 contains three 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 ZNF 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 ZNF domain that recognizes the 4–6th nucleotide position of the KLF DNA motif, KLF4 is mutated in the first ZNF 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 ZNF 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 WT KLF5 while also maintaining the binding sites of the WT 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 p53 toward novel binding sites while eliminating binding sites recognized by WT p53 (90, 91). This difference may be because, unlike the tumor suppressor p53, WT KLF5 itself is an oncogenic transcription factor and thus losing WT 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 superenhancers 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 germline 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 germline variants will likely uncover the specific mechanisms underlying their pathogenic features.

In addition to somatic genetic alterations, noncoding germline 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–associate 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 superenhancer 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 cancer therapeutics.

**METHODS**

**Pan-Cancer Copy-Number Alteration Analysis**

Genomic identification of significant targets in cancer (GISTIC) 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 2, 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 CCLE project (35) in 2015 and 2016. Cells tested negative for Mycoplasma and were maintained in RPMI-1640 medium supplemented with 10% heat-inactivated FBS and 1% penicillin streptomycin. Cell line identities were verified by RPMI-1640 medium supplemented with 10% heat-inactivated FBS and were maintained in Mycoplasma screening.

**ChIP-seq Analysis**

ChIP-seq assays were performed as previously described (102–106). 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 concubated with mixed Dynabeads A and G (Thermo Scientific). Antibodies that were used include H3K27ac (2 μg per ChIP; Abcam, ab4729), KLF5 (4 μg per ChIP; Abcam, ab137676), p300 (4 μg per ChIP; Thermo Fisher, R960-25), and V5 (4 μg per ChIP; 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 the Burrows-Wheeler Aligner (BWA; refs. 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 (Supplementary Fig. S6), we used the “Dnase 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 was normalized by edgedR pipeline (114, 115) and then log transformed. We performed k-means clustering for the top 10% most variable binding sites. To present the heat map, the normalized binding signal was scaled by rows. For identifying binding sites that are specific to KLF5 WT or E419Q in the LUSC 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 is over 10%, we randomly subsampled the larger sample by Sامتools (116) to normalize the signal. ChIP-seq data were uploaded to the Gene Expression Omnibus (GEO; GSE88976).

**Superenhancer Identification**

For each cancer type, H3K27ac ChIP-seq data from multiple cell lines were merged into one dataset. Based on the merged ChIP-seq results, including the aligned reads and MACS2 binding peaks, we identified superenhancers for each cancer type using the ROSE pipeline (39–41). To identify superenhancers that are gained by KLF5 E419Q bindings, we used Bedtools (113) to compare the superenhancers called from H3K27ac ChIP-seq signal in HCC95 cells over-expressing KLF5 WT and E419Q. We identified the superenhancers 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 superenhancers and the nearest genes are listed in Supplementary Table S1.

**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 superenhancer region. The 3C ligation products were quantified by SYBR Green-based PCR and the primer sequences are listed in Supplementary Table S2.

**CRISPR/Cas9-Mediated Enhancer Repression**

CRISPR/Cas9 sgRNAs were identified using the sgRNA designer tool from the Broad Institute (117) and control, nontargeting sgRNAs were selected from the GeCKOv2 library (118). The enhancer repression vector lent-KRAB–iCas9-blast was generated previously.
Oncogenic Activation of KLF5 by Diverse Somatic Alterations

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 ligated 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 Supplementary Table S2.

siRNA-Directed Gene Silencing

BICR31 cells were transfected with negative control, nontargeting 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. Prevented Silencer Select siRNAs (Thermo Scientific, Negative Control Nos.1 and 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 m 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

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-FLAG-puro and pLenti-EF1a-PKG-blasti, respectively, with or without the V5 tag fused to the N-terminus. Infected HEK293T and HCC95 cells were selected by 2 μg/mL puromycin or 10 μg/mL blastidicin. Overexpression was validated by RT-PCR and immunoblot analysis. Primers used for cloning are listed in Supplementary 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 μg/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 versus 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-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 GEO (GSE88977).

BETA Analysis to Combine ChIP-seq and RNA-seq Results

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 TSSs of each gene, and also based on the number of KLF5 binding sites ±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 upregulated, downregulated, 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 with 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 a Bio-Rad C1000-Touch Real-time PCR instrument. For TaqMan PCR, the following premade S’ nuclease probes were ordered from Integrated DNA technologies: KLF5 (Hs.PT.56a.4028297), KLF12 (Hs.PT.58.28103949), PIBF1 (Hs.PT.58.21509866), DISS (Hs.PT.58.39902044), ID1 (Hs.PT.58.18791272.g), and internal references HPRT1 (Hs.PT.58.45621572; for qPCR signal normalization) and GAPDH (Hs.PT.58.5898810.g). For SYPR green PCR, the primers used are listed in Supplementary Table S2.

CHX Chase Assays

HEK293T cells infected with KLF5 WT, P301S, S303P, or P304A were treated with 100 μg/mL 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.

Comunoprecipitation 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, 150 mmol/L NaCl, 50 mmol/L Tris-HCl pH 8.0) supplemented with protease and phosphotase inhibitor. Antibodies that were used include V5 (4 μg per IP; Thermo Fisher, R960-25) and HA (4 μg per IP; 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 mmol/L DTT before immunoblot analysis.

**EMSAs Assays**

KLF5 WT and E419Q proteins were translated by using the TNT Quick Coupled Transcription/Translation System (Promega LI170). 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 Supplementary Table S2. For the EMSA, the translated KLF5 proteins and the DNA probes were mixed, and incubated with binding reaction buffer (Final concentration: 10 mmol/L Tris-HCL pH 7.5, 50 mmol/L KCl, 2.5 mmol/L DTT, 0.05 mmol/L EDTA, 0.05 μg/μL Poly-dIdC, 0.25% Tween20) for 30 minutes 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.

**Acquisition Codes**

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

**Disclosure of Potential Conflicts of Interest**

J.M. Francis is a senior scientist at Gritstone Oncology and has ownership interest (including patents) in the same. G.F. Gao reports receiving commercial research support from Bayer AG. A.C. Berger reports receiving commercial research support from Bayer AG. A.D. Cherniack reports receiving a commercial research grant from Bayer AG. M. Meyerson reports receiving a commercial research grant from Bayer, has ownership interest in Origimed, and is a consultant/advisory board member for Origimed. No potential conflicts of interest were disclosed by the other authors.

**Authors’ Contributions**

Conception and design: X. Zhang, M. Meyerson Development of methodology: X. Zhang, W.C. Hahn, A.D. Cherniack Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): X. Zhang, F. Vazquez, J. Zhou, Z. Wu, M. Giannakis, W.C. Hahn Analysis and interpretation of data (e.g., statistical analysis, bio-statistics, computational analysis): X. Zhang, J.M. Francis, G.F. Gao, J.D. Campbell, Y. Mitsushi, G. Ha, J. Shih, F. Vazquez, A. Tsherniak, A.M. Taylor, A.C. Berger, M. Giannakis, A.D. Cherniack Writing, review, and/or revision of the manuscript: X. Zhang, P.S. Choi, J.M. Francis, G.F. Gao, J.D. Campbell, A. Ramachandran, Y. Mitsushi, A. Tsherniak, M. Giannakis, W.C. Hahn, M. Meyerson Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): X. Zhang, P.S. Choi Study supervision: W.C. Hahn, M. Meyerson

**Acknowledgments**

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

**Grant 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 P.S. Choi (1K99CA208028), and the Norman R. Seaman Endowment fund to M. Meyerson. M. Meyerson is an American Cancer Society Research Professor.

Received May 17, 2017; revised September 18, 2017; accepted September 26, 2017; published OnlineFirst September 29, 2017.

**REFERENCES**

Oncogenic Activation of KLF5 by Diverse Somatic Alterations


Oncogenic Activation of KLF5 by Diverse Somatic Alterations


Somatic Superenhancer Duplications and Hotspot Mutations Lead to Oncogenic Activation of the KLF5 Transcription Factor

Xiaoyang Zhang, Peter S. Choi, Joshua M. Francis, et al.

Cancer Discov  Published OnlineFirst September 29, 2017.

Updated version
Access the most recent version of this article at:
doi:10.1158/2159-8290.CD-17-0532

Supplementary Material
Access the most recent supplemental material at:
http://cancerdiscovery.aacrjournals.org/content/suppl/2017/09/29/2159-8290.CD-17-0532.DC1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://cancerdiscovery.aacrjournals.org/content/early/2017/12/14/2159-8290.CD-17-0532. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.