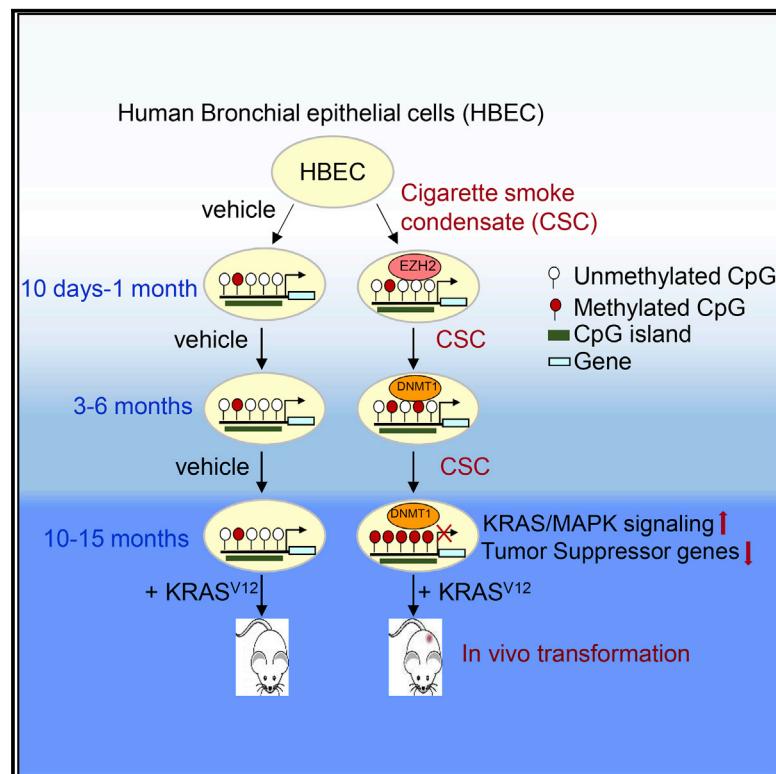


Cancer Cell

Chronic Cigarette Smoke-Induced Epigenomic Changes Precede Sensitization of Bronchial Epithelial Cells to Single-Step Transformation by KRAS Mutations

Graphical Abstract



Authors

Michelle Vaz, Stephen Y. Hwang,
Ioannis Kagiampakis, ...,
Victor E. Velculescu,
Hariharan P. Easwaran,
Stephen B. Baylin

Correspondence

heaswar2@jhmi.edu (H.P.E.),
sbaylin@jhmi.edu (S.B.B.)

In Brief

Vaz et al. show that long-term exposure of untransformed human bronchial epithelial cells to cigarette smoke condensate induces epigenetic changes, consistent with those commonly seen in smoking-related non-small cell lung cancer, that sensitize the cells to transformation with a single KRAS mutation.

Highlights

- Chronic smoke exposure causes sequential chromatin changes leading to gene silencing
- Silenced genes are normally polycomb controlled but adopt abnormal DNA methylation
- Gene methylation causes sequential upregulation of key signal transduction pathways
- Epigenetic alterations sensitize cells to transformation by a single oncogenic event

Chronic Cigarette Smoke-Induced Epigenomic Changes Precede Sensitization of Bronchial Epithelial Cells to Single-Step Transformation by KRAS Mutations

Michelle Vaz,¹ Stephen Y. Hwang,¹ Ioannis Kagiampakis,¹ Jillian Phallen,¹ Ashwini Patil,² Heather M. O'Hagan,^{3,5} Lauren Murphy,¹ Cynthia A. Zahnow,¹ Edward Gabrielson,⁴ Victor E. Velculescu,¹ Hariharan P. Easwaran,^{1,*} and Stephen B. Baylin^{1,6,*}

¹Department of Oncology, The Sidney Kimmel Comprehensive Cancer Center, The Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

²Krieger School of Arts and Sciences, Baltimore, MD 21218, USA

³Medical Sciences, Indiana University School of Medicine, Bloomington, IN 47405, USA

⁴Department of Pathology, The Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

⁵Indiana University Melvin and Bren Simon Cancer Center, Indianapolis, IN 46202, USA

⁶Lead Contact

*Correspondence: heaswar2@jhmi.edu (H.P.E.), sbaylin@jhmi.edu (S.B.B.)

<http://dx.doi.org/10.1016/j.ccr.2017.08.006>

SUMMARY

We define how chronic cigarette smoke-induced time-dependent epigenetic alterations can sensitize human bronchial epithelial cells for transformation by a single oncogene. The smoke-induced chromatin changes include initial repressive polycomb marking of genes, later manifesting abnormal DNA methylation by 10 months. At this time, cells exhibit epithelial-to-mesenchymal changes, anchorage-independent growth, and upregulated RAS/MAPK signaling with silencing of hypermethylated genes, which normally inhibit these pathways and are associated with smoking-related non-small cell lung cancer. These cells, in the absence of any driver gene mutations, now transform by introducing a single KRAS mutation and form adenosquamous lung carcinomas in mice. Thus, epigenetic abnormalities may prime for changing oncogene senescence to addiction for a single key oncogene involved in lung cancer initiation.

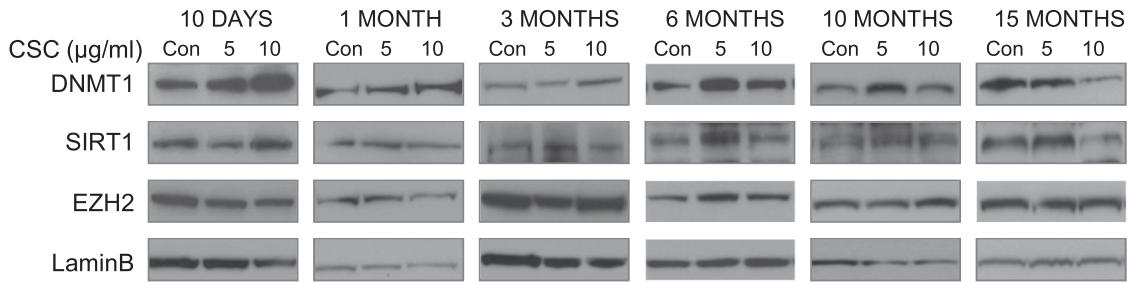
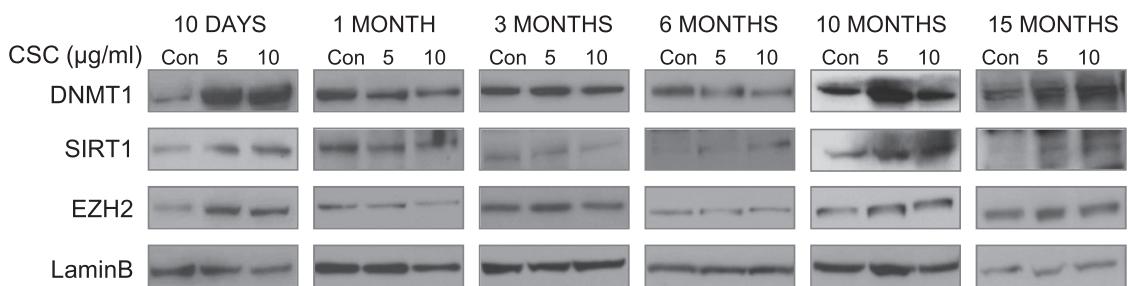
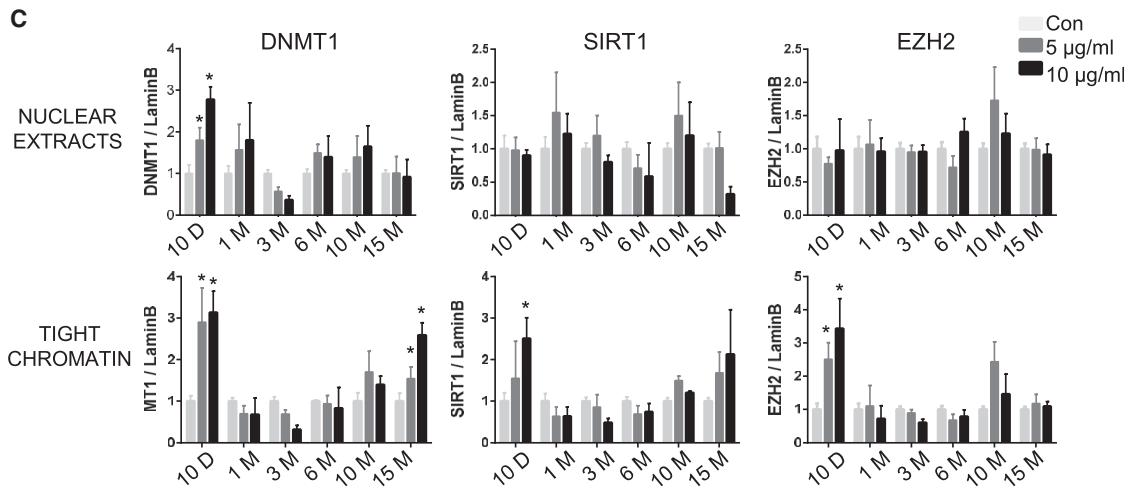
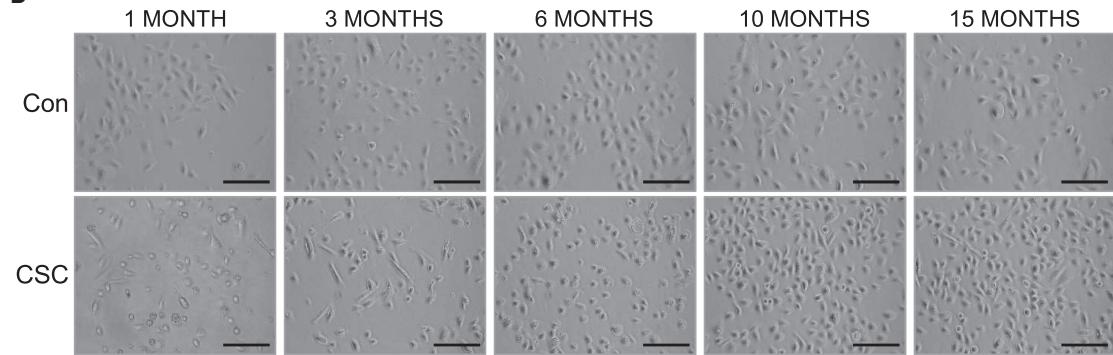
INTRODUCTION

It is well established that chronic exposure to various forms of stress can cause epigenetic as well as genetic alterations ultimately leading to the development of cancer. Cigarette smoke plays a key role in the development of lung cancer, which remains the leading cause of cancer-related deaths worldwide (Torre et al., 2015). The effect of cigarette smoke and its components in contributing to epigenetic changes in lung cancer is well documented (Belinsky et al., 2002; Damiani et al., 2008; Liu et al., 2010; Tellez et al., 2011; Tessema et al., 2014). In addition, a number of mutations seen in lung cancer patients are attributed

to cigarette smoke exposure (Cancer Genome Atlas Research Network, 2012; Govindan et al., 2012). It is now appreciated that these genetic abnormalities exist with epigenetic changes in all human cancers, and both presumably contribute to tumorigenesis through induction of abnormal regulation of multiple key signal transduction pathways (Baylin and Jones, 2011; Jones and Baylin, 2007; Macaluso et al., 2003; Shen and Laird, 2013; You and Jones, 2012). However, the exact order for the evolution of these molecular events and their specific contributions to steps in tumor initiation remains unclear. There are strong suggestions, but little direct evidence, that epigenetic changes might lead to altered regulation of key genes and their

Significance

Cancer development is a complex process involving both genetic and epigenetic abnormalities. However, the evolution of these changes has not been well delineated. We demonstrate that exposure of human bronchial epithelial cells to chronic cigarette smoke causes progressive gene promoter DNA hypermethylation that associates with activation of key signaling pathways driving the tumorigenic process. The methylated gene abnormalities highly overlap with those commonly seen in non-small cell lung cancer. Most importantly, these epigenetic changes may predispose cells to single step transformation mediated by a single oncogene, mutated KRAS, in the absence of other driver genetic changes. This study provides a paradigm in which epigenetic alterations may precede and sensitize for genetic events to drive the development of lung cancer.

A**B****C****D**

(legend on next page)

associated pathways, which then play a seminal role in tumor initiation (Baylin and Ohm, 2006; Suzuki et al., 2004). The direct demonstration of this possibility and the sequential events involved are difficult to study, however, especially in human cells. For the present study, we use human bronchial epithelial cells (HBECs), which are initially immortalized by their having been engineered for overexpression of human telomerase reverse transcriptase (hTERT) and cyclin-dependent kinase 4 (Cdk4) (Ramirez et al., 2004). The latter engineering causes the CDKN2A (p16) tumor-suppressor gene to be expressed at high levels but be unable to perform its normal roles of inhibiting the cell cycle and triggering cell senescence. However, these cells retain an intact p53 checkpoint, remain capable of responding to differentiation signals, are anchorage dependent, and cannot initiate tumor formation in immune-incompetent mice (Delgado et al., 2011; Ramirez et al., 2004). Moreover, they require exogenous expression of three or more driver gene mutations for inducing the above abnormal growth and tumorigenic phenotypes (Sato et al., 2006, 2013). In this context, our present study directly addresses one hypothesis we have put forth for the early role of abnormal epigenetic events in tumor initiation (Easwaran et al., 2014). Namely, these changes could alter signaling to up-regulate pathways downstream of key mutated oncogenes, allowing affected cells to subsequently bypass the normal oncogenic senescence response for the genetic abnormality and rather become addicted to it for tumorigenic effects.

RESULTS

Chronic CSC Exposure Induces DNA Damage-Related Chromatin Binding Changes

Earlier studies have shown that the transcription repressive proteins DNMT1, EZH2, and SIRT1 bind tightly to DNA at sites of DNA damage following induction of DNA double-strand breaks and/or acute oxidative stress (O'Hagan et al., 2008; O'Hagan et al., 2011). We treated HBECs with a commercially available cigarette smoke condensate (CSC) prepared as detailed in **STAR Methods**. CSC concentrations that did not significantly decrease cell viability were selected based on preliminary dose-response curves to define an appropriate concentration for long-term treatment. Treating HBECs with CSC for 10 days, as opposed to DMSO alone, induced chromatin binding of DNMT1, EZH2, and SIRT1 (Figures 1A–1C). While total nuclear protein levels of the maintenance DNA methylation enzyme, DNMT1, increased initially after CSC treatment, the levels decreased by 1 month and remained decreased for up to 15 months. There was no change in total nuclear levels of EZH2 and SIRT1 (Figures 1A and 1C). Importantly, the tight binding for each of the aforementioned proteins increased at 10 days following the acute stress of CSC exposure and then decreased once cells likely began to adapt to this stress (Figures 1B and 1C). The tightness of binding of DNMT1 to chromatin tended to

increase again after 10 and 15 months of chronic CSC exposure (Figures 1B and 1C). The expression of TET2, which is involved in DNA demethylation, is not altered by CSC treatment (Figure S1). Over the time course of treatment, HBECs underwent distinct and progressive morphological changes, defined in more detail later below, consisting of initial cell rounding indicative of cells under stress and, by 10 months, appearance of more easily visualized and flattened cells (Figure 1D).

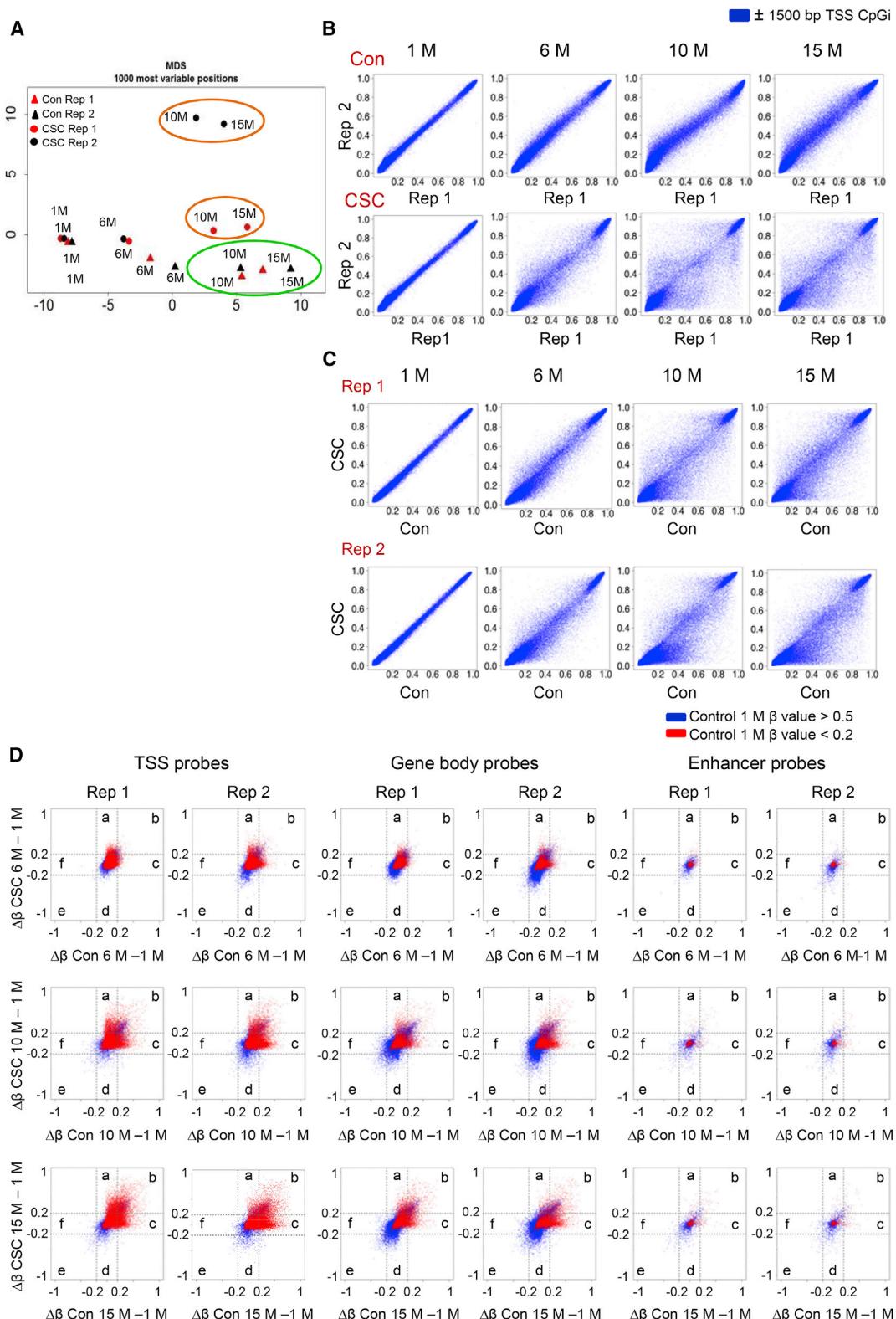
Chronic Exposure to CSC Causes Progressive, Genome-wide Alterations in DNA Methylation

HBEC cells chronically exposed to CSC evolve dramatic and progressive alterations in DNA methylation across the genome. These include, simultaneously, both global losses and more focal methylation gains, with the latter especially involving normally unmethylated promoter CpG islands. These time-dependent dynamics differ starkly from, and are far more extensive than those seen in untreated control cells grown on plastic for the same time periods. Thus, in principal component analysis plots for two independent biological replicates of each, control cells have relatively tightly clustered methylation changes throughout the 15 months of growth while the CSC-treated cells progressively vary significantly from these and also show, especially after 6 months, greater variation between their replicates (Figures 2A and 2B). For the latter cells, these changing methylation patterns suggest many probable stochastic events occurring heterogeneously (Figures 2A–2C). Thus, a time-dependent change in methylation is seen following CSC treatment and stark changes are observed in both replicates following 10 months of exposure (Figure 2C). Methylation changes do occur, but are far fewer and occur in a more coordinated pattern in control cells with little variation between replicates as compared with CSC-treated cells (Figure 2B). The patterns in the CSC-treated cells also indicate that while the replicate patterns diverge continually from each other, and from control cells, the 10- and 15-month time-point patterns for each replicate remain clustered together (orange versus green ellipses, Figure 2A).

At CpG islands within $\pm 1,500$ bp from transcription start sites (TSS), CSC-treated cells have virtually all gains in DNA methylation with time as observed by the increase in the number of low basal methylated probes in section “a” over the time course of the treatment (Figures 2D [red probes] and S2A [red bars]). In contrast, gene body regions predominantly underwent hypomethylation with time in CSC-treated cells as seen by the increase in number of probes in section “d” (Figures 2D [blue probes] and S2A [blue bars]). Putative enhancer regions showed both gains and losses in methylation with a trend toward the latter (Figures 2D and S2A). Methylation changes in the “shore regions,” which lie just outside the TSS-associated CpG islands, resembled changes in both TSS and gene body regions showing both gains and losses of methylation, whereas there were slight losses in

Figure 1. Expression of DNMT1, EZH2, and SIRT1 Following Chronic Exposure to CSC

(A and B) Immunoblots showing expression of DNMT1, SIRT1, and EZH2 in nuclear extracts (A) and tight chromatin fractions (B) obtained from cells treated with DMSO (control [Con]), 5 μ g/mL CSC, and 10 μ g/mL CSC over a period of 15 months at the indicated time points. (C) Quantitation of protein bands using ImageJ. The data presented are the mean of three independent experiments \pm SD. * $p < 0.05$ by two-tailed t test. (D) Phase-contrast photomicrographs showing morphological changes observed in control and CSC-treated (10 μ g/mL) HBECs over time. Scale bar, 100 μ m. See also Figure S1.



(legend on next page)

methylation in the shelf regions lying distal to the shores ([Figure S2B](#)).

Distinct Categories of Genes Become Abnormally

Methylated at Promoters with Chronic Exposure to CSC

Analysis of top variable TSS-associated CpG-island probes with low basal methylation levels in the control cells revealed that they grouped into three categories ([Figure 3A](#)). The first group (group I) contained probes that were slightly methylated in controls over time and further methylated by CSC. These changes strikingly resemble the promoter DNA methylation changes affecting genes in human cancers ([Issa, 2014; Maegawa et al., 2014](#)). The second group (group II) exhibited *de novo* methylation only in CSC-treated cells after 6 months. The third group (group III) gained methylation only in the control cells and remained unmethylated with CSC treatment. To establish the role of the CSC-mediated methylation changes, we obtained a list of CSC-specific methylated genes ([Table S1](#)) from both replicates at 6, 10, and 15 months relative to control cells at the same time points using the criteria detailed in [STAR Methods](#). Further categorization of these genes by gene ontology analysis shows them to be involved in stem cell fates concerned with development, cell fate decisions, and embryonic morphogenesis ([Figure 3B](#)). This is similar to what has been found for the majority of promoter hypermethylated genes in cancer ([Easwaran et al., 2012; Helman et al., 2012](#)). The individual genes fall into the functional categories of homeobox genes and transcription factors as well as genes involved in glycosylation. These gene categories remained constant at all time points, although the number of genes in each category increased with time ([Table S2](#)). Importantly, there is a very significant overlap between these CSC-specific hypermethylated genes and genes methylated in both lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) ([Figure 3C](#) and [Table S2](#)). Additionally these methylation changes are also observed in other oncologic states such as cancers of the breast, colon, and pancreas ([Figure S3A](#)). Genes hypermethylated in these cancers were obtained by selecting probes methylated in at least 25% of tumors and identifying corresponding genes. In a further link, for all the above changes to diseases associated with chronic cigarette smoking there is a moderate, but statistically significant, increase in methylation of our CSC-methylated genes in lung tissues obtained from people with chronic obstructive pulmonary disease and idiopathic lung fibrosis ([Figures S3B](#) and [S3C](#)). Finally, in addition to promoter-associated hypermethylation, CSC exposure also induced much less frequent hypomethylation events in these regions ([Figures 2D, S2A](#), and

[S3D](#)). Categorization of these genes shows them to be involved in cell signaling as well as negative regulation of cell growth ([Figure S3E](#)).

Dynamic Changes in Chromatin Marks and Binding of Chromatin-Associated Proteins Precede DNA Methylation

A building hypothesis for the evolution of promoter DNA hypermethylation in cancer is that the genes involved are primarily those associated with transcription regulation by polycomb group proteins (PcG) and not by promoter DNA methylation in normal tissues. These are typically developmental genes with unmethylated promoter CpG islands that are surrounded by bivalent chromatin, characterized by simultaneous presence of the repressive mark, H3K27me3 and the active mark, H3K4me3 around the TSS ([Azuara et al., 2006; Bernstein et al., 2006](#)) and associated with low basal, but poised, expression in embryonic and adult stem cells. It is hypothesized that during tumor initiation and progression, a quantitative shift occurs away from this chromatin state toward a more repressive one marked by DNA methylation ([Easwaran et al., 2012; Ohm and Baylin, 2007; Schlesinger et al., 2007; Widschwendter et al., 2007](#)). However, there are no data from dynamic models to confirm this. We investigated whether such a series of events occurs during chronic exposure of HBEC cells to CSC.

In initial chromatin immunoprecipitation (ChIP) studies in CSC-exposed versus control cells, we followed enrichment of H3K4me3 and H3K27me3, as well as binding of DNMT1 and EZH2, at proximal promoter CpG-island regions of candidate tumor-suppressor genes (*BMP3, SFRP2, GATA4*) known to be frequently methylated in lung and other cancers. By 10 days of CSC exposure, these genes accumulate bivalent chromatin with H3K4me3 and H3K27me3 marks at their promoters ([Figure 4A](#)). In stark contrast, genes that are not hypermethylated in cancer, including *ACTB*, a housekeeping gene with a CpG-island-containing promoter, and *NANOG*, an embryonic related gene with a CpG-poor promoter, do not develop this change. While the promoter protein binding and chromatin marks studied do not change for the latter two control genes over the entire duration of CSC exposure, *BMP3, SFRP2*, and *GATA4* exhibit initial increases in binding of the enzyme catalyzing H3K27me3, EZH2 ([Figure 4A](#)). Over the time course of CSC exposure and associated with evolution of promoter DNA hypermethylation, there is a quantitative switch from the presence of EZH2 to binding of the maintenance protein for DNA methylation, DNMT1, and a concomitant decrease in levels of H3K27me3 and the active mark, H3K4me3, at the 15-month time point

Figure 2. Progressive Changes in Methylation in Control and CSC-Treated Cells

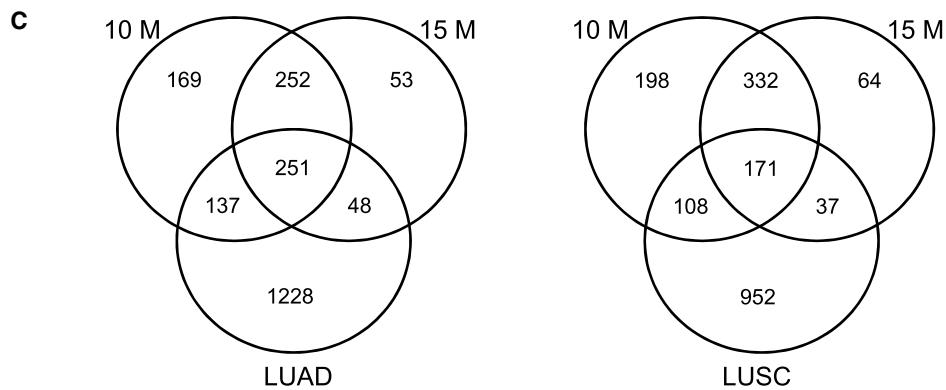
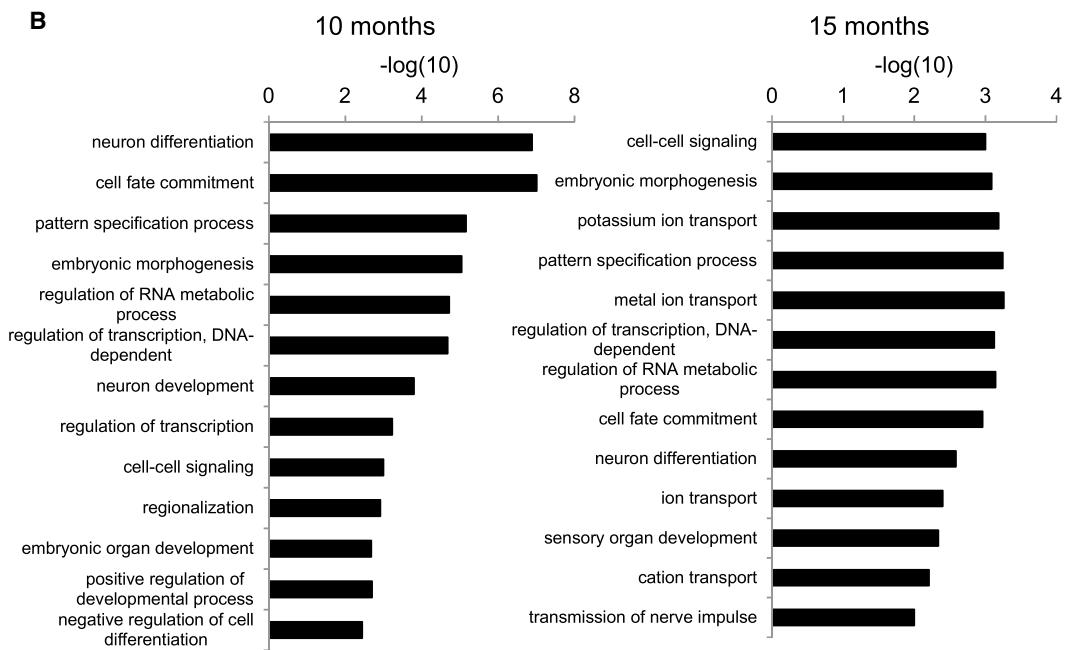
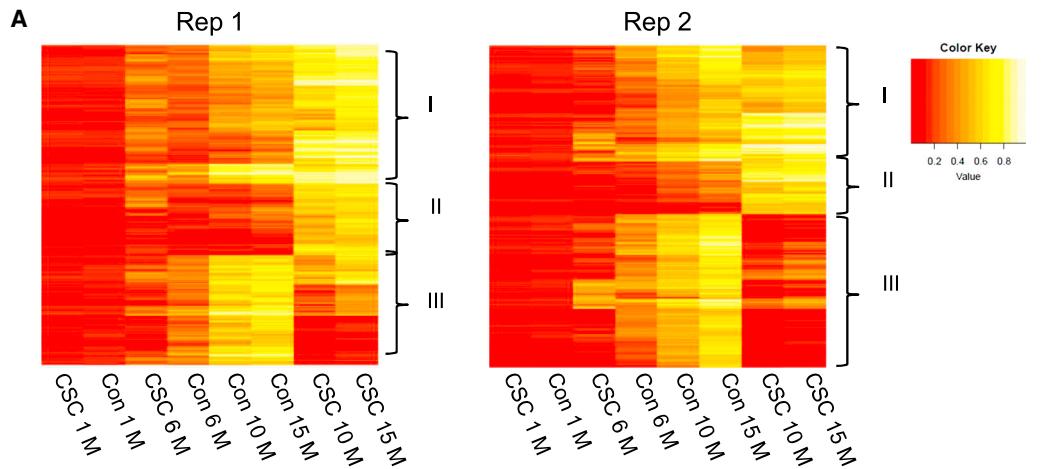
(A) Principal component analysis for control (Con) and 10 µg/mL CSC-exposed cells of the DNA methylation patterns, assayed by Illumina 450K arrays, for most variable probes at examined time points from two replicates in each treatment group. Green (Con) and orange (CSC) ellipses indicate groupings.

(B) Comparison of genome-wide DNA methylation changes, obtained from Illumina 450K arrays, at promoter CpG islands (CpGi) between replicates of the same treatment group.

(C) Comparison of genome-wide DNA methylation, obtained from Illumina 450K arrays, at promoter CpG islands (CpGi) in control (x axis) versus CSC-treated (y axis) cells across the indicated time points in two replicates.

(D) Scatter plots showing changes with time in methylation at promoter, gene body, and enhancer regions. Changes in β values for all CpG-island probes within the respective regions from 1 month (1 M) to the respective endpoint (6 M, 10 M and 15 M) are shown in control (x axis) and CSC-treated (y axis) samples. Red dots: probes with a β value <0.2 in control cells at 1 month; blue dots: probes with a β value >0.5 in control cells at 1 month.

Rep 1, replicate 1; Rep 2, replicate 2. See also [Figure S2](#).



(legend on next page)

(Figure 4A). While expression levels of the DNA hypermethylated genes remain unchanged at initial time points following CSC treatment, their expression decreases following 10 months of CSC treatment (Figure 4B).

The above studies are well reflected in genome-wide ChIP sequencing (ChIP-seq) studies (Figures 4C and S4A–S4G), which assessed enrichment of the aforementioned histone marks/epigenetic proteins at promoter CpG islands of genes (picked from Table S1), which progressively become DNA hypermethylated during chronic CSC exposure. This is best appreciated when these genes are compared with a control set of random CpG-island-containing genes not methylated by CSC. First, these latter genes have a much higher average basal expression (Figure S4H) and, importantly, a distinct, clear zone at the TSS indicating lack of DNMT1 binding in this region (Figure 4C, bottom right). In contrast, the genes that are hypermethylated by CSC have a much lower average basal expression (Figure S4H) and, under both control and CSC exposure conditions, DNMT1 binding encroaches upon the TSS region (Figure 4C). The levels of DNMT1 around the TSS in these genes initially decreases during acute CSC exposure (10 days) compared with the control cells, but later increases over time in the CSC-exposed cells as seen in the plots showing average read counts of DNMT1 enrichment in CSC-methylated genes (Figure 4C, bottom right). This initial decrease in DNMT1 levels may reflect an opening of the chromatin as the acute CSC exposure induces an increase in promoter bivalency as seen by the increase in the H3K4me3 mark in CSC cells at 10 days (Figure 4C, top left), described further below.

Similar to the chromatin changes observed in ChIP studies for candidate genes, ChIP-seq analysis of CSC-methylated genes at 10 days also showed increases in EZH2 (Figure 4C, top right) and H3K4me3 (Figure 4C, top left) around the TSS compared with the control cells. The bivalent nature of the CSC-methylated genes is also well appreciated when these genes are compared with the control set of random unmethylated genes, which have far higher levels of H3K4me3 and virtually absent H3K27me3 (Figure 4C, top and bottom left). By 10 months the CSC-methylated genes, despite maintaining distinct H3K27me3 marks in both control and CSC-exposed cells, have far lower levels of EZH2 than in control cells, which show a sharp increase in EZH2 levels with time (Figures 4C [top right] and S4A–S4G). All of these results indicate that chronic CSC exposure induces a dynamic, time-dependent switch from promoter Pcg control to promoter DNA methylation for a group of low-expression genes with initial promoter region bivalent chromatin. Interestingly, the increased EZH2 in the promoter region for these genes seen in control cells may indicate that their low expression controlled by Pcg may facilitate long-term growth as immortalized cells.

The ramifications of these changes are evident, as outlined below in gene expression and phenotypic differences between control and CSC-exposed cells.

Progressive Changes in Oncogenic Pathway Signaling Are Induced by CSC Exposure in the Absence of Driver Gene Mutations

The epigenetic changes chronicled above accompany CSC-induced alterations of signaling pathways. Initially, at 10 days and 1 month of CSC exposure, upregulation for cell-cycle and DNA replication events dominated followed by pathways for repair of DNA damage and extracellular environment interactions (Figures 5A and 5B). These changes may reflect adaptation of exposed cells to survive the initial stress induced by CSC treatment. Subsequently, with continued CSC exposure and consistent with the chromatin changes outlined earlier, there are switches in dominant pathway signaling at 3 and 6 months after CSC exposure, including downregulation of gene targets of the Pcg proteins BMI1 and EZH2 (Figures 5A and 5B). Thereafter, when abnormal DNA methylation is the dominant epigenetic change by 10–15 months of CSC treatment, top pathway signaling switches to activation of key potential transformation-promoting pathways. Thus, there is strong induction of KRAS and its downstream MAPK signaling, followed by WNT and EGFR signaling (Figures 5A, 5B, and S5). Importantly, exome sequencing at these time points reveals no mutations for key oncogenic driver genes, as discussed later (Table S3).

Key gene hypermethylation changes are linked to the aforementioned upregulation of the KRAS and WNT pathways (Tables 1 and S2), and importantly, these genes are also frequently hypermethylated, as mentioned earlier, in LUAD and LUSC (Figure 3C and Table S2). First, is BMP3, a member of the transforming growth factor (TGF- β) superfamily, which normally represses RAS-MAPK signaling (Kraunz et al., 2005). This gene is very frequently promoter hypermethylated in non-small cell lung cancer (NSCLC), and these events highly correlate with the presence of KRAS mutations (Dai et al., 2001; Kraunz et al., 2005). Second, CSC-related hypermethylation of a series of genes, including MSX1, WIF1, and associated anti-WNT genes, have important functional correlates (Table 1). In neural cells, MSX1 counters the upregulation of the WNT pathway (Revet et al., 2008, 2010), the other leading oncogenic pathway upregulated in our studies (Figures 5 and S5). MSX1 induces at least four different WNT pathway inhibitor genes: Dickkopf 1–3 (DKK1–3) and secreted frizzled-related protein 1 (SFRP1) and SFRP2 (Revet et al., 2010). These are all also hypermethylated in the late CSC-exposed cells and also frequently in primary lung cancers.

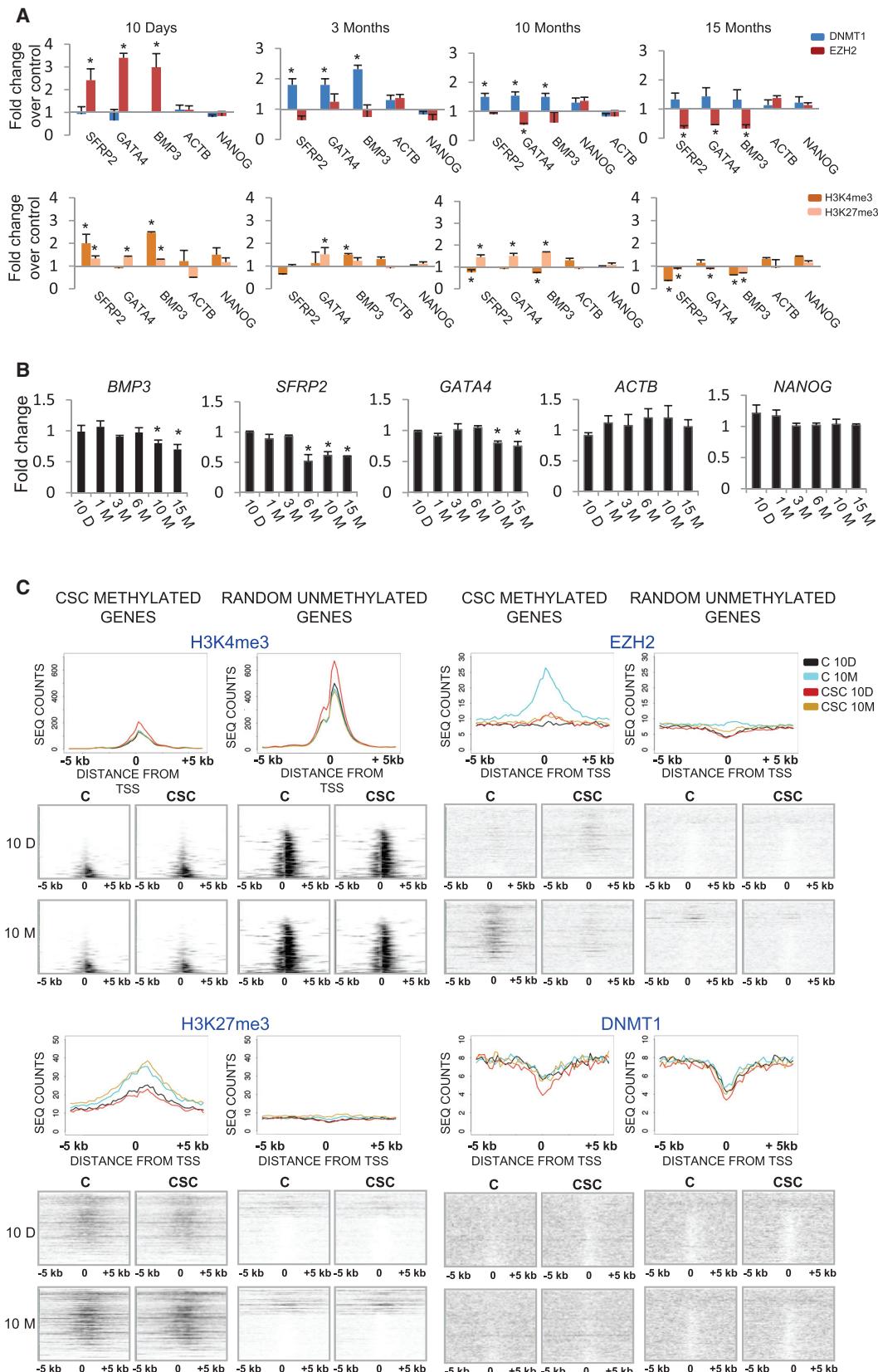
Figure 3. Identification and Characterization of Genes Methylated upon CSC Exposure

(A) Heatmaps of β values for Infinium probes in CpG islands within $\pm 1,500$ bp from TSS in control (Con) and 10 μ g/mL CSC-exposed cells from 1 to 15 months in replicates 1 and 2. Only probes with a starting β value of <0.2 in control cells at 1 month were selected. Three clusters of probes are identified: group I, methylated in controls and further methylated in CSC-treated cells; group II, methylated only in CSC-treated cells; group III, methylated only in controls.

(B) Top enriched biological processes in genes methylated in CSC-treated cells at 10 and 15 months of treatment.

(C) Venn diagrams showing overlap of methylated genes in cells treated with 10 μ g/mL CSC at 10 and 15 months with genes hypermethylated in lung adenocarcinoma (LUAD) and lung squamous cell carcinomas (LUSC). To obtain genes hypermethylated in LUAD and LUSC, we picked probes methylated in at least 25% of tumor samples and identified corresponding genes as described in STAR Methods.

M, months. See also Figure S3; Tables S1 and S2.



(legend on next page)

CSC Exposure Induces Time-Dependent Development of EMT, Anchorage-Independent Growth, and Susceptibility to KRAS-Induced Transformation

Tracking with the progression of epigenetic changes induced by CSC, the HBEC cells manifest key changes suggesting a drift toward a pre-tumorigenic state. First, by 10–15 months the cells become anchorage independent and form clones in soft agar (Figures 6A and 6B). At this time point, when these CSC-exposed cells are grown in medium containing 5% fetal bovine serum they become more elongated, suggestive of an epithelial-to-mesenchymal (EMT)-like phenotype (Figure 6C), and this is accompanied by a stark increase in vimentin, a key protein upregulated in EMT, and a decrease in E-cadherin (Figures 6D and 6E). However, despite these changes, neither the CSC-exposed cells nor the isolated soft agar clones could form tumors *in vivo*.

We hypothesized that this progressive development of the epigenetic, signaling pathway and morphogenic changes induced by CSC exposure might sensitize the cells to transformation by a single oncogenic event. This would be highly significant, since earlier studies have shown that complete transformation of HBECs requires introduction of at least three genetic events: short hairpin RNA-mediated downregulation of TP53 paired with introduction of both KRAS mutations and c-MYC overexpression (Sato et al., 2013). We introduced individually either KRAS^{V12} (Figure 6F) or downregulated the tumor suppressor TP53 (Figure 6G) in 6- and 15-month CSC-treated cells. Only KRAS^{V12} alters the 15-month CSC-treated cells, while neither affects the 6-month CSC-exposed cells or their respective controls. There is a dramatic increase in anchorage-independent growth *in vitro* with more numerous, enlarged, and irregular colonies in CSC-exposed cells expressing mutant KRAS (Figures 6H and 6I). These cells continue to have increased KRAS and WNT signaling now joined by significant enrichment of mTOR (mammalian target of rapamycin), Hedgehog, and VEGF (vascular endothelial growth factor) signaling pathways (Figure 6J). Most important, with exogenous expression of KRAS^{V12} but not TP53 downregulation, the CSC-treated cells are now transformed. Neither of the genetic manipulations transformed control cells. KRAS^{V12}-expressing 15-month CSC-exposed cells from both replicates induce tumors at different rates in mice, which begin appearing 3 weeks post injection and grow to an average size ranging from 500 to 900 mm³ (Figures 6K and 6L). Expression of mutant KRAS^{V12} was detected in the mouse xenografts but not in parental CSC cells transfected with empty vector (Figure 7A). Tumor histology suggests the tumors have an adenosquamous phenotype with sheets of poorly differentiated cells and focal areas of both squamous and glandular differentiation (Figure 7B). The squamous differentiation

was confirmed by positive staining for p63 and cytokeratin (CK) 5/6. Although tumor cells stained negative for TTF1 (NKX2-1), a commonly used marker for pulmonary adenocarcinoma, these were positive for CK7, which is more commonly expressed in human pulmonary adenocarcinoma than squamous cell carcinoma (Camilo et al., 2006). Areas of tumor with glandular differentiation stained for mucin, further indicating their mixed adeno- and squamous differentiation.

Methylation analysis of control and CSC cells following KRAS^{V12} expression, as well as in xenografts, showed in both replicates that there is maintenance of and slight increases in the abnormal promoter methylation of the genes classified as CSC-methylated at 15 months (Table S1; Figures S6A and S6B). Moreover, analysis of methylation at promoters of a key set of CSC-methylated genes (Table S2) revealed an even greater increase in promoter methylation of these genes, accompanied by a decrease in their expression in the xenografts (Figures 7C, 7D, S6C, and S6D). Gene expression analysis in xenografts showed enrichment of KRAS, WNT, and Hedgehog pathways as observed earlier in the CSC treated cells expressing KRAS^{V12} (Figure 7E). In addition, upregulation of LEF1 signaling was also observed in the xenografts suggesting an important role of the WNT pathway in driving tumor formation, as WNT signaling through LEF1 mediates metastasis of LUAD (Nguyen et al., 2009). These results all underscore that stress-induced epigenetic alterations induce signaling events that prime cells for transformation and oncogenic addiction, rather than induction of senescence by a single oncogene. In this regard, whole exome sequencing of >20,000 genes in the parental HBECs, control, and CSC-treated cells at 15 months and tumor xenografts, with an average sequence coverage within the target regions of nearly 150-fold for each sample, reveals the absence of well-known mutated driver genes for NSCLC and other cancers except for our introduced KRAS^{V12} mutation (Table S3). While there was a median of 72 mutations in the 15-month CSC-treated cells, 114 in the xenografts, and 62 in the 15-month DMSO-treated control cells that were not present in the parental HBECs, these all likely reflect passenger mutations that arose in the cells prior to xenograft formation. To exclude the presence of rare clones with driver mutations, we also performed targeted deep sequencing of three xenografts with >600-fold coverage at each base using a 123-gene panel (Table S3-targeted genes) and did not find any additional mutations not previously detected by exome sequencing (Table S3). We observed in the exome and targeted analyses alterations in JAK1 and CREBBP genes, but mutations at the affected amino acid residues are rare in all cancers and not likely to have biological consequences.

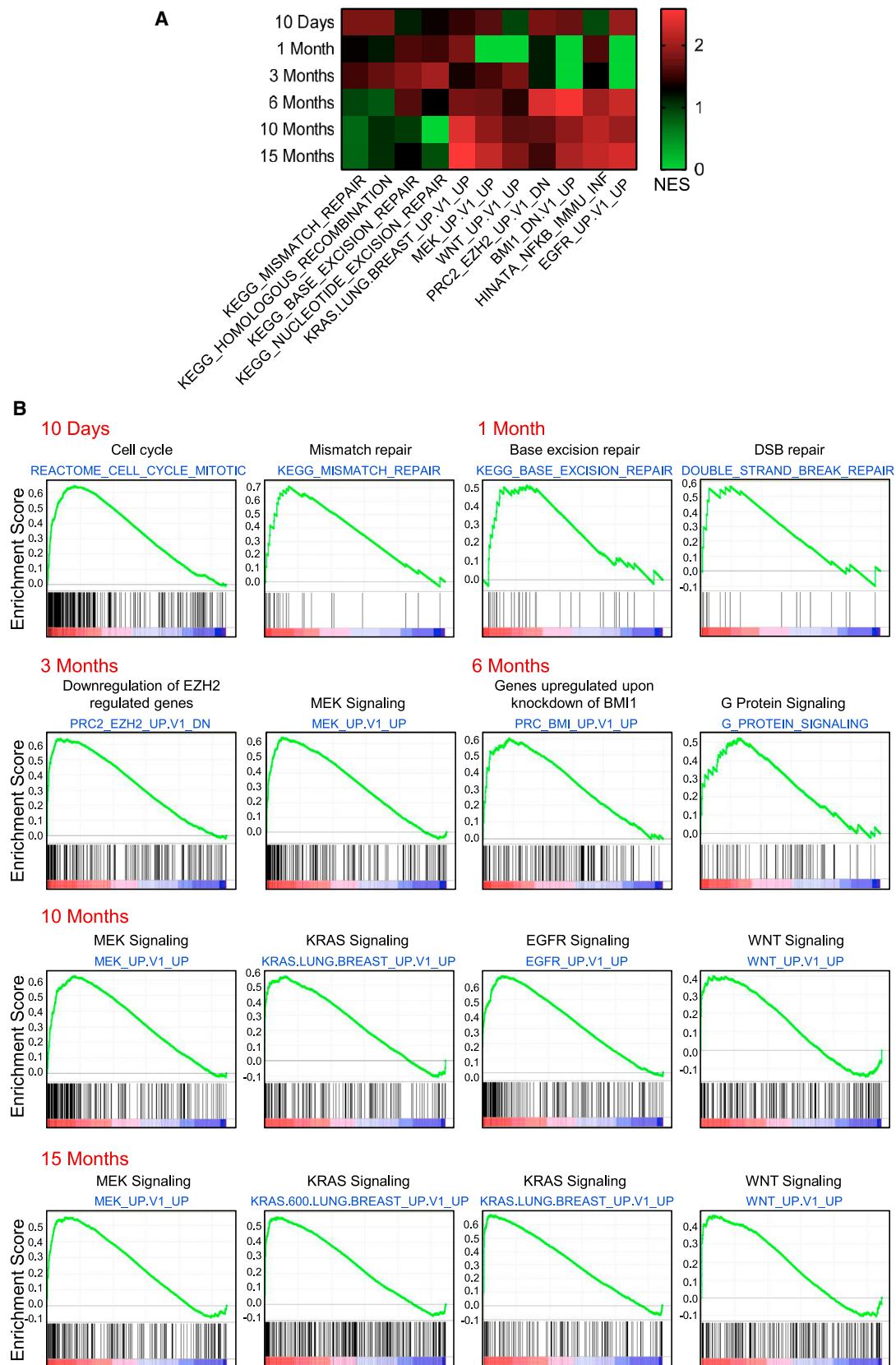
Figure 4. Progressive Changes in Chromatin Marks and Binding of Chromatin-Associated Proteins in CSC-Exposed Cells

(A) ChIP studies examining enrichment of H3K4me3 and H3K27me3, and binding of EZH2 and DNMT1, at representative genes in control and CSC-treated cells over time. Cells were treated with DMSO or 10 µg/mL CSC for the indicated time periods before processing for ChIP. Data are presented as mean ± SD (n = 3). *p < 0.05 by two-tailed t test.

(B) Gene expression of representative genes at the indicated time points following treatment with 10 µg/mL CSC. Values indicate fold change in CSC-treated over control cells. Data are presented as mean ± SD (n = 3). *p < 0.05 by two-tailed t test.

(C) Genome-wide ChIP-seq analysis in control (C) and CSC-treated cells at 10 days (10 D) and 10 months (10 M) post onset of CSC exposure showing enrichment of EZH2, DNMT1, H3K4me3, and H3K27me3 at promoters of genes methylated in CSC-treated cells (CSC-methylated genes) and a random set of unmethylated CpG-island-containing genes (random unmethylated genes). On x axis, 0 = TSS. For each mark analyzed, a graph showing quantitative data depicting the average read counts ±5 kb from the TSS for all genes in the respective groups is shown above the maps of read intensities for individual genes at 10 D and 10 M in control and CSC cells.

See also Figure S4.



(legend on next page)

Table 1. Candidate Genes with a Driver Function

Gene Name	Function
BMP3	member of the TGF- β superfamily and represses KRAS/MAPK signaling in lung cancer; inactivation is an early and frequent event in cancer development
WIF1	suppresses p38MAPK signaling, inhibits WNT pathway activation, suppresses stemness, and induces cellular senescence in salivary gland tumors
MSX1	essential for apoptotic function of intact p53; induces the WNT pathway antagonist genes DKK1, DKK2, DKK3, and SFRP1 in neuroblastoma cells
SFRP2	negative regulator of WNT signaling
SFRP5	negative regulator of WNT signaling
GATA4	zinc-finger transcription factor that regulates genes involved in embryogenesis; involved in G-protein-coupled receptor signaling pathway and is regulated by MAPK/ERK signaling
SLIT2	attenuated during lung cancer progression; deregulates β -catenin and E-cadherin and is associated with poor prognosis
GALR1	methylated in lung cancer; identified as a tumor suppressor

Candidate CSC-methylated genes with a driver function for the signaling events modulated by CSC treatment and which are frequently methylated in lung cancer. See also [Table S2](#).

CSC-Induced Methylation Changes Associate with Smoking-Related Non-Small Cell Lung Cancers with a Bias towards Those with KRAS Mutations

The aforementioned bias to KRAS-mediated transformation raises interesting questions with respect to smoking and the mutation distributions among subgroups of primary NSCLC. First, analyses of TCGA (The Cancer Genome Atlas) data from LUADs ([Cancer Genome Atlas Research Network, 2014](#)) suggest that the epigenetic changes in our model may well sensitize bronchial cells in humans to smoking-related tumorigenesis, which accounts for nearly 80%–90% of NSCLC. Lower methylation of our CSC-specific methylated genes are present in LUADs with EGFR mutations compared with tumors with KRAS or TP53 mutations ([Figure 7F](#)). It has been shown that EGFR mutant tumors very frequently occur in non-smokers, unlike KRAS and TP53 mutated tumors, which frequently occur in smokers ([Boch et al., 2013](#)). Second, consistent with the specificity for KRAS mutations outlined in the previous section, these methylation changes are higher in matched normal lung samples from KRAS mutated tumors than matched normal samples from TP53 mutated tumors ([Figure 7F](#)). Third, although the methylation of our genes are comparable between TP53 and KRAS mutant tumors, we see interesting differences between duration of smoking and the status of smokers (current versus reformed). Our CSC-specific methylated genes are more highly methylated in primary lung tumors from current smokers as compared with tumors from non-smokers or reformed smokers ([Figure 7G](#)). However, particularly intriguing is that, in tumors from patients who were reformed smokers for 15 years or more, there are starkly lower methylation levels of our CSC-specific hypermethylated genes in TP53 mutant tu-

mors versus those with KRAS mutations. This may indicate that the presence of a KRAS mutation leads to more maintenance of methylation for our CSC-methylated genes than for tumors harboring TP53 mutations without KRAS mutations. We hypothesize that although the changes seen in our model may correlate with smoking-induced sensitization to multiple mutations in all NSCLC, TP53 mutations may need to occur with additional genetic alterations. KRAS mutations, on the other hand, may more readily accomplish one-step transformation in the setting of the epigenetic and signaling changes we observe with CSC exposure.

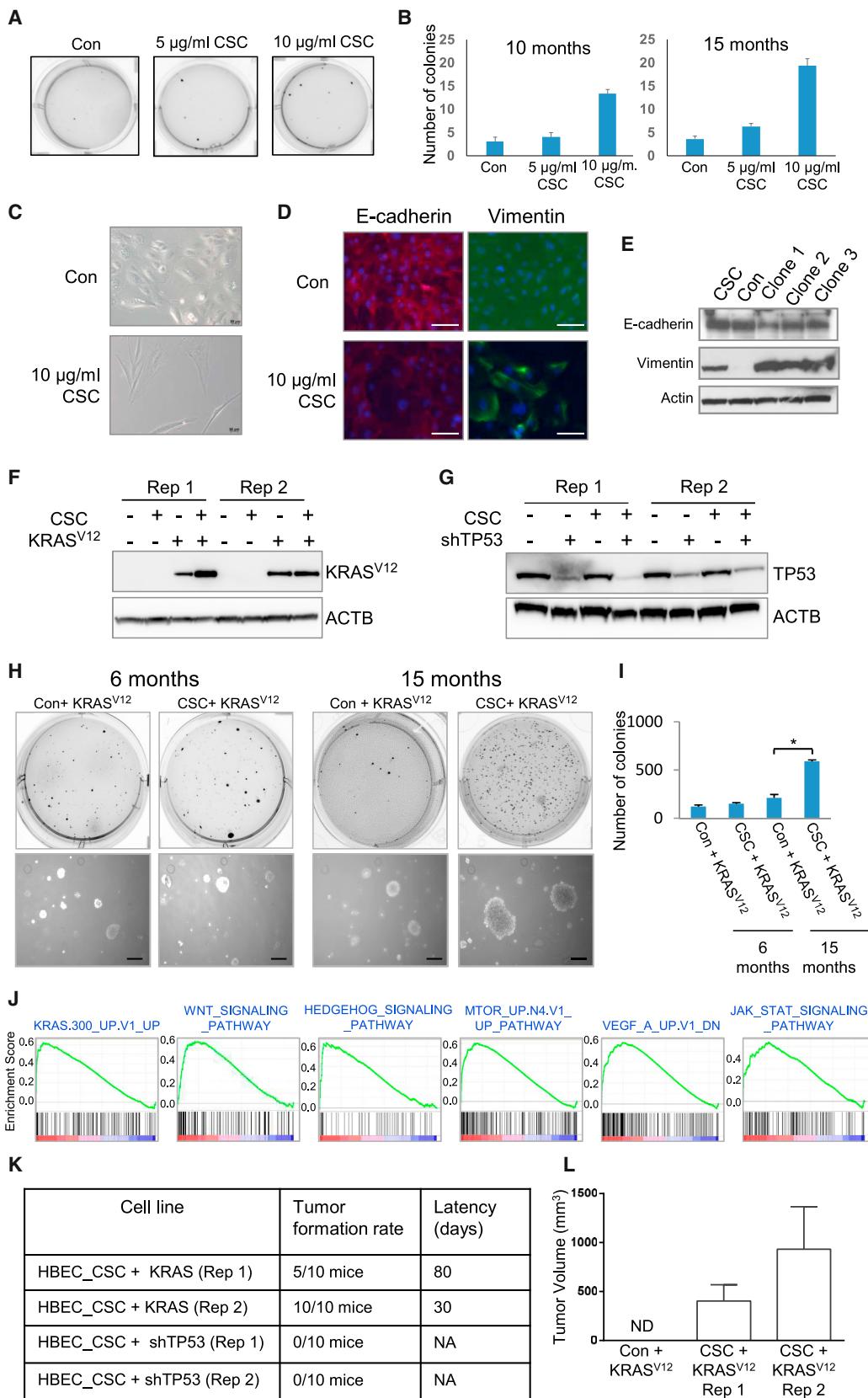
DISCUSSION

We have derived data suggesting how a progressive increase in epigenetic abnormalities during chronic cigarette smoke exposure may drive the initiation of lung cancer. This exposure pushes cells to more stem cell-like chromatin states associated with signaling events sensitizing the cells to oncogene-induced transformation. In this regard, epigenetic abnormalities now recognized in scenarios of chronic cellular stress, inflammation, and aging are suggested, but not proven, to drive early stages for cancer development ([Issa, 2014](#); [Maegawa et al., 2014](#); [Niwa and Ushijima, 2010](#)). Defining how these events interact with key genetic alterations, such as KRAS, is an important goal for cancer research. Our current data are consistent with our previous hypothesis that the types of epigenetic abnormalities observed in the present study could provide an “epigenetic substrate” allowing cells to become addicted, rather than senescent or dying, in responding to signaling from mutated driver genes such as KRAS ([Baylin, 2012](#); [Dimauro and David, 2010](#)).

Figure 5. Gene Set Enrichment Analysis of Genome-Wide Gene Expression Following CSC Exposure

(A) Genome-wide gene expression analysis of control and 10 μ g/ml CSC treated cells at the time points indicated over a period of 15 months. Ranked list of fold changes in gene expression in CSC-treated cells as compared with matched controls were calculated and analyzed using gene set enrichment analysis (GSEA). The top 20 enriched pathways were compared across the time points. Heatmap shows changes in enrichment (based on normalized enrichment score [NES]) of key pathways at various times following CSC exposure.

(B) Gene sets enriched in CSC-treated cells at the indicated time points. Log fold change in gene expression in CSC over control cells was determined and ranked lists were analyzed using GSEA (false discovery rate <0.25, p value <0.01). See also [Figure S5](#).



(legend on next page)

One key implication of our data is that smoke-induced stress results in a more stem-like state in the HBECs, facilitating transformation by a key oncogenic event. Other studies report that chronic stress could dedifferentiate mature cells to a more stem-like state (Lee et al., 2015; Stange et al., 2013; Tata et al., 2013). It has been observed that p53 or BRAF can exhibit their tumor-driving influence only in an epigenetically controlled stem cell state (Kaufman et al., 2016). Thus, our early CSC-induced epigenetic changes and other environmental factors could induce cell signaling pathways that cooperate with gene mutations in driving tumorigenesis.

The nature of the observed epigenetic abnormalities is very consistent with key elements of the above hypotheses. The initial chromatin changes, specifically tightening of DNMT1 and EZH2 to chromatin following a 10-day exposure to a low dose of CSC, are of the type we have previously associated with double-strand DNA breaks and/or reactive oxygen species exposure (O'Hagan et al., 2008; O'Hagan et al., 2011). The decrease in nuclear DNMT1 corresponds with previously published data (Liu et al., 2010), but our results emphasize that it is tight binding to chromatin that is central to the cell stress scenario. This places DNMT1 strategically for mediating maintenance of abnormal DNA methylation induced by chronic CSC exposure. The increased DNMT1 binding to chromatin at 10 days post exposure and no change over the next 3–6 months, followed by an increase after 10 months, is consistent with events for how the abnormal gene promoter DNA methylation evolves, and suggests possible clonal selection of cells with this abnormal methylation during CSC exposure.

Several key features of our data link lung cancer and the CSC-induced abnormally DNA-methylated genes. First, the genes overlap significantly with those commonly hypermethylated in NSCLC. Second, the increased methylation levels of these genes in adenocarcinomas from smokers versus reformed or non-smokers suggest their key role in the development of smoking-associated lung cancer. Third, as discussed earlier, our ChIP-seq studies are in line with hypothesized events for how abnormal promoter DNA methylation may reflect a molecular progression from polycomb to DNA methylation control. The involved genes are heavily biased to developmental and stem cell control, which appear most vulnerable to evolve abnormal promoter DNA methylation during cancer initiation and progres-

sion (Helman et al., 2012; Easwaran et al., 2012; Ohm and Baylin, 2007; Schlesinger et al., 2007; Widschwendter et al., 2007).

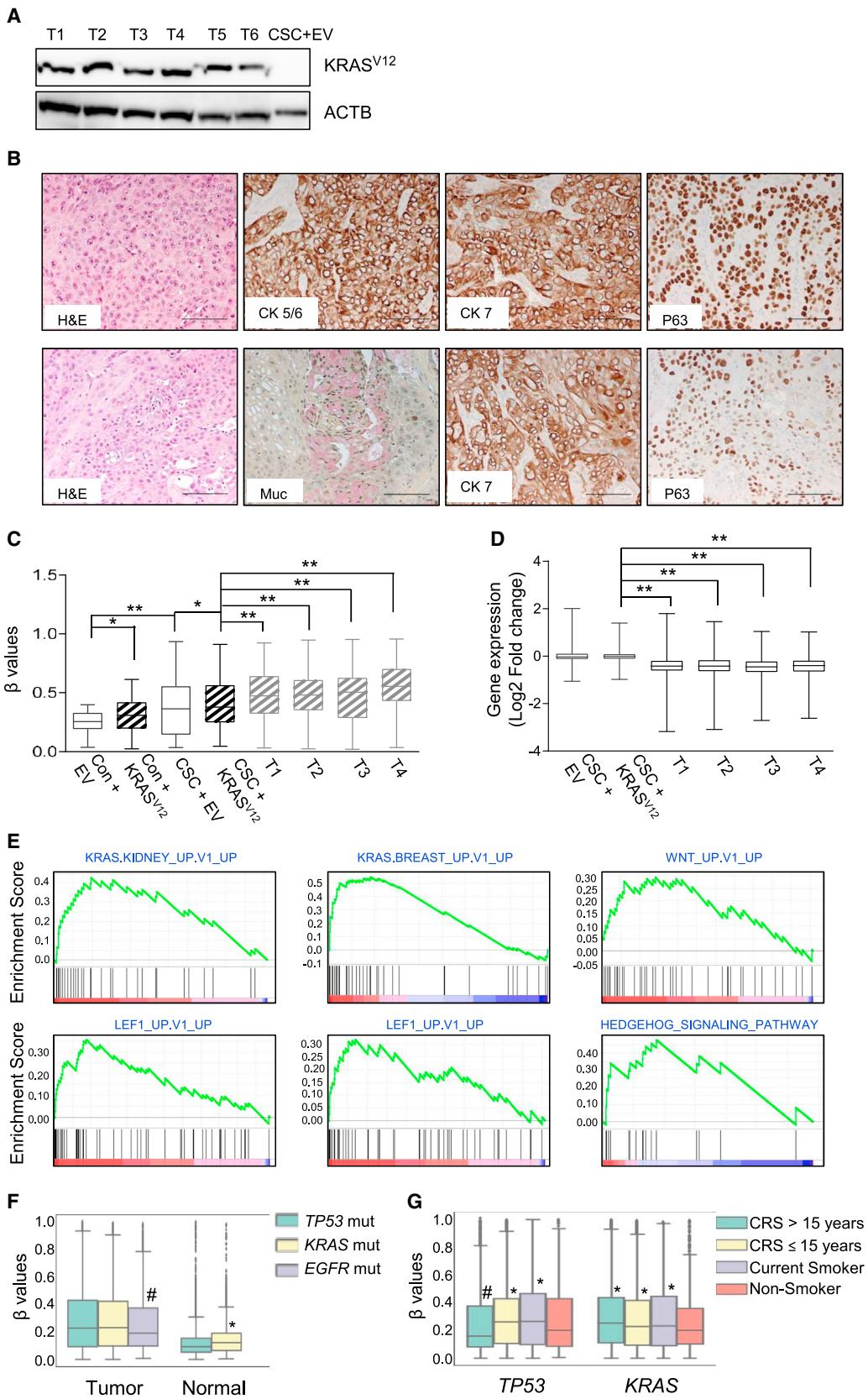
An important aspect of our data is linking the above vulnerability of involved genes to evolve such methylation to impingement of basal distribution of DNMT1 near the TSS as compared with highly transcribed genes with unmethylated promoter CpG islands. This suggests a strategic positioning for DNMT1 to mediate maintenance of promoter DNA methylation evolving with tumor initiation and/or progression. Importantly, DNMTs have scaffolding functions, which can exert transcriptional repressor functions (Clements et al., 2012; Fuks et al., 2000; Ha et al., 2011; Rountree et al., 2000). In addition, polycomb system components can recruit DNMTs to cancer-specific methylated promoters with simultaneous repression of their expression even prior to the appearance of abnormal DNA methylation (Mohammad et al., 2009).

Perhaps most importantly, we now suggest how these epigenetic abnormalities help CSC-exposed cells acquire transformed phenotypes. The augmentation of the KRAS/MAPK and WNT signaling pathways concomitant with peak abnormal gene promoter methylation at 10 months implies that the involved genes are key for driving these pathways. Indeed, functional studies have shown that these genes are important for normally preventing abnormal RAS-MAPK (Kraunz et al., 2005) and WNT pathway signaling (Revet et al., 2008, 2010; Mazieres et al., 2004; Nojima et al., 2007). Overexpression of these genes in cancer cells in which they are epigenetically silenced causes apoptosis and downregulation of key target genes such as c-MYC (Cowling et al., 2007; Suzuki et al., 2004).

Our model data may relate to the actual course of lung cancer initiation and progression. One caveat for our model system is that the cells utilized are already immortalized and the engineered loss of function of p16 would induce senescence bypass. However, we do not believe this negates our data for the temporal order of the other events we observe vis-a-vis their importance for the genesis of lung cancer. We stressed earlier that others have shown that HBECs need a combination of at least three oncogenic events to form tumors in mice (Sato et al., 2006, 2013). Indeed, insertion of KRAS mutations at 6 months of CSC exposure was not sufficient to transform the cells, consistent with the CSC-induced methylation changes playing a key role in enhancing susceptibility to one-step KRAS-induced transformation. We also hypothesize that the 15-month time span in our studies mimics the typical long course of smoking

Figure 6. Phenotypes of CSC-Exposed versus Control Cells before and after Expression of Mutant KRAS^{V12}

- (A) Soft agar assays to assess the anchorage-independent growth ability of HBEC cells following 10 months of CSC exposure.
- (B) Quantitation of soft agar colonies obtained with 10 and 15 month CSC treated cells. Values indicate average number of colonies \pm SD from replicates 1 and 2 at the indicated time points.
- (C) Phase-contrast photomicrographs showing morphological changes in control and CSC-treated cells when switched from defined serum-free to serum-containing medium for 96 hr. Scale bars, 50 μ m.
- (D) Immunofluorescent staining of E-cadherin and vimentin in cells grown in serum-containing medium for 96 hr. Scale bars, 100 μ m.
- (E) Immunoblot for EMT markers in cells (CSC and control [Con]) and soft agar clones obtained from 15-month CSC-treated cells, grown in serum-containing medium for 96 hr.
- (F and G) Expression of KRAS^{V12} (F) and TP53 (G) in 15-month control and CSC-treated cells 2 weeks following transduction with a KRAS^{V12}- or shTP53-expressing lentiviral vector, respectively. Stably infected cells were selected by growth in the respective antibiotic prior to examining gene expression.
- (H) Soft agar assays to assess the clonogenic potential of 6- and 15-month control and CSC-treated cells expressing mutant KRAS^{V12} (top panel). Representative images of soft agar colonies (bottom panel). Scale bars, 100 μ m.
- (I) Quantitation of colonies obtained in (H). Data are presented as average number of colonies \pm SD at the indicated time points. *p < 0.05 by two-tailed t test.
- (J) Signaling pathways enriched for in 15-month CSC-treated cells expressing KRAS^{V12} compared with respective empty vector controls, analyzed using GSEA.
- (K and L) Measurement of tumor latency (K) and average tumor volumes (L) of xenografts from mice injected with 15-month control and CSC-treated cells from replicate 1 (Rep 1) and replicate 2 (Rep 2) expressing KRAS^{V12} or shTP53. Data are presented as mean \pm SD (n = 5 for Rep 1 and n = 10 for Rep 2).



(legend on next page)

history for patients developing NSCLC. We have previously hypothesized that epigenetic abnormalities such as those induced by our CSC exposure provide a molecular environment facilitating cellular addiction to driver genetic events, and our current findings support this for NSCLC (Baylin, 2012). Our outlined relationships of CSC-hypermethylated genes in primary lung tumors with mutant KRAS and TP53 versus EGFR mutations and with smoking history further suggest the importance of this epigenetic change in evolution of smoking-associated NSCLC. Our modeled events may sensitize to mutations in addition to KRAS for NSCLC initiation and progression. However, they may have special relevance for helping KRAS mutations transform in a single step. Activating RAS/MAPK and WNT pathways are important for cigarette smoke-induced lung cancer (Lemjabbar-Alaoui et al., 2006; Stewart, 2014). It is also suggested that KRAS mutations alone are not a strong predictor for developing lung cancer, but associated increases in WNT signaling accelerate lung tumorigenesis (Pacheco-Pinedo et al., 2011; Pacheco-Pinedo and Morrisey, 2011). Finally, our findings stress that therapies to revert the types of epigenetic abnormalities we have now chronicled could be one key modality for preventing lung cancer in high-risk individuals.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cell Culture and Treatments
 - In Vivo Mouse Studies
- METHOD DETAILS
 - Cigarette Smoke Condensate
 - Nuclear Protein Extraction and Tight Chromatin Fractionation

- Genome-wide DNA-methylation Analysis
- Heatmaps with Methylation Data
- DAVID Analysis of Methylated Gene Lists (GO Analysis_Biological Processes)
- Comparing Methylated Gene Lists with TCGA Data (Used to Generate Venn Diagrams)
- Chromatin Immunoprecipitation (ChIP) and ChIP-seq Analysis
- Gene Expression Analysis
- Overexpression of KRAS^{V12} in HBEC Cells
- Downregulation of TP53 with shRNA
- Soft Agar Growth Assays
- In Vivo Tumorigenicity and Histologic Analysis
- Immunofluorescent Staining of Vimentin and E-cadherin
- Sample Preparation for Whole-Exome and Targeted Sequencing
- Analysis of Next-Generation Sequencing Data and Identification of Putative Somatic Mutations
- Generation of Box Plots for Analysis of Methylation Data from NSCLC Samples from TCGA for Correlation with Mutational and Smoking Status
- Generation of Box Plots for Analysis of Methylation Data from COPD (Chronic Lung Obstructive Disease) and IPF (Idiopathic Pulmonary Fibrosis) Samples

● QUANTIFICATION AND STATISTICAL ANALYSIS

● DATA AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.ccr.2017.08.006>.

AUTHOR CONTRIBUTIONS

M.V., H.P.E., and S.B.B. designed experiments, performed data analyses, and wrote the manuscript. M.V. performed experiments. S.Y.H., I.K., and A.P. performed bioinformatics analyses under the supervision of H.P.E. E.G. classified

Figure 7. Transformation of CSC-Exposed Cells Following Expression of Mutant KRAS^{V12}

(A) Expression of mutant KRAS^{V12} in xenograft tissues obtained from mice injected with CSC-treated cells expressing KRAS^{V12} (T1–T3, Rep 1 and T4–T6, Rep 2). CSC cells exposed for 15 months that were infected with empty vector (EV) are shown as a control.

(B) Representative histology of xenografts obtained from mice injected with CSC + KRAS^{V12} cells 4 weeks after appearance of the tumor. Five sections were stained from each replicate. Representative sections stained with H&E, mucicarmine, CK5/6, CK7, and p63 are shown. Scale bars, 100 μm.

(C) Comparison of methylation changes in key CSC-methylated genes from Table S2 in EV- and KRAS^{V12}-expressing control and CSC-treated cells and xenografts (T1–T4) harvested 4 weeks after appearance of tumors from replicate 2. Boxplots generated with β values of promoter CpG-island probes of selected genes. Only probes with a starting β value of ≤0.4 in EV-containing control cells were selected. Boxplots represent the 25th and 75th percentiles, with midlines indicating the median values. Whiskers extend to the minimum/maximum values of the data sample. *p < 0.01, **p < 0.0001 by Wilcoxon's rank test.

(D) Boxplots showing average log₂ fold change in expression of CSC-methylated genes from Table S2 in EV- and KRAS^{V12}-expressing CSC-exposed cells and xenografts (T1–T4) from replicate 2 as compared with their respective control cells expressing EV or KRAS^{V12}. Boxplots represent the 25th and 75th percentiles, with midlines indicating the median values. Whiskers extend to the minimum/maximum values of the data sample. **p < 0.0001 by Wilcoxon's rank test.

(E) Signaling pathways enriched for in xenografts compared with control cells expressing KRAS^{V12} analyzed with GSEA.

(F) Boxplots of β values of promoter CpG-island probes of CSC-methylated genes in lung adenocarcinoma samples from TCGA with KRAS (n = 49), TP53 (n = 49), and EGFR (n = 6) mutations and corresponding matched normal tissues (n = 4 for KRAS, n = 8 for TP53). Data are presented as the average of all probes in all patients within the respective group. *p < 0.05 by Welch's t test indicates significant increase in matched normal from KRAS tumors compared with matched normals from TP53 tumors. #p < 0.05 by Welch's t test indicates significant decrease in EGFR mutant tumors compared to tumors with KRAS and/or TP53 mutation.

(G) Boxplots of β values of promoter CpG-island probes of CSC-methylated genes in lung adenocarcinoma samples from TCGA with KRAS and TP53 mutations categorized by smoking status. *p < 0.05 by Welch's t test indicates significant increase compared with non-smoker tumor within mutation group. #p < 0.05 by Welch's t test indicates significant decrease compared with non-smoker tumor within mutation group. CRS, current reformed smoker (n = 8 current smokers, 19 CRS <15 years, 14 CRS >15 years, 4 non-smokers for KRAS tumors); n = 14 current smokers, 20 CRS <15 years, 12 CRS >15 years, 3 non-smokers for TP53 tumors). Top and bottom of boxplots shown in (F) and (G) represent the 25th and 75th percentiles (interquartile range; IQR), with midlines indicating the median values. The whiskers extend to points that lie within 1.5 IQR of the lower and upper quartile. Observations that fall outside this range are displayed independently. See also Figure S6.

xenograft histopathology. J.P. and V.E.V. performed and analyzed exome sequencing. H.M.O'H. assisted with data analyses and manuscript review. C.A.Z. and L.M. assisted with *in vivo* experiments.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Cancer Institute under award number R01CA043318 (S.B.B.), R01CA170550 (S.B.B.), NIH R01CA121113 (V.E.V.), P30CA006973 (S.B.B., V.E.V.), U10CA180950 (V.E.V.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Support for this research was also provided by grants from the National Institute of Environmental Health Sciences, National Institutes of Health R01ES011858 (S.B.B.), R01ES023183 (H.M.O'H.). Further support: The Hodson Trust (S.B.B.), the Commonwealth Foundation (V.E.V.); The Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (V.E.V., S.B.B.); The SU2C-DCS International Translational Cancer Research Dream Team (SU2C-AACR-DT1415; V.E.V.), Stand Up To Cancer is a program of the Entertainment Industry Foundation, Administered by AACR; and Flight Attendant Medical Research Institute (M.V.). We thank Kathy Bender for help with manuscript preparation. V.E.V. is a founder of Personal Genome Diagnostics, is a member of its Scientific Advisory Board and Board of Directors, and owns Personal Genome Diagnostics stock, which is subject to certain restrictions under university policy. V.E.V. is also on the Scientific Advisory Board for Ignyta. The terms of these arrangements are managed by the Johns Hopkins University in accordance with its conflict of interest policies.

Received: August 30, 2016

Revised: June 21, 2017

Accepted: August 11, 2017

Published: September 11, 2017

REFERENCES

- Aran, D., Sabato, S., and Hellman, A. (2013). DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol.* **14**, R21.
- Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., and Fisher, A.G. (2006). Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8**, 532–538.
- Baylin, S.B. (2012). The cancer epigenome: its origins, contributions to tumorigenesis, and translational implications. *Proc. Am. Thorac. Soc.* **9**, 64–65.
- Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome—biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734.
- Baylin, S.B., and Ohm, J.E. (2006). Epigenetic gene silencing in cancer—a mechanism for early oncogenic pathway addiction? *Nat. Rev. Cancer* **6**, 107–116.
- Belinsky, S.A., Palmisano, W.A., Gilliland, F.D., Crooks, L.A., Divine, K.K., Winters, S.A., Grimes, M.J., Harms, H.J., Tellez, C.S., Smith, T.M., et al. (2002). Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res.* **62**, 2370–2377.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326.
- Bertotti, A., Papp, E., Jones, S., Adleff, V., Anagnostou, V., Lupo, B., Sausen, M., Phallen, J., Hruban, C.A., Tokheim, C., et al. (2015). The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature* **526**, 263–267.
- Bibikova, M., and Fan, J.B. (2009). GoldenGate assay for DNA methylation profiling. *Methods Mol. Biol.* **507**, 149–163.
- Boch, C., Kollmeier, J., Roth, A., Stephan-Falkenau, S., Misch, D., Gruning, W., Bauer, T.T., and Mairinger, T. (2013). The frequency of EGFR and KRAS mutations in non-small cell lung cancer (NSCLC): routine screening data for central Europe from a cohort study. *BMJ Open* **3**, e002560.
- Camilo, R., Capelo, V.L., Siqueira, S.A., and Del Carlo Bernardi, F. (2006). Expression of p63, keratin 5/6, keratin 7, and surfactant-A in non-small cell lung carcinomas. *Hum. Pathol.* **37**, 542–546.
- Campeau, E., Ruhl, V.E., Rodier, F., Smith, C.L., Rahmberg, B.L., Fuss, J.O., Campisi, J., Yaswen, P., Cooper, P.K., and Kaufman, P.D. (2009). A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS One* **4**, e6529.
- Cancer Genome Atlas Research Network. (2012). Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525.
- Cancer Genome Atlas Research Network. (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550.
- Clements, E.G., Mohammad, H.P., Leadem, B.R., Easwaran, H., Cai, Y., Van Neste, L., and Baylin, S.B. (2012). DNMT1 modulates gene expression without its catalytic activity partially through its interactions with histone-modifying enzymes. *Nucleic Acids Res.* **40**, 4334–4346.
- Cowling, V.H., D'Cruz, C.M., Chodosh, L.A., and Cole, M.D. (2007). c-Myc transforms human mammary epithelial cells through repression of the Wnt inhibitors DKK1 and SFRP1. *Mol. Cell Biol.* **27**, 5135–5146.
- Dai, Z., Lakshmanan, R.R., Zhu, W.G., Smiraglia, D.J., Rush, L.J., Fruhwald, M.C., Brena, R.M., Li, B., Wright, F.A., Ross, P., et al. (2001). Global methylation profiling of lung cancer identifies novel methylated genes. *Neoplasia* **3**, 314–323.
- Damiani, L.A., Yingling, C.M., Leng, S., Romo, P.E., Nakamura, J., and Belinsky, S.A. (2008). Carcinogen-induced gene promoter hypermethylation is mediated by DNMT1 and causal for transformation of immortalized bronchial epithelial cells. *Cancer Res.* **68**, 9005–9014.
- Delgado, O., Kaisani, A.A., Spinola, M., Xie, X.J., Batten, K.G., Minna, J.D., Wright, W.E., and Shay, J.W. (2011). Multipotent capacity of immortalized human bronchial epithelial cells. *PLoS One* **6**, e22023.
- Dimauro, T., and David, G. (2010). Ras-induced senescence and its physiological relevance in cancer. *Curr. Cancer Drug Targets* **10**, 869–876.
- Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191.
- Easwaran, H., Johnstone, S.E., Van Neste, L., Ohm, J., Mosbruger, T., Wang, Q., Aryee, M.J., Joyce, P., Ahuja, N., Weisenberger, D., et al. (2012). A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res.* **22**, 837–849.
- Easwaran, H., Tsai, H.C., and Baylin, S.B. (2014). Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol. Cell* **54**, 716–727.
- Fuks, F., Burgers, W.A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat. Genet.* **24**, 88–91.
- Govindan, R., Ding, L., Griffith, M., Subramanian, J., Dees, N.D., Kanchi, K.L., Maher, C.A., Fulton, R., Fulton, L., Wallis, J., et al. (2012). Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **150**, 1121–1134.
- Ha, K., Lee, G.E., Palii, S.S., Brown, K.D., Takeda, Y., Liu, K., Bhalla, K.N., and Robertson, K.D. (2011). Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery. *Hum. Mol. Genet.* **20**, 126–140.
- Helman, E., Naxerova, K., and Kohane, I.S. (2012). DNA hypermethylation in lung cancer is targeted at differentiation-associated genes. *Oncogene* **31**, 1181–1188.
- Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13.
- Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.
- Issa, J.P. (2014). Aging and epigenetic drift: a vicious cycle. *J. Clin. Invest.* **124**, 24–29.
- Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. *Cell* **128**, 683–692.
- Kaufman, C.K., Mosimann, C., Fan, Z.P., Yang, S., Thomas, A.J., Ablain, J., Tan, J.L., Fogley, R.D., van Rooijen, E., Hagedorn, E.J., et al. (2016). A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science* **351**, aad2197.

- Kraunz, K.S., Nelson, H.H., Liu, M., Wiencke, J.K., and Kelsey, K.T. (2005). Interaction between the bone morphogenetic proteins and Ras/MAP-kinase signalling pathways in lung cancer. *Br. J. Cancer* 93, 949–952.
- Leary, R.J., Cummins, J., Wang, T.L., and Velculescu, V.E. (2007). Digital karyotyping. *Nat. Protoc.* 2, 1973–1986.
- Lee, E., Yang, J., Ku, M., Kim, N.H., Park, Y., Park, C.B., Suh, J.S., Park, E.S., Yook, J.I., Mills, G.B., et al. (2015). Metabolic stress induces a Wnt-dependent cancer stem cell-like state transition. *Cell Death Dis.* 6, e1805.
- Lemjabbar-Alaoui, H., Dasari, V., Sidhu, S.S., Mengistab, A., Finkbeiner, W., Gallup, M., and Basbaum, C. (2006). Wnt and Hedgehog are critical mediators of cigarette smoke-induced lung cancer. *PLoS One* 1, e93.
- Liu, F., Killian, J.K., Yang, M., Walker, R.L., Hong, J.A., Zhang, M., Davis, S., Zhang, Y., Hussain, M., Xi, S., et al. (2010). Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. *Oncogene* 29, 3650–3664.
- Macaluso, M., Paggi, M.G., and Giordano, A. (2003). Genetic and epigenetic alterations as hallmarks of the intricate road to cancer. *Oncogene* 22, 6472–6478.
- Maegawa, S., Gough, S.M., Watanabe-Okochi, N., Lu, Y., Zhang, N., Castoro, R.J., Estecio, M.R., Jelinek, J., Liang, S., Kitamura, T., et al. (2014). Age-related epigenetic drift in the pathogenesis of MDS and AML. *Genome Res.* 24, 580–591.
- Mazieres, J., He, B., You, L., Xu, Z., Lee, A.Y., Mikami, I., Reguart, N., Rosell, R., McCormick, F., and Jablons, D.M. (2004). Wnt inhibitory factor-1 is silenced by promoter hypermethylation in human lung cancer. *Cancer Res.* 64, 4717–4720.
- Mohammad, H.P., Cai, Y., McGarvey, K.M., Easwaran, H., Van Neste, L., Ohm, J.E., O'Hagan, H.M., and Baylin, S.B. (2009). Polycomb CBX7 promotes initiation of heritable repression of genes frequently silenced with cancer-specific DNA hypermethylation. *Cancer Res.* 69, 6322–6330.
- Nguyen, D.X., Chiang, A.C., Zhang, X.H., Kim, J.Y., Kris, M.G., Ladanyi, M., Gerald, W.L., and Massague, J. (2009). WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. *Cell* 138, 51–62.
- Niwa, T., and Ushijima, T. (2010). Induction of epigenetic alterations by chronic inflammation and its significance on carcinogenesis. *Adv. Genet.* 71, 41–56.
- Nojima, M., Suzuki, H., Toyota, M., Watanabe, Y., Maruyama, R., Sasaki, S., Sasaki, Y., Mita, H., Nishikawa, N., Yamaguchi, K., et al. (2007). Frequent epigenetic inactivation of SFRP genes and constitutive activation of Wnt signaling in gastric cancer. *Oncogene* 26, 4699–4713.
- O'Hagan, H.M., Mohammad, H.P., and Baylin, S.B. (2008). Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. *PLoS Genet.* 4, e1000155.
- O'Hagan, H.M., Wang, W., Sen, S., Destefano Shields, C., Lee, S.S., Zhang, Y.W., Clements, E.G., Cai, Y., Van Neste, L., Easwaran, H., et al. (2011). Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. *Cancer Cell* 20, 606–619.
- Ohm, J.E., and Baylin, S.B. (2007). Stem cell chromatin patterns: an instructive mechanism for DNA hypermethylation? *Cell Cycle* 6, 1040–1043.
- Pacheco-Pinedo, E.C., Durham, A.C., Stewart, K.M., Goss, A.M., Lu, M.M., Demayo, F.J., and Morrisey, E.E. (2011). Wnt/beta-catenin signaling accelerates mouse lung tumorigenesis by imposing an embryonic distal progenitor phenotype on lung epithelium. *J. Clin. Invest.* 121, 1935–1945.
- Pacheco-Pinedo, E.C., and Morrisey, E.E. (2011). Wnt and Kras signaling-dark siblings in lung cancer. *Oncotarget* 2, 569–574.
- Ramirez, R.D., Sheridan, S., Girard, L., Sato, M., Kim, Y., Pollack, J., Peyton, M., Zou, Y., Kurie, J.M., DiMaio, J.M., et al. (2004). Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. *Cancer Res.* 64, 9027–9034.
- Revet, I., Huizenga, G., Chan, A., Koster, J., Volckmann, R., van Sluis, P., Ora, I., Versteeg, R., and Geerts, D. (2008). The MSX1 homeobox transcription factor is a downstream target of PHOX2B and activates the Delta-Notch pathway in neuroblastoma. *Exp. Cell Res.* 314, 707–719.
- Revet, I., Huizenga, G., Koster, J., Volckmann, R., van Sluis, P., Versteeg, R., and Geerts, D. (2010). MSX1 induces the Wnt pathway antagonist genes DKK1, DKK2, DKK3, and SFRP1 in neuroblastoma cells, but does not block Wnt3 and Wnt5A signalling to DVL3. *Cancer Lett.* 289, 195–207.
- Rountree, M.R., Bachman, K.E., and Baylin, S.B. (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat. Genet.* 25, 269–277.
- Sato, M., Larsen, J.E., Lee, W., Sun, H., Shames, D.S., Dalvi, M.P., Ramirez, R.D., Tang, H., DiMaio, J.M., Gao, B., et al. (2013). Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations. *Mol. Cancer Res.* 11, 638–650.
- Sato, M., Vaughan, M.B., Girard, L., Peyton, M., Lee, W., Shames, D.S., Ramirez, R.D., Sunaga, N., Gazdar, A.F., Shay, J.W., and Minna, J.D. (2006). Multiple oncogenic changes (K-RAS(V12), p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer Res.* 66, 2116–2128.
- Sausen, M., Leary, R.J., Jones, S., Wu, J., Reynolds, C.P., Liu, X., Blackford, A., Parmigiani, G., Diaz, L.A., Jr., Papadopoulos, N., et al. (2013). Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. *Nat. Genet.* 45, 12–17.
- Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., et al. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat. Genet.* 39, 232–236.
- Shen, H., and Laird, P.W. (2013). Interplay between the cancer genome and epigenome. *Cell* 153, 38–55.
- Singh, A., Sweeney, M.F., Yu, M., Burger, A., Greninger, P., Benes, C., Haber, D.A., and Settleman, J. (2012). TAK1 inhibition promotes apoptosis in KRAS-dependent colon cancers. *Cell* 148, 639–650.
- Stange, D.E., Koo, B.K., Huch, M., Sibbel, G., Basak, O., Lyubimova, A., Kujala, P., Bartfeld, S., Koster, J., Geahlen, J.H., et al. (2013). Differentiated Troy⁺ chief cells act as reserve stem cells to generate all lineages of the stomach epithelium. *Cell* 155, 357–368.
- Stewart, D.J. (2014). Wnt signaling pathway in non-small cell lung cancer. *J. Natl. Cancer Inst.* 106, djt356.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Suzuki, H., Watkins, D.N., Jair, K.W., Schuebel, K.E., Markowitz, S.D., Chen, W.D., Pretlow, T.P., Yang, B., Akiyama, Y., Van Engeland, M., et al. (2004). Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nat. Genet.* 36, 417–422.
- Tata, P.R., Mou, H., Pardo-Saganta, A., Zhao, R., Prabhu, M., Law, B.M., Vinarsky, V., Cho, J.L., Breton, S., Sahay, A., et al. (2013). Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature* 503, 218–223.
- Tellez, C.S., Juri, D.E., Do, K., Bernauer, A.M., Thomas, C.L., Damiani, L.A., Tessema, M., Leng, S., and Belinsky, S.A. (2011). EMT and stem cell-like properties associated with miR-205 and miR-200 epigenetic silencing are early manifestations during carcinogen-induced transformation of human lung epithelial cells. *Cancer Res.* 71, 3087–3097.
- Tessema, M., Yingling, C.M., Liu, Y., Tellez, C.S., Van Neste, L., Baylin, S.S., and Belinsky, S.A. (2014). Genome-wide unmasking of epigenetically silenced genes in lung adenocarcinoma from smokers and never smokers. *Carcinogenesis* 35, 1248–1257.
- Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J. Clin.* 65, 87–108.
- Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., and Laird, P.W. (2007). Epigenetic stem cell signature in cancer. *Nat. Genet.* 39, 157–158.
- You, J.S., and Jones, P.A. (2012). Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* 22, 9–20.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-DNMT1	Sigma-Aldrich	Cat#D4692; RRID: AB_262096
Rabbit monoclonal anti-EZH2 XP (D2C9)	Cell Signaling Technology	Cat#5246; RRID: AB_10694683
Rabbit polyclonal anti-SIRT1 (C19)	Delta Biolabs	Cat#DB083; RRID: AB_2239223
Goat polyclonal anti-Lamin B (C-20)	Santacruz Biotechnology	Cat#sc-6216; RRID:AB_648156
Rabbit polyclonal anti-H3K4me3	Millipore	Cat# 07-473; RRID:AB_1977252
Rabbit polyclonal anti-H3K27m3	Millipore	Cat# 07-449; RRID:AB_310624
Mouse monoclonal anti-E-cadherin	BD Transduction laboratories	Cat# 610181; RRID:AB_397580
Rabbit monoclonal anti-Vimentin (D21H3)	Cell Signaling Technology	Cat#5741; RRID:AB_10695459
Mouse monoclonal anti-TP53 (Ab-6)	Millipore	Cat# OP43; RRID:AB_10682938
Rabbit monoclonal anti-Ras (G12V Mutant Specific)	Cell Signaling Technology	Cat#14412
Mouse monoclonal anti-β-Actin	Sigma-Aldrich	Cat#A5441; RRID:AB_476744
Goat anti-Rabbit IgG Alexa Fluor 488	Thermo Fisher	Cat# A-11034; RRID:AB_2576217
Goat anti-Rabbit IgG Alexa Fluor 594	Thermo Fisher	Cat#A-11032; RRID:AB_2534091
Bacterial and Virus Strains		
pLenti-PGK-KRAS4B(G12V)	Singh et al., 2012	Addgene Plasmid #35633
pLenti PGK Hygro DEST (w530-1)	Campeau et al., 2009	Addgene Plasmid #19066
pRFP-CB-shLenti Lentiviral vector	OriGene	Cat#TR30032
Chemicals, Peptides, and Recombinant Proteins		
Cigarette smoke condensate	Murty pharmaceuticals	NA
Lipofectamine® 3000 Transfection Reagent	Thermo Fisher	Cat#L3000015
Complete EDTA free protease inhibitor	Sigma Aldrich	Cat#11873580001
NuPAGE® LDS Sample Buffer (4X)	ThermoFisher	Cat#NP0007
Hygromycin B	Corning Cellgro	Cat#30240CR
Blasticidin S HCl	ThermoFisher	Cat#A1113903
Phosphatase Inhibitor Cocktail 2	Sigma Aldrich	Cat#P5726
N-ethylmaleimide	Sigma	Cat#E3876
Pefabloc SC AEBSF	Roche Applied Science	Cat#11427393103
Dynabeads® Protein A	ThermoFisher	Cat#10001D
Dynabeads® Protein G	ThermoFisher	Cat#10003D
Critical Commercial Assays		
EZ DNA Methylation™ Kit	Zymo Research	Cat#D5002
RNeasy Mini kit	Qiagen	Cat#74104
DNeasy Blood & Tissue Kit	Qiagen	Cat#69504
qScript cDNA SuperMix	Quanta Biosciences	Cat#95048
Deposited Data		
Raw and analyzed data for DNA methylation, gene expression array, ChIP-seq, Exome-seq and targeted Deep-seq	This paper	GEO: GSE101864
Experimental Models: Cell Lines		
Human: HBEC3KT	Obtained from Drs. Shay and Minna, Southwestern Medical Center, Dallas, TX	Ramirez et al., 2004; RRID:CVCL_X491
Experimental Models: Organisms/Strains		
Mouse: NOD scid gamma (NOD.Cg-Prkdc ^{scid} II2rg ^{tm1Wjl} /SzJ)	The Jackson Laboratory	Cat#005557

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
Primers for qRT-PCR, see Table S4	This paper	N/A
shRNA targeting sequence for TP53: (ACGGCGCACAGAGGAAGAGAAT) cloned in pRFP-CB-shLenti	Origene	N/A
Primers for ChIP, see Table S4	This paper	N/A
Software and Algorithms		
ImageJ software	Open source	https://imagej.nih.gov/ij/ . RRID:SCR_003070
GraphPad Prism	GraphPad Software	http://www.graphpad.com/ . RRID:SCR_002798
Gene set enrichment Analysis	(Subramanian et al., 2005)	http://www.broadinstitute.org/gsea/ . RRID:SCR_003199
DAVID Bioinformatics resources database (v6.7)	(Huang da et al., 2009a; Huang da et al., 2009b).	http://david.abcc.ncifcrf.gov/ . RRID:SCR_001881

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Stephen Baylin (sbaylin@jhmi.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Cell Culture and Treatments**

Human bronchial epithelial cells (HBEC-3KT) were obtained from Drs. Shay and Minna, Southwestern Medical Center, Dallas, TX (Ramirez et al., 2004). Cells were maintained at 37°C in the presence of 5% CO₂ in keratinocyte serum-free medium (KSFM; Life Technologies Inc.) media containing 50 mg/mL of bovine pituitary extract (BPE; Life Technologies Inc.) and 5 ng/mL of EGF (Life Technologies Inc.) (Ramirez et al., 2004).

Cells were passaged every 2-3 days and fresh CSC and DMSO were added to the CSC and control treated cells respectively at each passage. Cells were treated with 5 µg/ml and 10 µg/ml CSC (Murthy pharmaceuticals) or corresponding concentrations of DMSO (Controls).

Two independent sets of treatments (replicate 1 and replicate 2) were done to obtain the two technical repeats that were used for all studies. To determine if CSC treated cells exhibited EMT like characteristics, CSC treated cells at 10 and 15 months along with their respective controls were grown in medium with 5% FBS for 96 hr as described earlier (Sato et al., 2013).

In Vivo Mouse Studies

We used male NSG mice (5-6 weeks old) (Jackson laboratories) for these studies. Mice were housed at the Johns Hopkins Animal care facility. All animal experiments were approved by the Johns Hopkins Animal Care and Use Committee. All animal care and protocols followed were in accordance with guidelines of the institutional Animal Care and Use Committee (IACUC). Mice were injected subcutaneously in the flank with 2 x 10⁶ viable CSC and control cells obtained from the 15 month treated cells from both replicates expressing KRAS^{V12} or shTP-53 in 0.2 mL of KSFM with matrigel. Mice were monitored every 3 days for tumor formation for up to 5 months.

METHOD DETAILS**Cigarette Smoke Condensate**

CSC was purchased from Murthy pharmaceuticals (Lexington, KY, USA). According to the manufacturer's brochure, the condensate is obtained by smoking University of Kentucky's Standard Research cigarettes (3R4F; University of Kentucky, KY, USA) on an FTC Smoke Machine. The total particulate matter was collected on Cambridge glass fibre filters and the amount obtained determined by weight increase of the filter. CSC was prepared by dissolving the collected smoke particulates in dimethyl sulfoxide (DMSO) to yield a 4% (40 mg/mL) solution (w/v).

Nuclear Protein Extraction and Tight Chromatin Fractionation

CSC or control (DMSO treated) cells were collected at the indicated time points. To prepare nuclear extracts, cell pellets were lysed in CEBN buffer (10 mM HEPES [pH 7.8], 10 mM KCl, 1.5 mM MgCl₂, 0.34 M sucrose, 10% glycerol, 0.2 % NP-40, 1X protease inhibitor

cocktail (Invitrogen), 1X phosphatase Inhibitor cocktail, N-ethylmaleimide (Sigma) and Pefabloc SC AEBSF (Roche Applied Science)) and pellets were washed with CEB buffer (CEBN buffer without NP-40). For preparation of total nuclear protein extracts, nuclear pellets were lysed with 4% SDS and passed through QIAshredder (Qiagen). For preparation of Tight Chromatin fractions, nuclear pellets obtained after washing with CEB buffer, were further washed with soluble nuclear buffer (3 mM EDTA, 0.2 mM EGTA, inhibitors as described above), salt buffer with 0.45M NaCl concentration (50 mM Tris pH 8.0, 0.05% NP40, 0.45M NaCl). The pellet was then lysed in 4% SDS buffer using a Qiashredder column (Qiagen) and referred to as tight chromatin. Western blots of the protein extracts were probed with DNMT1 (Sigma), EZH2 XP (Cell Signaling), SIRT1 (Delta Biolabs), Lamin B (Santacruz Biotechnology). Band densitometry for western blots were quantitated using ImageJ software.

Genome-wide DNA-methylation Analysis

Genome-wide DNA methylation was studied in the control and CSC treated cells at 1, 6, 10 and 15 months as well as in the 15 month control and CSC treated cells expressing empty vector (EV) or KRAS^{V12} and the tumor xenografts . These samples were collected from both replicates. DNA was extracted from cells and xenograft tissues using the Qiagen DNeasy blood and tissue kit (Qiagen, CA).The Infinium HumanMethylation450k BeadChip array (Illumina) (Bibikova and Fan, 2009) was used to analyze bisulfite-treated DNA (EZ DNA-Hypermethylation Kit, Zymo Research) as previously described (Easwaran et al., 2012). The R Bioconductor package minfi was used to process the raw IDAT files. Probes with a detection p value > 0.01 and probes associated with sex chromosomes were removed after the data was SWAN-normalized. B values were computed as the signal of the methylation-specific probe over the sum of the signals of the methylation- and unmethylated-specific probes. The data is then subsetted into further groups: cpg islands, body, shelf and shore probes using the annotation information available from Bioconductor IlluminaHumanMethylation450k.db library. To further get a comprehensive list of all probes that are around the TSS, the CpG coordinates for the Infinium 450K array were obtained from IlluminaHumanMethylation450k.db and mapped to \pm 1500 bp around all transcripts in hg19 (UCSC) using R package BiomaRt (Durinck et al., 2009). Probes around the TSS were split by whether or not they are within a CpG island (obtained from hg19 genome from UCSC (<https://genome.ucsc.edu/>)) to obtain the CpG-island and non-CPG island containing TSS/promoter probes. Probes in putative enhancer regions were obtained by selecting for probes from Infinium HumanMethylation450k BeadChip array that matched with enhancer regions identified in lung cancer in an earlier published study (Aran et al., 2013). This list was further filtered to exclude probes that were within \pm 2 kb of a TSS. Scatter plots to examine methylation changes occurring between control and CSC treated cells at each time point were generated by plotting CpG-island probes mapped to \pm 1500bp of TSS.

To obtain a list of CSC methylated genes, we first chose probes that had a starting β value of \leq 0.2 in the untreated cells at 1 month. We then filtered for those probes whose β values did not increase by more than 0.2 between the control at 1M and the respective control end point and showed an increase of at least 0.3 in β values in CSC treated cells as compared to β values in the control cells at the respective end point (6 months, 10 months and 15 months).

Hypermethylated gene lists for each time point of replicates 1 and 2 were generated using the following equations:

6 months: Con(1 M) \leq 0.2 AND ($\Delta\beta$ Con (6 M-1 M) \leq 0.2 and $\Delta\beta$ CSC 6 M - Con 6 M \geq 0.25),

10 months: Con(1 M) \leq 0.2 AND ($\Delta\beta$ Con (10 M-1 M) \leq 0.2 and $\Delta\beta$ CSC 10 M - Con10 M \geq 0.25),

15 months: Con(1 M) \leq 0.2 AND ($\Delta\beta$ Con (15 M - 1 M) \leq 0.2 and $\Delta\beta$ CSC 15 M-Con 15 M \geq 0.25).

Hypomethylated genes lists for each time point of replicates 1 and 2 were generated using the following equations:

6 months: Con(1 M) \geq 0.5 AND ($\Delta\beta$ Con (6 M -1 M) $>$ - 0.2 AND $\Delta\beta$ CSC 6 M-Con 6 M \leq -0.2

10 months: Con(1 M) \geq 0.5 AND ($\Delta\beta$ Con (10 M -1 M) $>$ - 0.2 AND $\Delta\beta$ CSC 10 M-Con 10 M \leq -0.2

15 months: Con(1 M) \geq 0.5 AND ($\Delta\beta$ Con (15 M - 1 M) $>$ - 0.2 AND $\Delta\beta$ CSC 15 M - Con 10 M \leq -0.2

Heatmaps with Methylation Data

Heat maps are based on hierarchical clustering of β values using Euclidean distance and Ward's algorithm (R package gplots). Heatmaps for hypermethylated probes were generated by clustering top variable CpG-island probes mapped to \pm 1500bp of TSS that had a starting β value of \leq 0.2 in controls at one month. Heatmaps for hypomethylated probes were generated by clustering the top 2.5% variable CpG-island probes mapped to \pm 1500bp of TSS and had a β value of $>$ 0.5 in the control cells are one month.

DAVID Analysis of Methylated Gene Lists (GO Analysis_Biological Processes)

Probes with a $\Delta\beta$ cut off of \geq 0.3 picked from methylated gene list obtained as described above, from replicate 1 and 2 (Table 1) were combined and analyzed for Gene Ontology (GO) enrichment for biological processes (BP) using the DAVID Bioinformatics resources database (Huang da et al., 2009a; Huang da et al., 2009b). Only categories that were below the false discovery rate (Benjamini and Hochberg) of 0.05 are reported. The methylated gene list (Table S1) was also analyzed using the DAVID Functional Annotation Tool (version 6.7) to obtain enriched functional categories at each time point and a list of genes enriched for in a particular functional category was generated as reported in Table S2.

Comparing Methylated Gene Lists with TCGA Data (Used to Generate Venn Diagrams)

Infinium methylation array data for primary NSCLC, breast, colon and pancreatic tumors were used to test if genes identified as methylated in CSC treated cells are also methylated in primary tumors. Data were downloaded from the Cancer Genome Atlas (TCGA; <http://tcga.cancer.gov/dataportal>). β values were calculated as above followed by quantile normalization of samples within the same tumor type. A conservative cut-off of β values > 0.5 in the TCGA data was used to call samples as methylated. Hypermethylated probes in LUAD and LUSC were identified as follows: For each probe that have mean β value in normal cells ≤ 0.2 , we estimated the proportion of samples in the tumors that have β value greater than 2 standard deviations from the mean β value of normal and that the difference in β value for the probe in the sample relative to the mean β value of normal is greater than 0.2. Those probes that are methylated in at least 25 % of the tumor samples were defined as hypermethylated and the corresponding genes were identified as described earlier.

Chromatin Immunoprecipitation (ChIP) and ChIP-seq Analysis

For ChIP and ChIP-seq studies, cells were crosslinked using 1% formaldehyde. Nuclei were isolated using CEBN buffer as described above, lysed and the chromatin was sonicated to obtain fragments ranging from 150-250 bp. Chromatin was immunoprecipitated by antibodies to DNMT1(Sigma), EZH2 XP (Cell Signaling), H3K4Me3 (Millipore) and H3K27Me3 (Millipore) and captured on ProteinA/G magnetic beads (DynaBeads). Bound chromatin was washed in low-salt, high-salt buffers and TE and subjected to elution, de-crosslinking, and Proteinase-K treatment at 65°C in elution buffer followed by column purification (Qiagen, Min Elute). Primers used for ChIP are indicated in [Table S4](#). ChIP was done with cells from both replicates at all time points indicated.

For ChIP-seq studies purified DNA was then used to prepare a sequencing library and sequenced on Applied Biosystems SOLiD (V3). Sequencing reads were aligned to hg19 (UCSC) using bowtie1.0.0, which guarantees finding all alignments between the first 25 bp of the read (seed) and the reference sequence with up to two mismatches. Each match is extended to the full length of the read, scoring 1 point for matching and -2 points for mismatching bases. The read is trimmed to the length with the highest score. If there is only one alignment or if an alignment scores significantly higher than the others for the same read, it is considered unique and reported. ChIP-seq reads were normalized to total read counts in each sample. CSC methylated genes used for this analysis were picked from the 10 month methylated gene list from replicate 1 since ChIP-seq analysis was performed with cells obtained from this replicate. The methylated gene list was filtered to pick genes which showed $\Delta\beta$ Con (10 M-1 M) ≤ 0.2 and $\Delta\beta$ CSC 10 M - Con 10 M ≥ 0.35 . A similar number of random CpG-island containing but unmethylated genes were picked as a control set. For the analysis using heatmaps in [Figure 4C](#), for the CSC methylated and a random set of unmethylated genes, we obtained the ChIP-seq reads in 200bp bins (obtained using BedTools) ± 5 Kb around the TSS. The rows in the heatmap corresponding to the TSS region were ordered by lowest to highest (top to bottom) number of reads in the H3K4me3 ChIP-seq in the control samples at 10 days. This order was maintained for all the other ChIP-seq data plotted in these heatmaps. Average plots were generated by averaging the normalized ChIP-seq reads per bin for the set of genes in each plot.

Gene Expression Analysis

Gene expression was studied in the control and CSC treated cells at 10 days, 1, 3, 6, 10 and 15 months as well as in the 15 month control and CSC treated cells expressing empty vector (EV) or KRASV¹² and the tumor xenografts. These samples were collected from both replicates. RNA was extracted using the RNeasy Mini kit (Qiagen) and processed for hybridization on an Human Gene Expression v2 4x44K Microarray chip. The R/Bioconductor package limma was used to process expression data. Within- and between-array normalizations were performed using the loess and aquantile methods, respectively. The normexp option was used for background correction. Raw files were read in using the read.maimages function. Log fold change in gene expression for CSC treated over respective controls was obtained for each sample at each time point studied. Ranked lists of log fold change were analyzed using Gene Set Enrichment Analysis (GSEA) by the Broad Institute and data packages (Reactome, KEGG, oncogenic) ([Subramanian et al., 2005](#)). Pathways enriched with a false discovery rate less than 0.25 and p value < 0.05 were chosen. For analysis of gene expression in mouse xenografts obtained from replicate 1 and 2, all genes that were upregulated/downregulated by 2 fold or higher as compared to the respective control cells with KRAS^{G12V} were selected. Ranked lists were prepared with these genes and analyzed by GSEA using the same data packages described above. Gene expression changes were also plotted for a key set of methylated genes identified in [Table S2](#) ([Figures 7D and S6](#)). Fold change in gene expression of these genes in the tumor xenografts was calculated by comparing with their expression levels in the respective control cells expressing KRAS^{V12}.

qRT-PCR to validate expression of select genes analyzed in ChIP studies in control and CSC treated cells at the indicated time points was performed using the primers in [Table S4](#).

Overexpression of KRAS^{12V} in HBEC Cells

Control and CSC treated cells at 6 and 15 months were transduced with Lentiviral vectors expressing mutant KRAS^{12V} or an empty vector control. Vectors were gifts from Daniel Haber (Addgene plasmid # 35633 and # 19066) ([Singh et al., 2012](#)). Transduced cells were selected with hygromycin. Expression of KRAS^{V12} was confirmed by western blot using a KRAS G12V mutant specific antibody (Cell Signaling).

Downregulation of TP53 with shRNA

The following shRNA sequence for TP-53 (ACGGCGCACAGAGGAAGAGAAT) was cloned in the pRFP-CB-shLenti Lentiviral vector (TR30032) by OriGene. This sequence was obtained from SIGMA mission shRNA sequences shown to have an efficiency of knocking down TP53 by almost 90%. A control scrambled shRNA cloned in the similar vector (TR30033) obtained from OriGene was used as a negative control. Cells were transduced with lentiviral constructs expressing either shTP53 or scrambled shRNA control and selected by growth in medium with Blasticidin S.

Soft Agar Growth Assays

Clonogenic potential of control and CSC treated cells at 10 months was determined by soft agar assays. 1,000 viable cells were suspended and plated in 0.4% agar in K-SFM medium in 6-well plates, and were layered over a 0.6% agar base in the same medium. Colonies were stained with ethidium bromide prior to enumeration. Similar assays were also done to determine clonogenicity of 6 month and 15 month control and CSC treated cells expressing KRAS^{V12}.

In Vivo Tumorigenicity and Histologic Analysis

2×10^6 viable CSC and control cells expressing KRAS^{V12} in 0.2 mL of KSFM with matrigel were injected subcutaneously in the flank of male NSG mice (5–6 weeks). Mice were monitored every 3 days for tumor formation. Formalin-fixed, paraffin-embedded (FFPE) xenograft tumor tissue was sectioned and stained with hematoxylin and eosin (H&E) and mucicarmine for histologic analysis. Immunohistochemical staining for p63, TTF1, CK5 and CK7 was done in the Johns Hopkins Medicine Department of Pathology Immunohistochemistry laboratory.

Immunofluorescent Staining of Vimentin and E-cadherin

Cells grown on chamber slides were fixed in 4% formaldehyde for 15 min, following which they were washed thrice in 1X PBS for 5 min each. Cells were then blocked for 1 hr in blocking buffer (1X PBS/5% normal goat serum/0.3% Triton X-100). Primary antibodies (E-cadherin and Vimentin) were added at 1.100 dilution overnight at 4°C. Slides were then washed thrice with 1X PBS and fluoro-chrome-conjugated secondary antibodies added at 1:500 dilution for 1–2 hr at RT. Slides were then washed thrice with 1X PBS and mounted in Prolong Gold Antifade with DAPI (Cell Signaling Technology).

Sample Preparation for Whole-Exome and Targeted Sequencing

Samples were prepared, sequenced and analyzed as described earlier (Bertotti et al., 2015). DNA was extracted from cells and xenograft tissues using the Qiagen DNeasy blood and tissue mini kit (Qiagen, CA). Fragmented genomic DNA from tumor and normal samples was used for Illumina TruSeq library construction (Illumina, San Diego, CA). Exonic regions or targeted regions (targeted genes—Table S3) were captured in solution using the Agilent SureSelect v.4 kit (Agilent, Santa Clara, CA) according to the manufacturer's instructions as previously described (Bertotti et al., 2015; Sausen et al., 2013). Paired-end sequencing, resulting in 100 or 150 bases from each end of the DNA fragments for the exome or targeted sequencing libraries, respectively, was performed using Illumina HiSeq 2000/2500 instrumentation (Illumina, San Diego, CA).

Analysis of Next-Generation Sequencing Data and Identification of Putative Somatic Mutations

Genetic alterations were identified using VariantDx custom software for identifying mutations in samples by removing background mutations by using a specified control as previously described (Bertotti et al., 2015; Sausen et al., 2013). Before mutation calling, primary processing of sequence data both the sample and for control was performed using Illumina CASAVA software (version 1.8), including masking of adaptor sequences. Sequence reads were aligned against the human reference genome (version hg19) using ELAND with additional realignment of select regions using the Needleman-Wunsch method. Candidate mutations, consisting of point mutations, insertions, and deletions, were then identified using VariantDx across the whole exome. VariantDx examines sequence alignments of samples against a control while applying filters to exclude alignment and sequencing artifacts. Potential alterations were compared with mouse sequences from experimentally obtained mouse whole-exome and targeted sequence data as well as the reference mouse genome (mm9) to remove mouse-specific variants. Amplification analyses were performed using the digital karyotyping approach (Leary et al., 2007) by comparing the number of reads mapping to a particular gene with the average number of reads mapping to each gene in the panel, along with a minor allele fraction analysis of heterozygous single nucleotide polymorphisms contained within each gene. Candidate somatic mutations were further filtered on the basis of gene annotation to identify those occurring in protein coding regions. Functional consequences were predicted using snpEff and a custom database of CCDS, RefSeq and Ensembl annotations using the latest transcript versions available on hg19 from UCSC (<https://genome.ucsc.edu/>). Predictions were ordered to prefer transcripts with canonical start and stop codons and CCDS or Refseq transcripts over Ensembl when available.

Generation of Box Plots for Analysis of Methylation Data from NSCLC Samples from TCGA for Correlation with Mutational and Smoking Status

Infinium methylation array data for primary lung adenocarcinoma and adjacent normal tissue was used to examine the methylation status of the CSC methylated genes in these tumors and normal samples (depicted in Figures 7F and 7G) when separated by mutational status and/or smoking history. Mutational data for generating Figures 7F and 7G was obtained from previously published data

(Cancer Genome Atlas Research Network, 2014) and methylation data for the same samples was obtained from the Cancer Genome Atlas (TCGA; <http://tcga.cancer.gov/dataportal>). A key set of CSC-methylated genes from Table S2 used earlier for analysis in Figures 7E and 7F were used to test the methylation status of these genes in the primary normal and tumor tissues. Box plots are made by plotting all the CpG-island probes mapped to \pm 1500bp of TSS of these genes. Data are presented as the average of all probes in all patients within the respective group.

Generation of Box Plots for Analysis of Methylation Data from COPD (Chronic Lung Obstructive Disease) and IPF (Idiopathic Pulmonary Fibrosis) Samples

Methylation data for analysis of COPD samples depicted in Figure S3B were obtained from the Lung Genomics Research consortium data portal (<http://www.lung-genomics.org>). Methylation in these samples was generated using the CHARM platform. The normalized CHARM values and the meta-data was downloaded from Lung Genomics Research Consortium data portal. CHARM data was annotated using RefSeq Genes (NCBI36/gh18) and Canonical CpG Islands (NCBI36/gh18) from UCSC Genome Bowser (genome.ucsc.edu). Box plots were generated by plotting all the CpG-island probes mapping to \pm 1500bp of TSS of 15month CSC-methylated genes from rep1 and rep2 (Table S1) in COPD and normal samples.

Methylation data for analysis of COPD, IPF and lung cancer samples depicted in Figure S3C were obtained from the GEO database the GEO database (GSE63704). Infinium 450K methylation data was available for these samples. Box plots were generated by averaging all the CpG-island probes mapped to \pm 1500bp of TSS for each individual and plotting the average β values of promoter CpG-island probes. Python package seaborn was used for all of the box plots analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

All western blot and ChIP data are presented as the mean \pm standard deviation (SD). These data are evaluated by two-tailed t test using Excel software. Box and whisker plots (Figures 7C, 7D, S4H, and S6) were generated in GraphPad Prism® (GraphPad Software, Inc.) and analyzed using the Wilcoxon matched-pairs signed rank test. Box plots in Figures 7F, 7G, S3B, and S3C were generated using Python package seaborn and p values were calculated using Welch's t-test. p values of < 0.05 were considered statistically significant for all tests. Where appropriate, figure legends define n, which indicates the number of individual biological replicates.

DATA AVAILABILITY

All DNA methylation, gene expression arrays, ChIP-seq, Exome-seq and targeted Deep-seq data has been deposited in GEO under the series reference GSE101864.

Supplemental Information

Chronic Cigarette Smoke-Induced Epigenomic Changes

Precede Sensitization of Bronchial Epithelial Cells

to Single-Step Transformation by *KRAS* Mutations

Michelle Vaz, Stephen Y. Hwang, Ioannis Kagiampakis, Jillian Phallen, Ashwini Patil, Heather M. O'Hagan, Lauren Murphy, Cynthia A. Zahnow, Edward Gabrielson, Victor E. Velculescu, Hariharan P. Easwaran, and Stephen B. Baylin

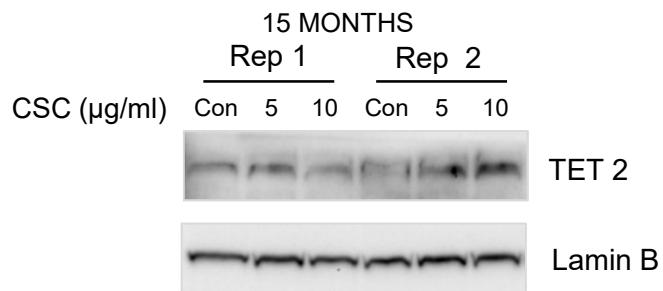
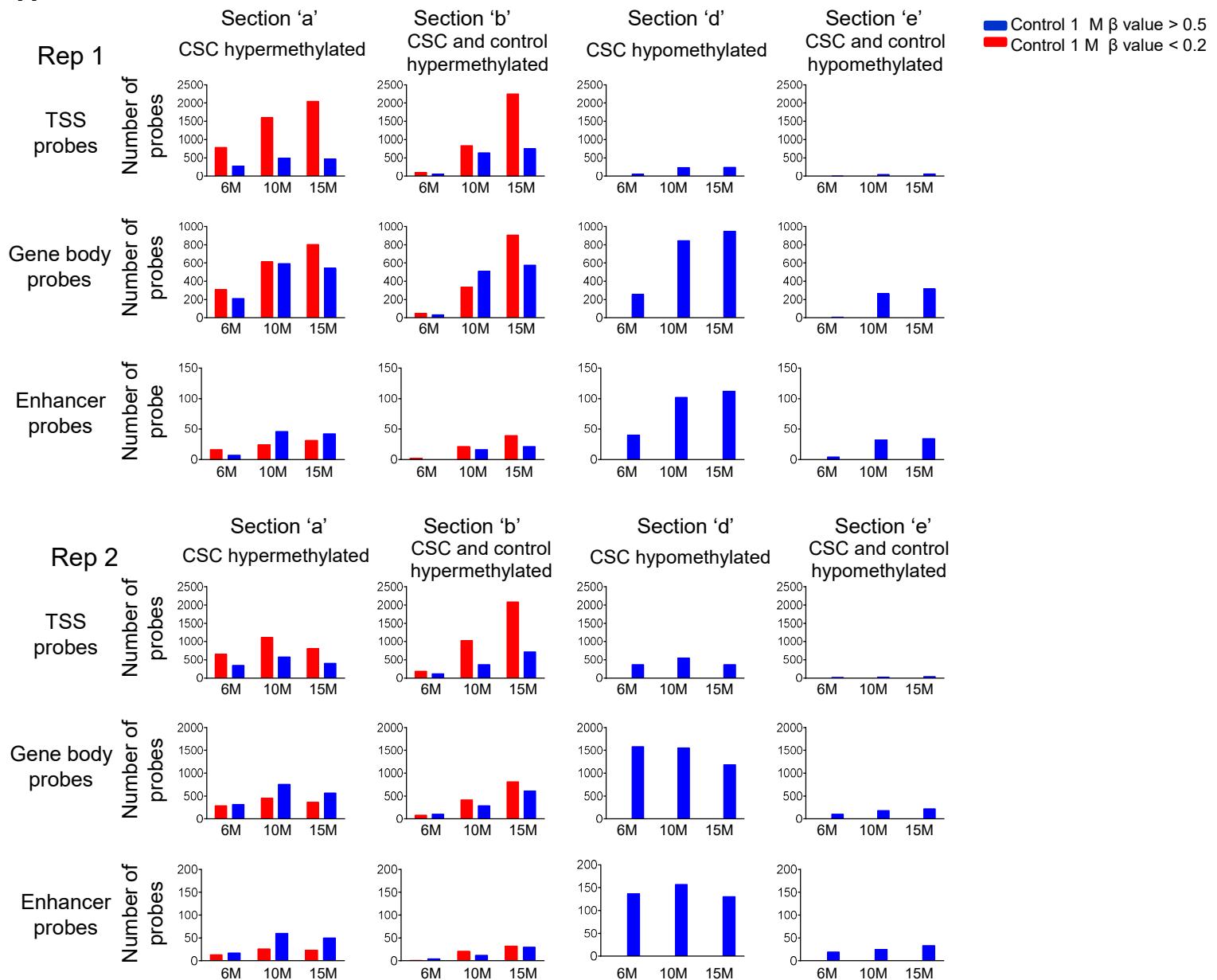
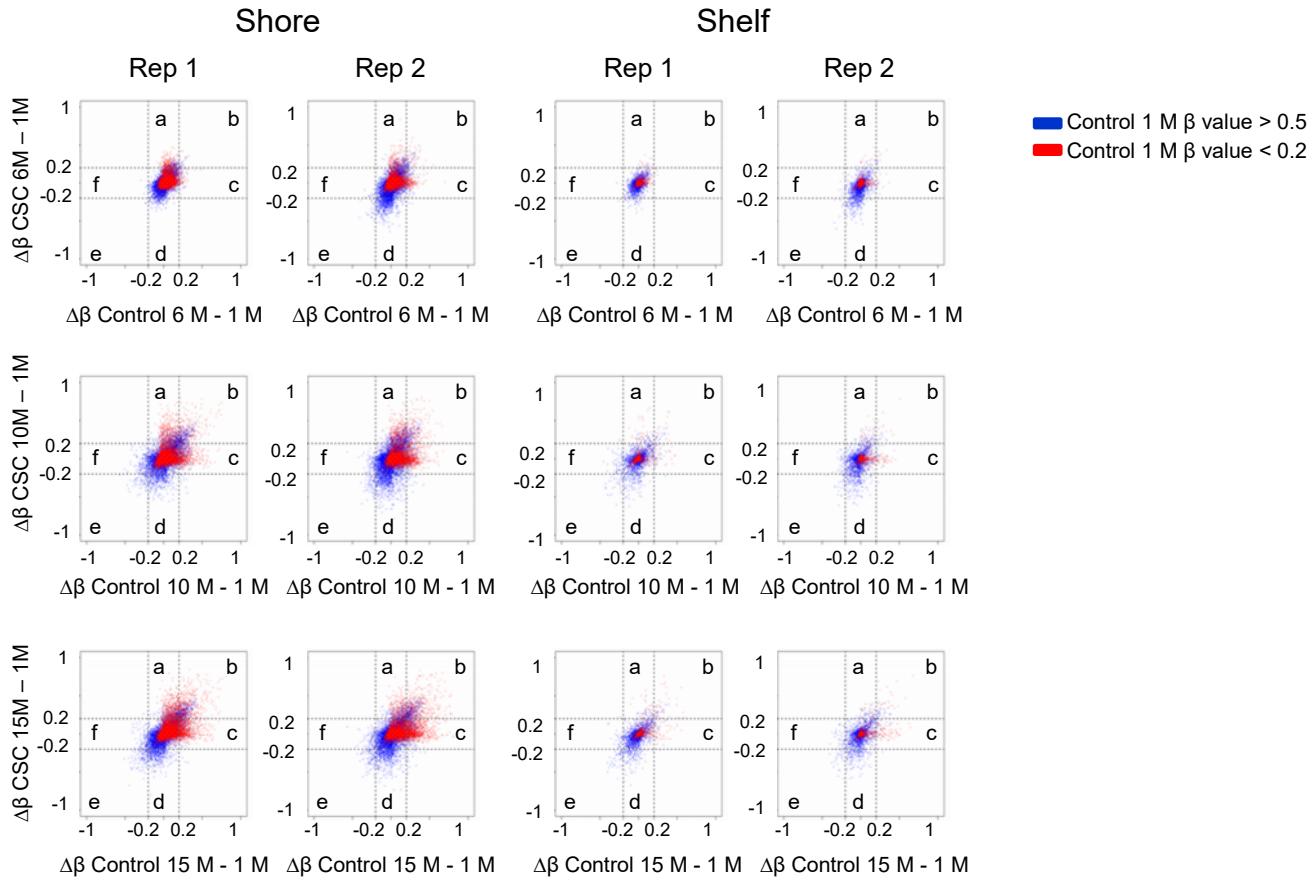


Figure S1, Related to Figure 1: Expression of TET2 following exposure to CSC for 15 months.

Immunoblots showing expression of TET 2 in nuclear extracts from cells treated with DMSO (Con), 5 µg/ml and 10 µg/ml CSC for 15 months.

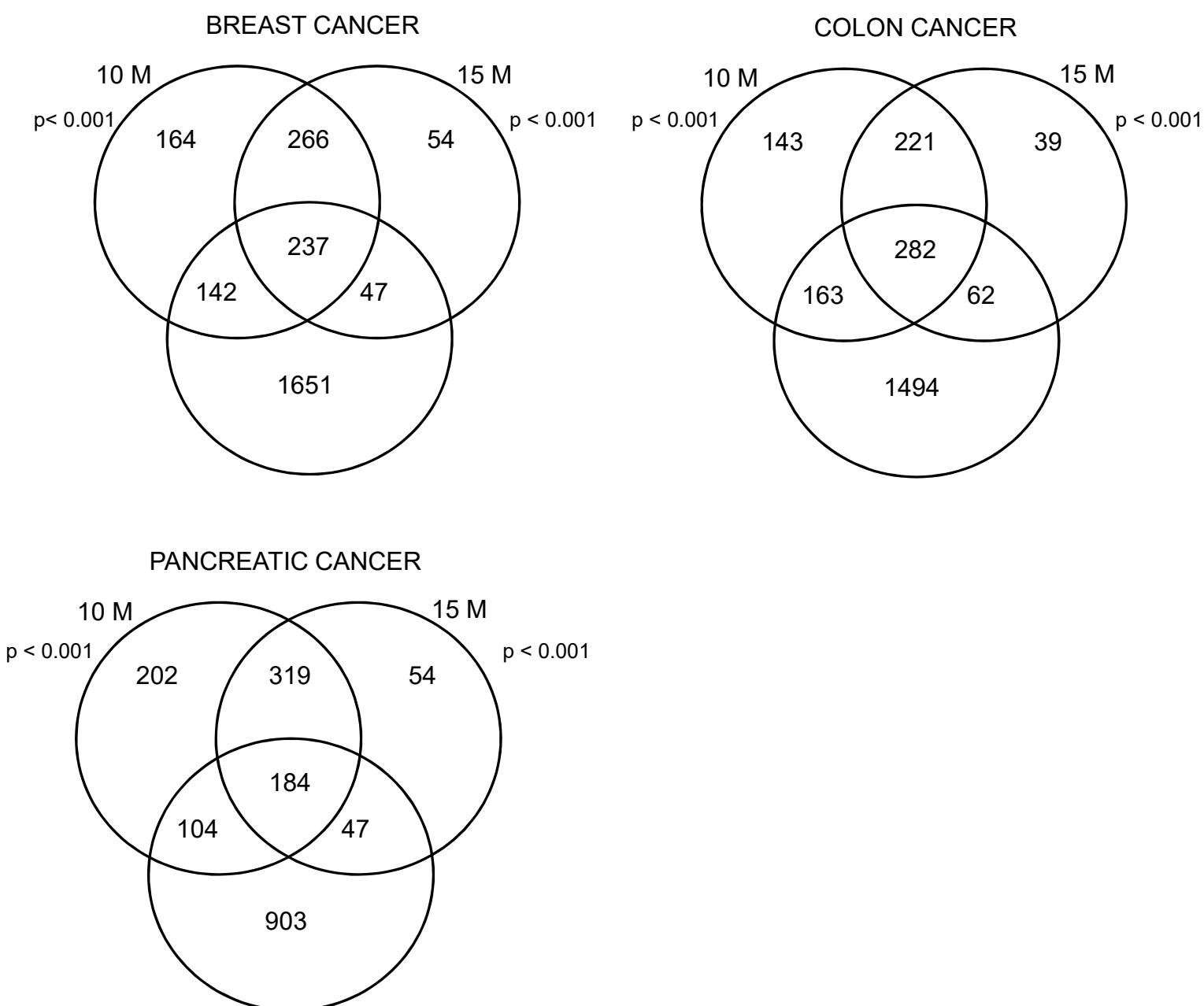
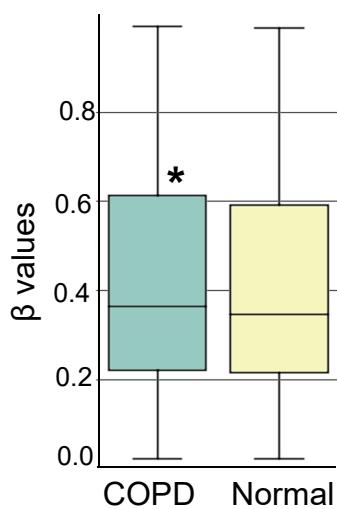
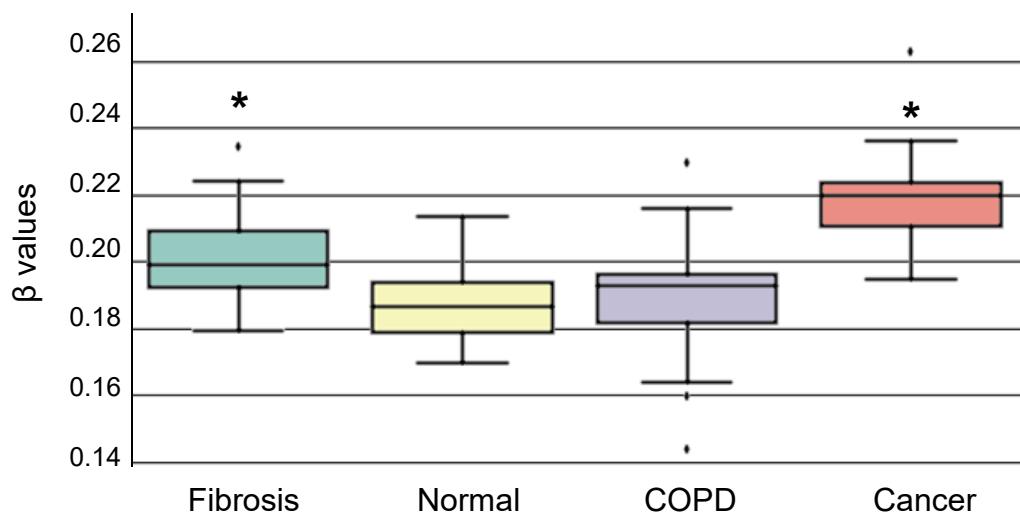
A

B

- a: CSC hypermethylated;
- b: CSC and control hypermethylated;
- c: Control hypermethylated;
- d: CSC hypomethylated;
- e: CSC and control hypomethylated;
- f: Control hypomethylated

Figure S2, Related to Figure 2: Progressive changes in methylation in control and CSC treated cells.

(A) Bar graphs showing change in number of probes undergoing CSC-induced hypermethylation (section ‘a’ and ‘b’) and hypomethylation (section ‘d’ and ‘e’) at TSS, gene body and enhancer regions at the indicated time points in replicate 1 and 2. Hypermethylation in probes with low basal methylation levels is shown in red and represents probes with starting β value < 0.2 . Hypomethylation in probes with high basal methylation is shown in blue and represents probes with starting β value > 0.5). (B) Scatter plots showing changes with time in methylation at shelf and shore regions. Changes in β values for all CpG island probes within the respective regions from 1 month (1 M) to the respective end point (6 M, 10 M and 15 M) are shown in control (X-axis) and CSC treated (Y-axis) cells. Red dots: probes with a β value < 0.2 in control cells at 1 month; blue dots: probes with β value > 0.5 in control cells at 1 month. Rep 1: replicate 1; Rep2: replicate 2.

A**B****C**

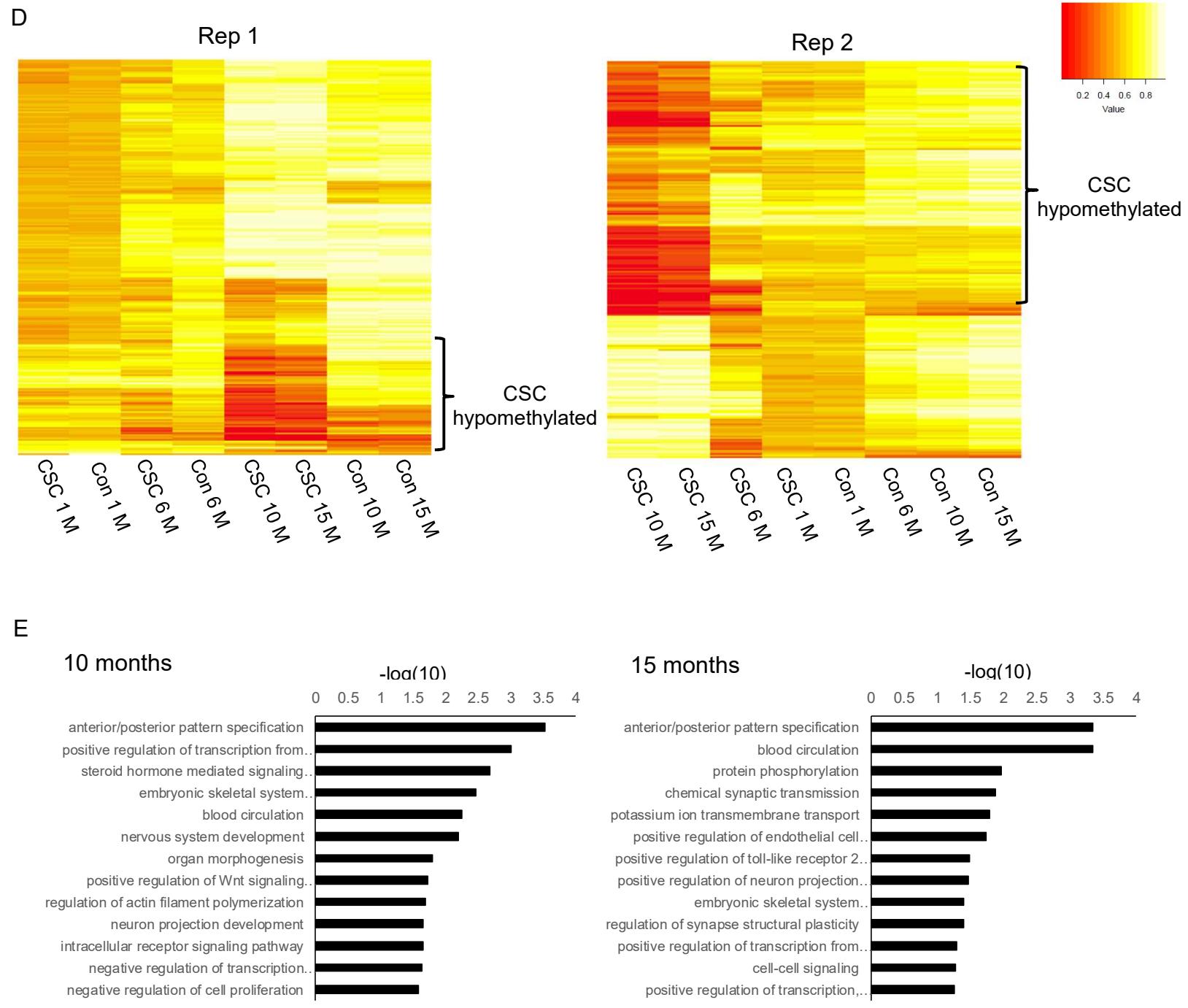


Figure S3, Related to Figure 3: Analysis of methylation levels of CSC methylated genes in cancers and smoking associated lung diseases and analysis of genes showing TSS-specific hypomethylation in CSC treated cells.

(A) Venn diagrams showing overlap of methylated genes in 10 μ g/ml CSC treated cells at 10 and 15 months with genes hypermethylated in breast, colon and pancreatic cancer samples from TCGA databases. To obtain genes hypermethylated in the indicated cancers, we picked probes methylated in at least 25% of tumor samples and identified corresponding genes as described in STAR methods. (B and C) Methylation levels of CSC-methylated genes following 15 months of exposure in lung tissue samples obtained from normals ($n = 72$) and COPD patients ($n = 315$) from the Lung Genomics Research Consortium (LGRC) (B) and lung tissues obtained from patients with pulmonary fibrosis ($n = 37$), COPD ($n = 32$), lung cancer ($n = 17$) and normal lung ($n = 43$) obtained from GEO database (C). Box plots in B and C show average β values of promoter CpG island probes of 15 month CSC methylated genes per person. Top and bottom of Box plots represent the 25th and 75th percentiles (interquartile range; IQR), with midlines indicating the median values. The whiskers extend to points that lie within 1.5 IQR's of the lower and upper quartile. Observations that fall outside this range are displayed independently. * $p < 0.05$ comparing disease state versus normal by Welch's t-test. (D) Heat maps of β values of top 2.5% variable CpG-island probes mapped within ± 1500 bp from TSS in control (Con) and 10 μ g/ml CSC exposed cells from 1 month to 15 months in replicates 1 and 2. Only probes with a starting β value of > 0.5 in control cells at 1 month were selected. (E) Top enriched biological processes in genes hypomethylated by CSC at 10 M and 15 M of treatment. M: months.

Table S1, Related to Figure 3: CSC methylated genes from Replicate 1 (Rep1) and Replicate 2 (Rep 2) over the time course of the 15 month treatment period. Provided as an Excel File.

Table S2, Related to Figure 3 and Table 1: Functional categories of genes methylated by CSC treatment over time.

Time point	Functional categories	Genes Involved
6 months	Developmental protein	BMP3, GSX1, TBX5, ZIC2, MSX1, NEUROD1, NEUROG1, POU3F3, POU4F1, OLIG1, WIF1, PAX6.
	Homeobox	GSX1, ONECUT2, PITX2, POU3F3, POU4F1, POU4F2, HOXA9, MSX1, PROX1, ZEB1, ZEB2.
	Transcription factor	GATA4, NEUROD1, POU3F3, POU4F2, TBX21, MSX1, TBX5, IRF8, ONECUT2, PITX2, ZEB2.
	N-linked glycosylation	ADAMTS16, ADAMTS9, GFRA1, ST6GALNAC5.
10 months	Homeobox	ALX4, BARX1, ISL2, NKX2-4, NKX6-1, NKX6-2, POU4F1, CDX2, DBX2, EMX1, MSX1, ONECUT2, ONECUT3, RAX, SHOX2, HOXC12, IRX4, PROX1, ZEB1, ZEB2.
	Transcription factor	GATA4, IKZF1, ISL2, NKX2-4, NKX6-1, NKX6-2, POU4F1, LHX8, ID4, PROX1, SALL1, ZEB2, TBX5, TAL1, FOXB2, FOXI3, GFI1, MSX1, NEUROD1, PITX2.
	Glycoproteins/signaling	ADAMTS16, ADAMTS18, ADAMTS9, PR83, GPR149, GPRC5B, IHH, ST6GALNAC, EPHA5, ICAM5, ITGB3, GFRA1, BMP3, BMP8B, PCDH8, WIF1, SFRP2, SFRP5, SLIT1, SLIT2, WNT9B.
	Embryonic morphogenesis	ALX4, TBX5, MSX1, NEUROG1, NEUROD1, SLIT2, MAFB, ZIC2, GFI1, ZEB1, ZEB2.
15 months	Glycoproteins/signaling	ADAMTS16, ADAMTS18, ADAMTS9, BMP3, ST6GALNAC5, GPR83, GPRC5B, GFRA1, ITGA8, ICAM5, PTCHD2, RPRM, GALR1, GALR2, SFRP2, SFRP5, IHH.
	Transcription factor	GATA4, GATA5, GSX1, ISL2, LHX8, POU4F1, NKX6-1, TBX5, CDX2, FOXP1, FOXI3, MSX1, MYOD1, NEUROD1, SALL1, ZEB2, TOX3, ZIC2, OLIG3, PROX1, PITX2, SALL1, SHOX2, MAFB.
	Homeobox	ALX4, GSX1, ISL2, LHX8, NKX6-1, POU4F1, CDX2, EMX1, MSX1, ONECUT2, ONECUT3, PROX1, SHOX2, ZEB1, ZEB2, PITX2.
	Embryonic morphogenesis	ALX4, TBX5, ZIC2, FOXP1, ITGA8, MSX1, NEUROD1, PROX1, MAFB, ZEB2.

^aCSC methylated genes at the indicated time points from Replicate 1 and 2 were analyzed using the ‘Functional Annotation’ tool in DAVID Bioinformatics resources database. All these genes are also methylated in lung adenocarcinoma (LUAD) and lung squamous cell carcinomas (LUSC) samples analyzed from the TCGA database. To obtain genes hypermethylated in LUAD and LUSC, we picked probes methylated in at least 25% of tumor samples and identified corresponding genes as described in STAR methods. Genes in bold are those identified to drive the key signaling pathways described in Table 1.

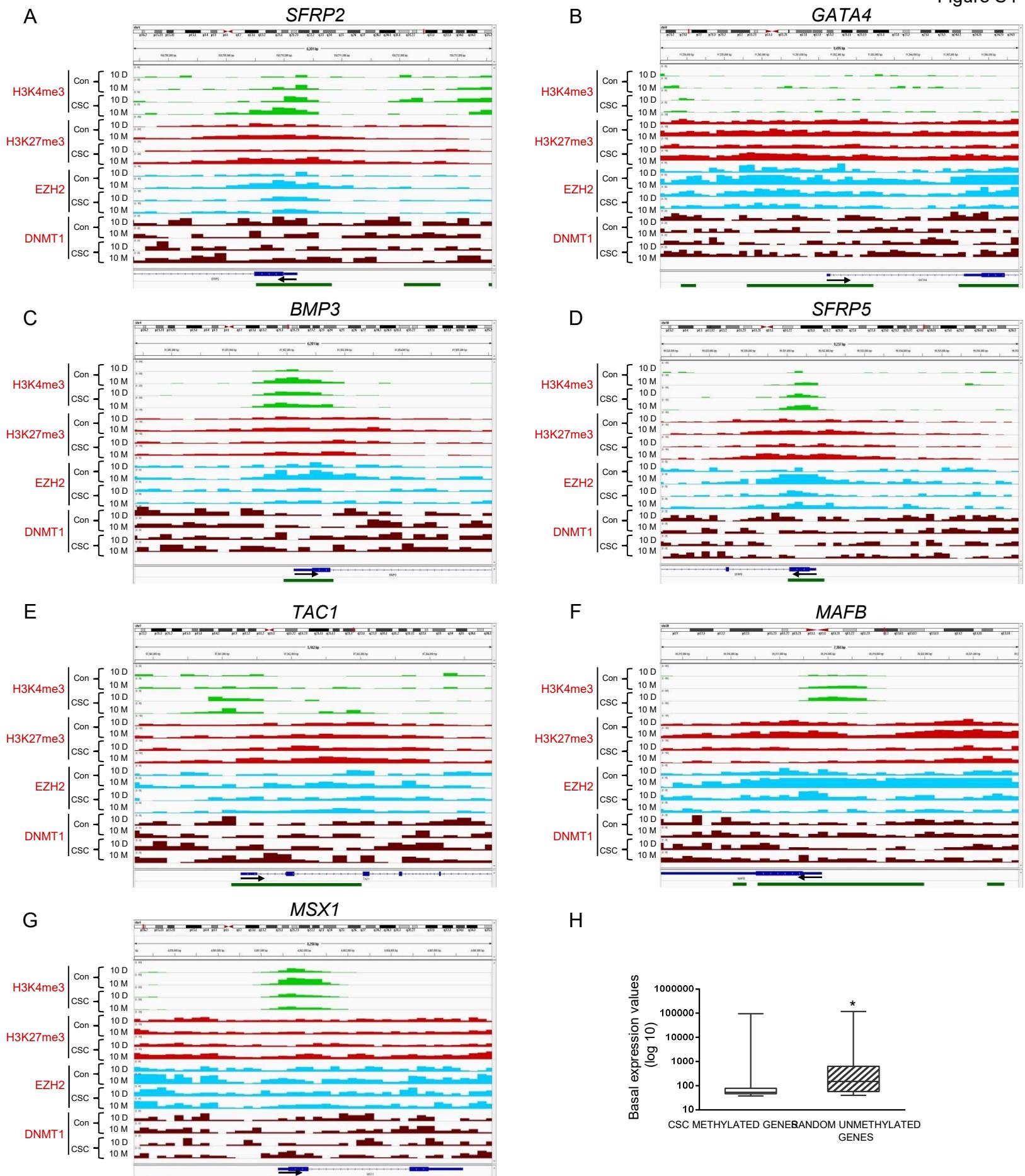


Figure S4, Related to Figure 4: Enrichment of H3K4me3, H3K4me27, EZH2 and DNMT1 at select genes. (A-G) Plots showing enrichment of H3K4me3, H3K4me27 and binding of EZH2 and DNMT1 analyzed by ChIP-seq at promoters of select genes visualized using IGV program. CpG islands are marked in green and arrows denote TSS and direction of transcription. D: days; M: months. (H) Average basal expression of CSC methylated and random unmethylated genes used for analysis in Figure 4C. Box plots represent the 25th and 75th percentiles, with midlines indicating the median values. Whiskers extend to the minimum/maximum values of the data sample. * $p < 0.01$ by Wilcoxon rank test.

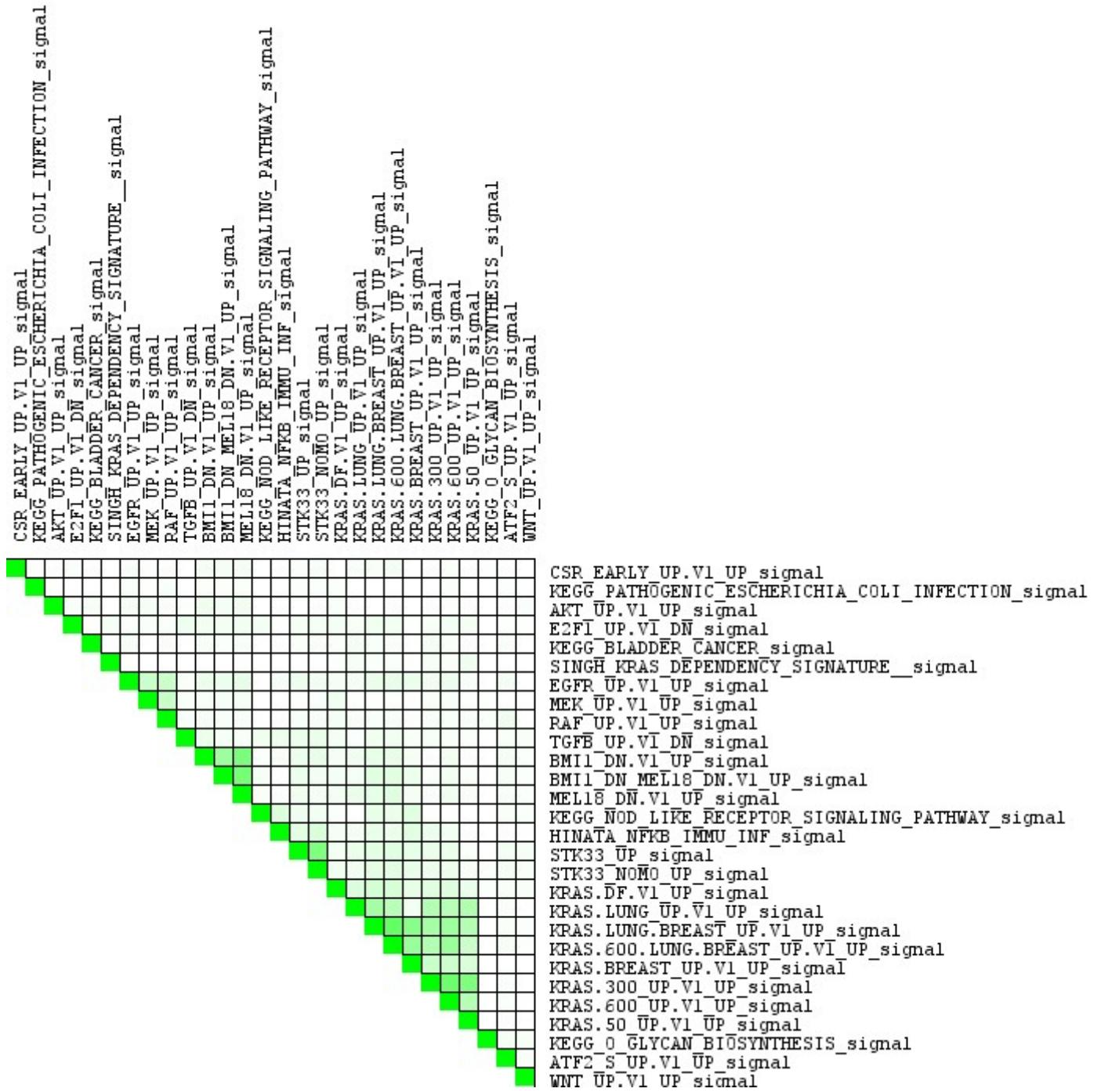
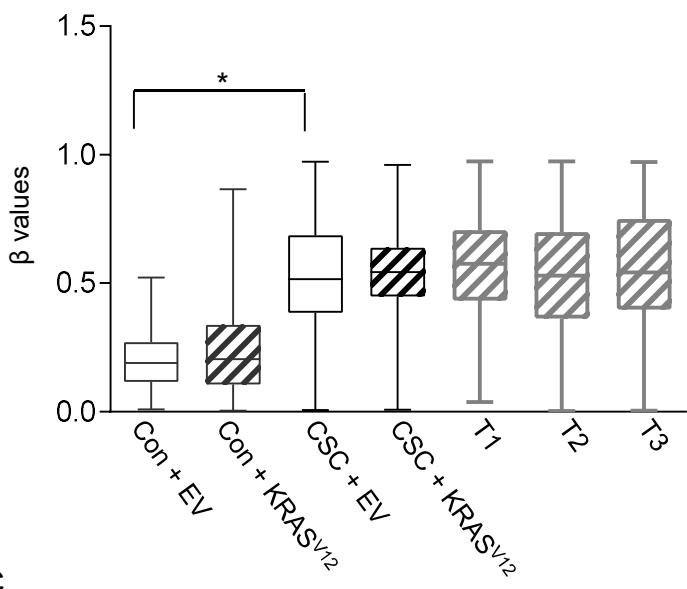


Figure S5, Related to Figure 5: Leading edge analysis of most highly enriched pathways (NES > 1.8, FDR < 1%) in cells exposed to CSC for 15 months

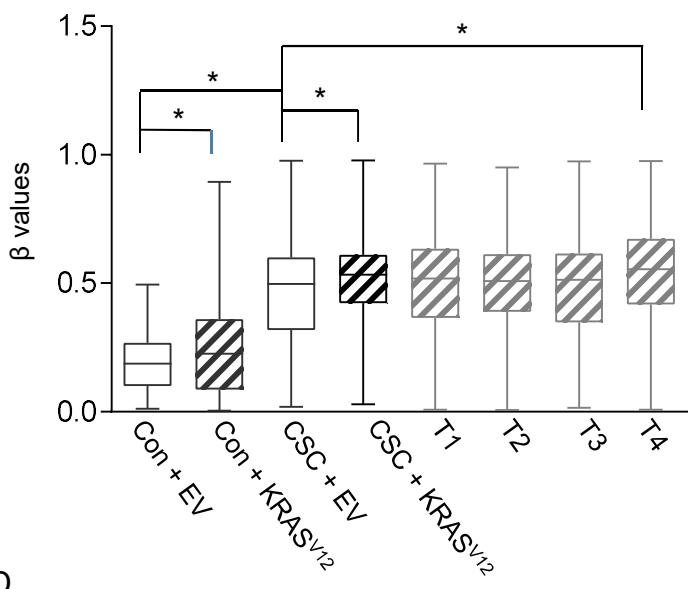
A

Replicate 1



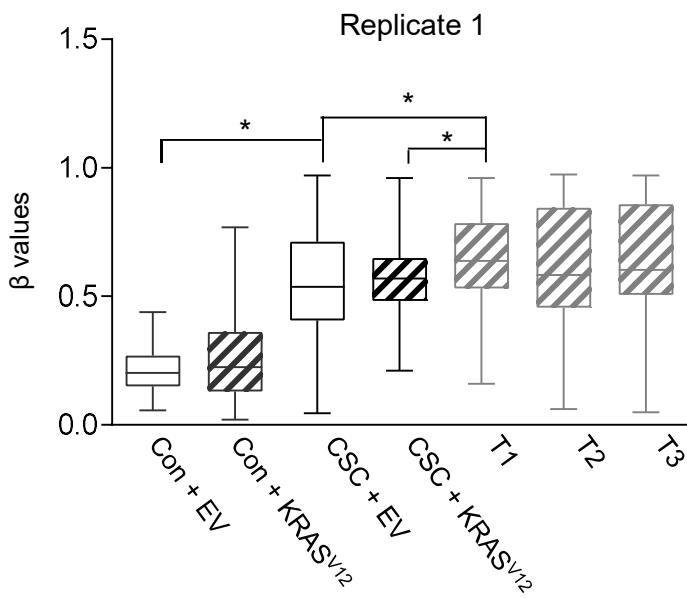
B

Replicate 2



C

Replicate 1



D

Replicate 1

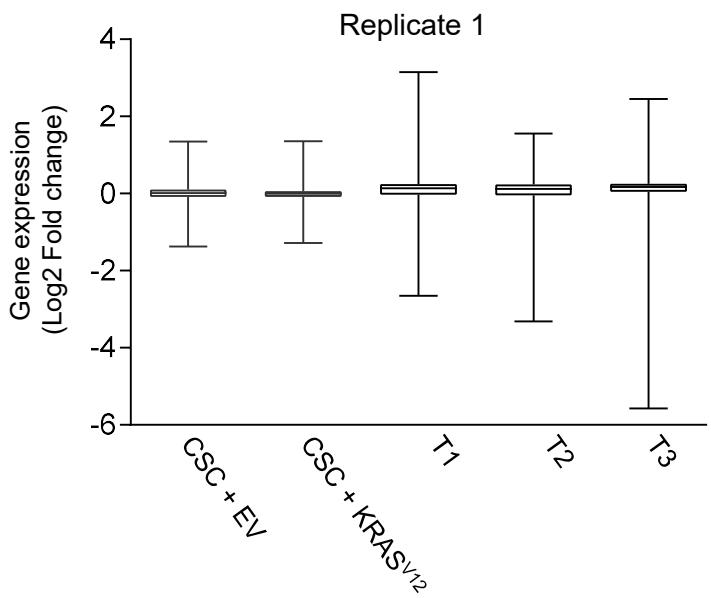


Figure S6, Related to Figure 7: Methylation and gene expression changes in control and CSC treated cells with empty vector (EV) or KRAS^{V12} and in tumor xenografts. (A and B) Methylation levels of 15 month CSC methylated genes in empty vector (EV) and KRAS^{V12} expressing control and CSC treated cells and xenografts harvested 4 weeks after appearance of tumors from replicate 1 (A) and replicate 2 (B). β values of promoter CpG islands probes in genes methylated by CSC at 15 months are plotted. Only probes with a starting β value of ≤ 0.4 in EV containing control cells were selected (C) Methylation changes in key CSC-methylated genes from Table S2 in empty vector (EV) and KRAS^{V12} expressing control and CSC treated cells and xenografts (T1-T3) harvested 4 weeks after appearance of tumors from replicate 1. β values of promoter CpG island probes are plotted. Only probes with a starting β value of ≤ 0.4 in EV containing control cells were selected. (D) Box plots showing average log2 fold change in gene expression of CSC-methylated genes from Table S2 in the EV and KRAS^{V12} expressing CSC cells and xenografts as compared to their respective control cells expressing EV or KRAS^{V12}. All box plots were generated using GraphPad Prism software. Top and bottom of boxplots represent the 25th and 75th percentiles, with midlines indicating the median values. Whiskers extend to the minimum/maximum values of the data sample. T: xenografts obtained after injecting CSC + KRAS^{V12} cells in mice. * $p < 0.01$ by Wilcoxon rank test.

Table S3, Related to Figure 7 : Sequencing analysis (Exome-seq and targeted deep-seq) of HBEC cells and xenografts. Provided as an Excel file

Table S4, Related to STAR methods: Primer sequences used in the study.

<u>Primers for ChIP</u>		
Gene	Forward Primer	Reverse primer
<i>SFRP2</i>	GCAAGGGAGGAGCCAATGAA	AACGCAAAACCAGACCCAGA
<i>GATA4</i>	AGCTTCCGCACAGTCCC	CTGTGGGAGTCACGTGCAA
<i>BMP3</i>	TCCCTGGAAAAGGCAATCGAG	GCGCGAGGGACTAGGACT
<i>NANOG</i>	TTGTTGCTGGGTTGTCTTCAGGTT	CGTCTACCAGTCTCACCAAGGCC
<i>ACTB</i>	CTCCCTCCTCCTCTTCCTCA	TCGAGCCATAAAAGGCAACTT
<u>Primers for qRT-PCR</u>		
Gene	Forward Primer	Reverse primer
<i>SFRP2</i>	TCGGTGCTGTGGCTCAAAG	GCTCCCCACCCCTGTTCTGT
<i>GATA4</i>	AGGAGGAAGGAGGCCAGCCTA	TATTGGGGGCAGAACGACGGA
<i>BMP3</i>	GGCAGCAGCAGAACTCTTG	ACAGAAATACAGTGTGGCAGACA
<i>NANOG</i>	CAGTCTGGACACTGGCTGAA	CTCGCTGATTAGGCTCCAAC
<i>ACTB</i>	GAGCACAGAGCCTGCCTT	TCATCATCCATGGTGAGCTGG