Published in final edited form as:

Cancer Cell. 2017 September 11; 32(3): 360–376.e6. doi:10.1016/j.ccell.2017.08.006.

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], **Lauren Murphy**[1], **Cynthia A Zahnow**[1], **Edward Gabrielson**[4], **Victor E Velculescu**[1], **Hariharan P Easwaran**[1,5], and **Stephen B Baylin**[1,5,6]

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

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

[3]Medical Sciences, Indiana University School of Medicine, Bloomington, IN 47405, USA Melvin and Bren Simon Cancer Center, Indianapolis, IN 46202, USA

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

SUMMARY

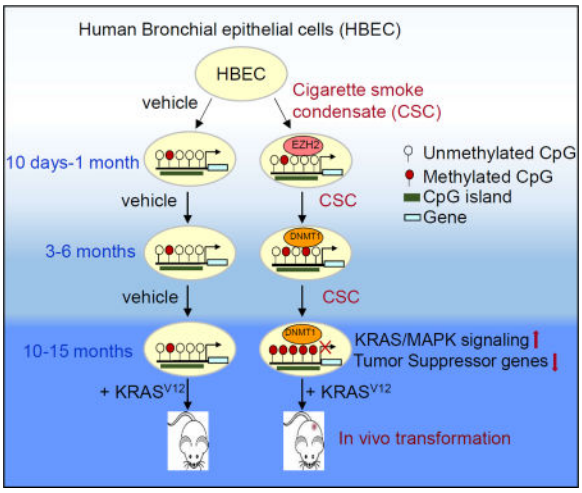
We define how chronic cigarette smoke-induced time-dependent epigenetic alterations can sensitize human bronchial epithelial cells for transformation by a single oncogene. The smoke-induced chromatin changes include initial repressive polycomb marking of genes, later manifesting abnormal DNA methylation by 10 months. At this time, cells exhibit epithelial to mesenchymal changes, anchorage-independent growth and upregulated RAS/MAPK signaling with silencing of hyper-methylated genes, which normally inhibit these pathways and are associated with smoking related NSCLC. These cells, in the absence of any driver gene mutations, now transform by introducing a single *KRAS* mutation and form adeno-squamous 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.

eTOC blurb/In Brief

[5]To whom all correspondence should be addressed: **Stephen B. Baylin**, Department of Oncology, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, sbaylin@jhmi.edu, **Hariharan P Easwaran**, Department of Oncology, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD 21287, USA. heaswar2@jhmi.edu.
[5]Lead Contact, **Stephen B. Baylin**, Department of Oncology, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD 21287, USA. sbaylin@jhmi.edu.

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 supervision of H.P.E. E.G classified xenograft histopathology. J.P. and V.E.V. performed and analyzed exome sequencing. HOH assisted with data analyses and manuscript review. C.A.Z and L.M assisted with in vivo experiments.



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

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, 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 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 for human cells. For the present study, we use human bronchial epithelial cells (HBECs), which are initially immortalized via 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., 2013; Sato et al., 2006). 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 upregulate 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) that is 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 1 A–C). While total nuclear protein levels of the maintenance DNA methylation enzyme, DNMT1, increased initially after CSC treatment, the levels decreased by one 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 above 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–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 (PCA) 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 (Figure 2A and 2B). For the latter cells, these changing methylation patterns suggest many probable stochastic events occurring heterogeneously (Figures 2A–C). 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 to 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 circles versus green circle - Figure 2A).

At CpG islands within ± 1500 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 (Figure 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’ (Figure 2D **- blue probes and** S2A **- blue bars**). Putative enhancer regions showed both gains and losses in methylation with a trend towards 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 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 passaged at 1 month, 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 times points using the criteria detailed in the STAR methods section. 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, 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 adeno- (LUAD) and squamous cell (LUSC) carcinomas (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 the above 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 (COPD) 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 (ESC) and adult stem cells. It is hypothesized that during tumor initiation and progression, a quantitative shift occurs away from this chromatin state towards a more repressive one marked by DNA methylation (Easwaran et al., 2012; Ohm and Baylin, 2007; Schlesinger et al., 2007; Widschwendter et al., 2007). Yet, 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 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 house keeping 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 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 time (Figure 4B).

The above studies are well reflected in genome wide ChIP-seq studies (Figures 4C **and** S4A–G), which assessed enrichment of the above mentioned 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 to 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 get 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 to 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 as compared to the control cells. The bivalent nature of the CSC-methylated genes is well appreciated when these genes are compared to 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–G). All of the above 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 (Figure 5A and B). 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 (Figure 5A and B). Thereafter, when abnormal DNA methylation is the dominant epigenetic change by 10 to 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 (Figure 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 above 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 TGF-β superfamily, which normally represses RAS-MAPK signaling (Kraunz et al., 2005). This gene is very frequently promoter hypemethylated in 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; Revet et al., 2010), the other leading oncogenic pathway upregulated in our studies (Figure 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.

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 towards a pre-tumorigenic state. First, by 10 to 15 months, the cells become anchorage independent and form clones in soft agar (Figure 6A and 6B). At this time point, when these CSC exposed cells are grown in 5% FBS containing medium, they become more elongated, suggestive of an 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 (Figure 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 the above 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 HBEC cells requires introduction of at least three genetic events: shRNA-mediated downregulation of TP53 paired with introduction of both *KRAS* mutations and CMYC overexpression (Sato et al., 2013). We introduced individually either $KRAS^{V12}$ (Figure 6F) or downregulate 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 impact 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, Hedgehog and VEGF 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 appear 3 weeks post injection and grow to an average size ranging from 500–900 mm^3 (Figure 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 adeno-squamous 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 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 (Figure 7C, 7D, S6C and S6D). Gene expression analysis in xenografts and in CSC treated cells expressing $KRAS^{V12}$ showed enrichment of KRAS, WNT and Hedgehog pathways (Figure 7E). Upregulation of LEF1 signaling suggests an important role of WNT pathway in driving tumor formation as WNT signaling through LEF1 mediates metastasis of lung adenocarcinoma (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 HBEC cells, 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 HBEC cells, 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.

CSC induced methylation changes associate with smoking related non-small cell lung cancers with a bias towards those with KRAS mutations

The above bias to KRAS-mediated transformation raises interesting questions with respect to smoking and the mutation distributions among sub-groups of primary NSCLC. First, analyses of TCGA data from lung adenocarcinomas (Cancer Genome Atlas Research, 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 lung adenocarcinomas with *EGFR* mutations compared to 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 to 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 tumors 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, like *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 to, rather than senescing or dying, in responding to signaling from mutated driver genes, like *KRAS* (Baylin, 2012; Dimauro and David, 2010).

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 are 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 ROS 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 which 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 suggests their key role in the development of smoking-associated lung cancer. Third, as discussed earlier, our ChIP data at selected candidate genes and the genome-wide 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 progression (Helman et al., 2012).

An 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 to 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, DNMT’s 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 DNMT’s 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 the above 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 the involved genes are key for driving these pathways. Indeed, functional studies have shown these genes are important for normally preventing abnormal RAS-MAPK (Kraunz et al., 2005) and WNT pathway signaling (Revet et al., 2008; Revet et al., 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 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 HBEC cells need a combination of at least three oncogenic events to form tumors in mice (Sato et al., 2013; Sato et al., 2006). 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 fifteen month time span in our studies mimics the typical long course of smoking 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

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 2×10^6 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 PBS 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% 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 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 ± 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 ± 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 ± 1500bp of TSS.

To obtain a list of CSC methylated genes, we first chose probes that had a starting β value of ≤ 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) ≤ 0.2 AND (Δβ Con (6 M-1 M) ≤ 0.2 and Δβ CSC 6 M - Con 6 M ≥ 0.25),

10 months: Con(1 M) ≤ 0.2 AND (Δβ Con (10 M-1 M) ≤ 0.2 and Δβ CSC 10 M - Con10 M ≥ 0.25),

15 months: Con(1 M) ≤ 0.2 AND (Δβ Con (15 M - 1 M) ≤ 0.2 and Δβ CSC 15 M-Con 15 M ≥ 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) ≥ 0.5 AND (Δβ Con (6 M -1 M) > - 0.2 AND Δβ CSC 6 M-Con 6 M ≤ −0.2

10 months: Con(1 M) ≥ 0.5 AND (Δβ Con (10 M -1 M) > - 0.2 AND Δβ CSC 10 M-Con 10 M ≤ - 0.2

15 months: Con(1 M) ≥ 0.5 AND (Δβ Con (15 M - 1 M) > - 0.2 AND Δβ CSC 15 M - Con 10 M ≤ - 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 ± 1500bp of TSS that had a starting β value of ≤ 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 ± 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 Δβ cut off of ≥ 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 col umn 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 4B, for the CSC methylated and a random set of unmethylated genes, we obtained the ChIP-seq reads in 200bp bins (obtained using BedTools) ± 5Kb 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 $KRASV^{12}$ 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 (Figure 7 and Figure S4). 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 on 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 PBS 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 anitbodies (E-cadherin and Vimentin) were added at 1.100 dilution overnight at 4°C. Slides were then washed thrice with 1X PBS and flurochrome-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, 2014) and methyaltion 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 ± 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 generate 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 ± 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 ± 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 ± standard deviation (SD). These data are evaluated by two-tailed t test using Excel software. Box and whisker plots (Figures 7C, 7D, S4H and Figure S6) were generated in GraphPad Prism® (GraphPad Software, Inc.) and analyzed using the Wilcoxon matched-pairs signed rank test. Box plots in Figure 7F, 7G and 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 microarray, methylation and ChIP-seq data has been deposited in GEO under the series reference GSE101864.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute under award number R01CA043318 (S.B.B.), R01CA185357(H.E.), National Institutes of Health R01CA121113 (V.E.V.), P30CA006973 (S.B.B.) (V.E.V.), U10CA180950 (V.E.V), NIEHS, R01ES011858 (S.B.B), R01ES023183 (H.M.O'H). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. 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.). 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. Flight Attendant Medical Research Institute (M.V.). We thank Kathy Bender for help with manuscript preparation.

References

Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol. 2013; 14:R21. [PubMed: 23497655]

Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG. Chromatin signatures of pluripotent cell lines. Nature cell biology. 2006; 8:532–538. [PubMed: 16570078]

Baylin SB. The cancer epigenome: its origins, contributions to tumorigenesis, and translational implications. Proceedings of the American Thoracic Society. 2012; 9:64–65. [PubMed: 22550245]

Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. Nature reviews Cancer. 2011; 11:726–734. [PubMed: 21941284]

Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nature reviews Cancer. 2006; 6:107–116. [PubMed: 16491070]

Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, Grimes MJ, Harms HJ, Tellez CS, Smith TM, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer research. 2002; 62:2370–2377. [PubMed: 11956099]

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006; 125:315–326. [PubMed: 16630819]

Bertotti A, Papp E, Jones S, Adleff V, Anagnostou V, Lupo B, Sausen M, Phallen J, Hruban CA, Tokheim C, et al. The genomic landscape of response to EGFR blockade in colorectal cancer. Nature. 2015; 526:263–267. [PubMed: 26416732]

Bibikova M, Fan JB. GoldenGate assay for DNA methylation profiling. Methods Mol Biol. 2009; 507:149–163. [PubMed: 18987813]

Boch C, Kollmeier J, Roth A, Stephan-Falkenau S, Misch D, Gruning W, Bauer TT, Mairinger T. 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. 2013; 3

Camilo R, Capelozzi VL, Siqueira SA, Del Carlo Bernardi F. Expression of p63, keratin 5/6, keratin 7, and surfactant-A in non-small cell lung carcinomas. Human pathology. 2006; 37:542–546. [PubMed: 16647951]

Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012; 489:519–525. [PubMed: 22960745]

Cancer Genome Atlas Research, N. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014; 511:543–550. [PubMed: 25079552]

Clements EG, Mohammad HP, Leadem BR, Easwaran H, Cai Y, Van Neste L, Baylin SB. DNMT1 modulates gene expression without its catalytic activity partially through its interactions with histone-modifying enzymes. Nucleic acids research. 2012; 40:4334–4346. [PubMed: 22278882]

Cowling VH, D'Cruz CM, Chodosh LA, Cole MD. c-Myc transforms human mammary epithelial cells through repression of the Wnt inhibitors DKK1 and SFRP1. Molecular and cellular biology. 2007; 27:5135–5146. [PubMed: 17485441]

Dai Z, Lakshmanan RR, Zhu WG, Smiraglia DJ, Rush LJ, Fruhwald MC, Brena RM, Li B, Wright FA, Ross P, et al. Global methylation profiling of lung cancer identifies novel methylated genes. Neoplasia. 2001; 3:314–323. [PubMed: 11571631]

Damiani LA, Yingling CM, Leng S, Romo PE, Nakamura J, Belinsky SA. Carcinogen-induced gene promoter hypermethylation is mediated by DNMT1 and causal for transformation of immortalized bronchial epithelial cells. Cancer research. 2008; 68:9005–9014. [PubMed: 18974146]

Delgado O, Kaisani AA, Spinola M, Xie XJ, Batten KG, Minna JD, Wright WE, Shay JW. Multipotent capacity of immortalized human bronchial epithelial cells. PloS one. 2011; 6:e22023. [PubMed: 21760947]

Dimauro T, David G. Ras-induced senescence and its physiological relevance in cancer. Current cancer drug targets. 2010; 10:869–876. [PubMed: 20718709]

Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature protocols. 2009; 4:1184–1191. [PubMed: 19617889]

Easwaran H, Johnstone SE, Van Neste L, Ohm J, Mosbruger T, Wang Q, Aryee MJ, Joyce P, Ahuja N, Weisenberger D, et al. A DNA hypermethylation module for the stem/progenitor cell signature of cancer. Genome research. 2012; 22:837–849. [PubMed: 22391556]

Easwaran H, Tsai HC, Baylin SB. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. Molecular cell. 2014; 54:716–727. [PubMed: 24905005]

Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nature genetics. 2000; 24:88–91. [PubMed: 10615135]

Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012; 150:1121–1134. [PubMed: 22980976]

Ha K, Lee GE, Palii SS, Brown KD, Takeda Y, Liu K, Bhalla KN, Robertson KD. 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. Human molecular genetics. 2011; 20:126–140. [PubMed: 20940144]

Helman E, Naxerova K, Kohane IS. DNA hypermethylation in lung cancer is targeted at differentiation-associated genes. Oncogene. 2012; 31:1181–1188. [PubMed: 21804601]

Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research. 2009a; 37:1–13. [PubMed: 19033363]

Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009b; 4:44–57. [PubMed: 19131956]

Issa JP. Aging and epigenetic drift: a vicious cycle. The Journal of clinical investigation. 2014; 124:24–29. [PubMed: 24382386]

Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007; 128:683–692. [PubMed: 17320506]

Kaufman CK, Mosimann C, Fan ZP, Yang S, Thomas AJ, Ablain J, Tan JL, Fogley RD, van Rooijen E, Hagedorn EJ, et al. A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. Science. 2016; 351:aad2197. [PubMed: 26823433]

Kraunz KS, Nelson HH, Liu M, Wiencke JK, Kelsey KT. Interaction between the bone morphogenetic proteins and Ras/MAP-kinase signalling pathways in lung cancer. British journal of cancer. 2005; 93:949–952. [PubMed: 16175182]

Leary RJ, Cummins J, Wang TL, Velculescu VE. Digital karyotyping. Nat Protoc. 2007; 2:1973–1986. [PubMed: 17703209]

Lee E, Yang J, Ku M, Kim NH, Park Y, Park CB, Suh JS, Park ES, Yook JI, Mills GB, et al. Metabolic stress induces a Wnt-dependent cancer stem cell-like state transition. Cell death & disease. 2015; 6:e1805. [PubMed: 26136078]

Lemjabbar-Alaoui H, Dasari V, Sidhu SS, Mengistab A, Finkbeiner W, Gallup M, Basbaum C. Wnt and Hedgehog are critical mediators of cigarette smoke-induced lung cancer. PloS one. 2006; 1:e93. [PubMed: 17183725]

Liu F, Killian JK, Yang M, Walker RL, Hong JA, Zhang M, Davis S, Zhang Y, Hussain M, Xi S, et al. Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. Oncogene. 2010; 29:3650–3664. [PubMed: 20440268]

Macaluso M, Paggi MG, Giordano A. Genetic and epigenetic alterations as hallmarks of the intricate road to cancer. Oncogene. 2003; 22:6472–6478. [PubMed: 14528270]

Maegawa S, Gough SM, Watanabe-Okochi N, Lu Y, Zhang N, Castoro RJ, Estecio MR, Jelinek J, Liang S, Kitamura T, et al. Age-related epigenetic drift in the pathogenesis of MDS and AML. Genome research. 2014; 24:580–591. [PubMed: 24414704]

Mazieres J, He B, You L, Xu Z, Lee AY, Mikami I, Reguart N, Rosell R, McCormick F, Jablons DM. Wnt inhibitory factor-1 is silenced by promoter hypermethylation in human lung cancer. Cancer research. 2004; 64:4717–4720. [PubMed: 15256437]

Mohammad HP, Cai Y, McGarvey KM, Easwaran H, Van Neste L, Ohm JE, O'Hagan HM, Baylin SB. Polycomb CBX7 promotes initiation of heritable repression of genes frequently silenced with cancer-specific DNA hypermethylation. Cancer research. 2009; 69:6322–6330. [PubMed: 19602592]

Nguyen DX, Chiang AC, Zhang XH, Kim JY, Kris MG, Ladanyi M, Gerald WL, Massague J. WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. Cell. 2009; 138:51–62. [PubMed: 19576624]

Niwa T, Ushijima T. Induction of epigenetic alterations by chronic inflammation and its significance on carcinogenesis. Advances in genetics. 2010; 71:41–56. [PubMed: 20933125]

Nojima M, Suzuki H, Toyota M, Watanabe Y, Maruyama R, Sasaki S, Sasaki Y, Mita H, Nishikawa N, Yamaguchi K, et al. Frequent epigenetic inactivation of SFRP genes and constitutive activation of Wnt signaling in gastric cancer. Oncogene. 2007; 26:4699–4713. [PubMed: 17297461]

O'Hagan HM, Mohammad HP, Baylin SB. Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. PLoS genetics. 2008; 4:e1000155. [PubMed: 18704159]

O'Hagan HM, Wang W, Sen S, Destefano Shields C, Lee SS, Zhang YW, Clements EG, Cai Y, Van Neste L, Easwaran H, et al. Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. Cancer cell. 2011; 20:606–619. [PubMed: 22094255]

Ohm JE, Baylin SB. Stem cell chromatin patterns: an instructive mechanism for DNA hypermethylation? Cell Cycle. 2007; 6:1040–1043. [PubMed: 17457052]

Pacheco-Pinedo EC, Durham AC, Stewart KM, Goss AM, Lu MM, Demayo FJ, Morrisey EE. Wnt/beta-catenin signaling accelerates mouse lung tumorigenesis by imposing an embryonic distal progenitor phenotype on lung epithelium. The Journal of clinical investigation. 2011; 121:1935–1945. [PubMed: 21490395]

Pacheco-Pinedo EC, Morrisey EE. Wnt and Kras signaling-dark siblings in lung cancer. Oncotarget. 2011; 2:569–574. [PubMed: 21753228]

Ramirez RD, Sheridan S, Girard L, Sato M, Kim Y, Pollack J, Peyton M, Zou Y, Kurie JM, Dimaio JM, et al. Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. Cancer research. 2004; 64:9027–9034. [PubMed: 15604268]

Revet I, Huizenga G, Chan A, Koster J, Volckmann R, van Sluis P, Ora I, Versteeg R, Geerts D. The MSX1 homeobox transcription factor is a downstream target of PHOX2B and activates the Delta-Notch pathway in neuroblastoma. Experimental cell research. 2008; 314:707–719. [PubMed: 18201699]

Revet I, Huizenga G, Koster J, Volckmann R, van Sluis P, Versteeg R, Geerts D. 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 letters. 2010; 289:195–207. [PubMed: 19815336]

Rountree MR, Bachman KE, Baylin SB. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nature genetics. 2000; 25:269–277. [PubMed: 10888872]

Sato M, Larsen JE, Lee W, Sun H, Shames DS, Dalvi MP, Ramirez RD, Tang H, DiMaio JM, Gao B, et al. Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations. Molecular cancer research : MCR. 2013; 11:638–650. [PubMed: 23449933]

Sato M, Vaughan MB, Girard L, Peyton M, Lee W, Shames DS, Ramirez RD, Sunaga N, Gazdar AF, Shay JW, Minna JD. 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 research. 2006; 66:2116–2128. [PubMed: 16489012]

Sausen M, Leary RJ, Jones S, Wu J, Reynolds CP, Liu X, Blackford A, Parmigiani G, Diaz LA Jr, Papadopoulos N, et al. Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. Nature genetics. 2013; 45:12–17. [PubMed: 23202128]

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nature genetics. 2007; 39:232–236. [PubMed: 17200670]

Shen H, Laird PW. Interplay between the cancer genome and epigenome. Cell. 2013; 153:38–55. [PubMed: 23540689]

Singh A, Sweeney MF, Yu M, Burger A, Greninger P, Benes C, Haber DA, Settleman J. TAK1 inhibition promotes apoptosis in KRAS-dependent colon cancers. Cell. 2012; 148:639–650. [PubMed: 22341439]

Stange DE, Koo BK, Huch M, Sibbel G, Basak O, Lyubimova A, Kujala P, Bartfeld S, Koster J, Geahlen JH, et al. Differentiated Troy+ chief cells act as reserve stem cells to generate all lineages of the stomach epithelium. Cell. 2013; 155:357–368. [PubMed: 24120136]

Stewart DJ. Wnt signaling pathway in non-small cell lung cancer. Journal of the National Cancer Institute. 2014; 106:djt356. [PubMed: 24309006]

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–15550. [PubMed: 16199517]

Suzuki H, Watkins DN, Jair KW, Schuebel KE, Markowitz SD, Chen WD, Pretlow TP, Yang B, Akiyama Y, Van Engeland M, et al. Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. Nature genetics. 2004; 36:417–422. [PubMed: 15034581]

Tata PR, Mou H, Pardo-Saganta A, Zhao R, Prabhu M, Law BM, Vinarsky V, Cho JL, Breton S, Sahay A, et al. Dedifferentiation of committed epithelial cells into stem cells in vivo. Nature. 2013; 503:218–223. [PubMed: 24196716]

Tellez CS, Juri DE, Do K, Bernauer AM, Thomas CL, Damiani LA, Tessema M, Leng S, Belinsky SA. 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 research. 2011; 71:3087–3097. [PubMed: 21363915]

Tessema M, Yingling CM, Liu Y, Tellez CS, Van Neste L, Baylin SS, Belinsky SA. Genome-wide unmasking of epigenetically silenced genes in lung adenocarcinoma from smokers and never smokers. Carcinogenesis. 2014; 35:1248–1257. [PubMed: 24398667]

Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA: a cancer journal for clinicians. 2015; 65:87–108. [PubMed: 25651787]

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, Laird PW. Epigenetic stem cell signature in cancer. Nature genetics. 2007; 39:157–158. [PubMed: 17200673]

You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin? Cancer cell. 2012; 22:9–20. [PubMed: 22789535]

SIGNIFICANCE

Cancer development is a complex process involving both genetic and epigenetic abnormalities. However, the evolution of these changes have 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.

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.

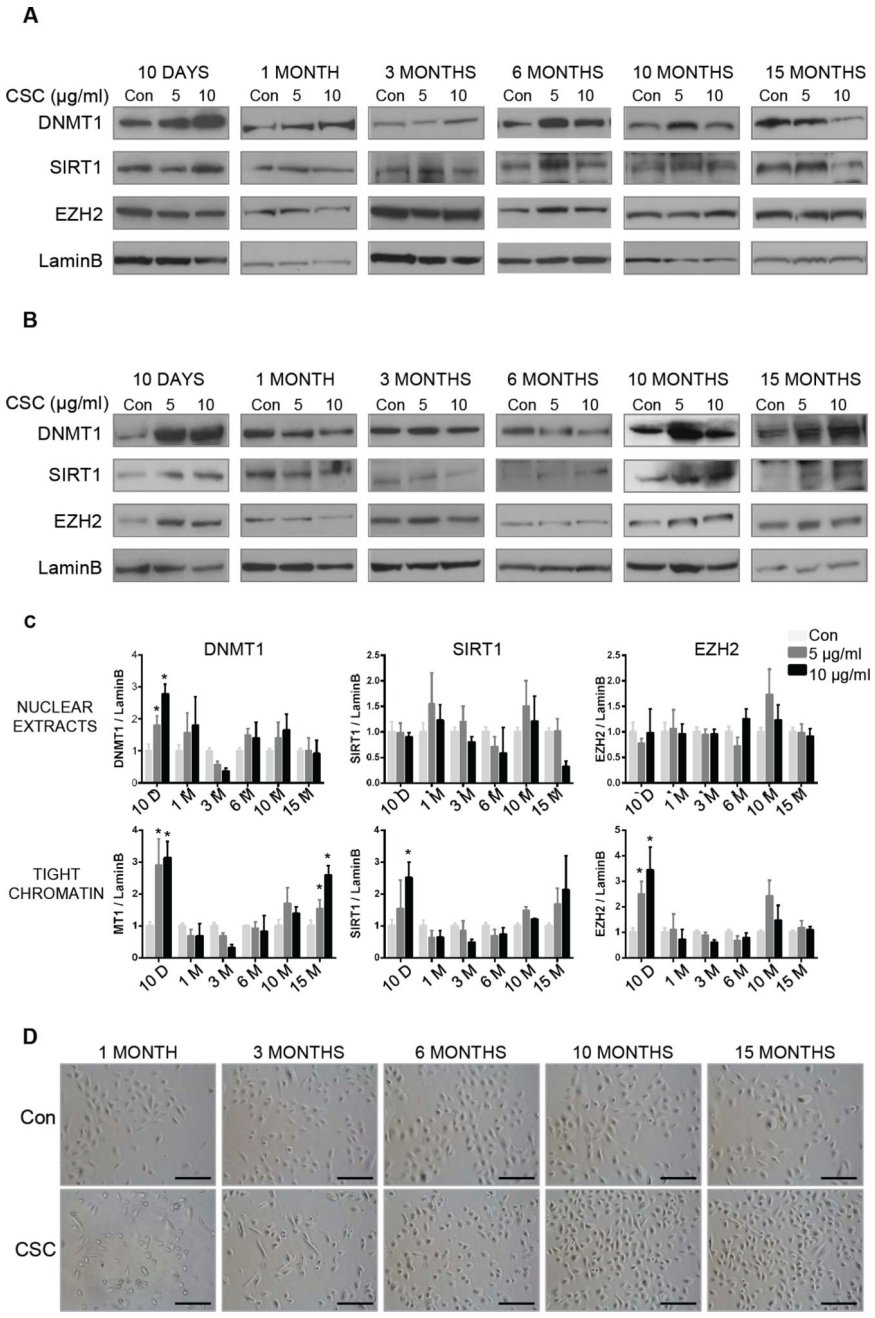


Figure 1. Expression of DNMT1, EZH2 and SIRT1 following chronic exposure to CSC

(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 (Con), 5 μg/ml and 10 μg/ml CSC over a period of 15 months collected at the indicated time points. (C) Quantitation of protein bands using ImageJ. The data presented is the mean of three independent experiments ± SD. * $p < 0.05$ by two tailed t-test. D) Phase contrast photomicrographs showing morphological changes observed in Control and 10 μg/ml CSC treated HBECs over time. Scale bar = 100 μm. See also Figure S1.

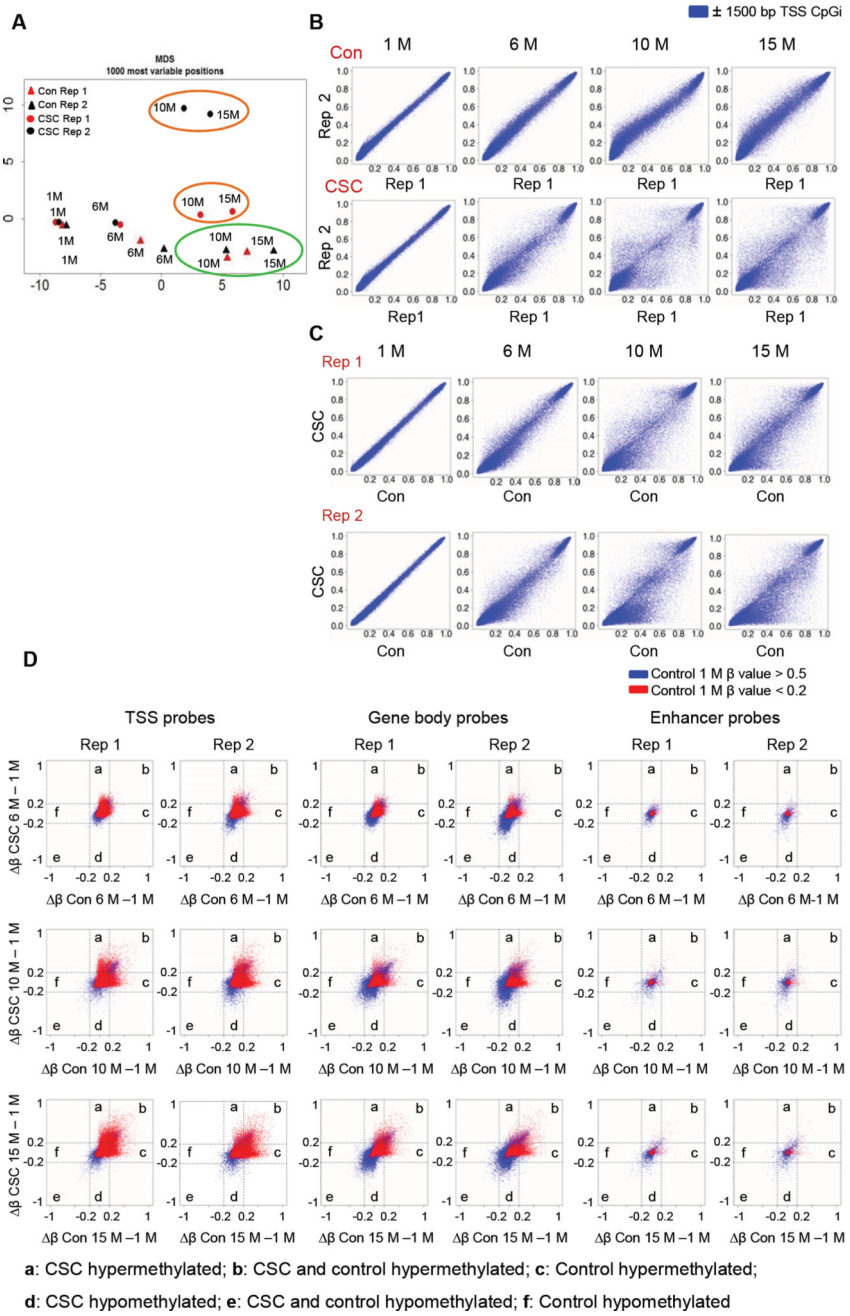


Figure 2. Progressive changes in methylation in control and CSC treated cells

(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) circles indicate groupings. (B) Comparison of genome-wide DNA methylation changes, 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 CpG islands in proximal promoter regions in control (x axis) versus CSC (y axis) treated 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 end point (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 β value > 0.5 in control cells at 1 month. Rep 1: replicate 1; Rep2: replicate 2. See also Figure S2.

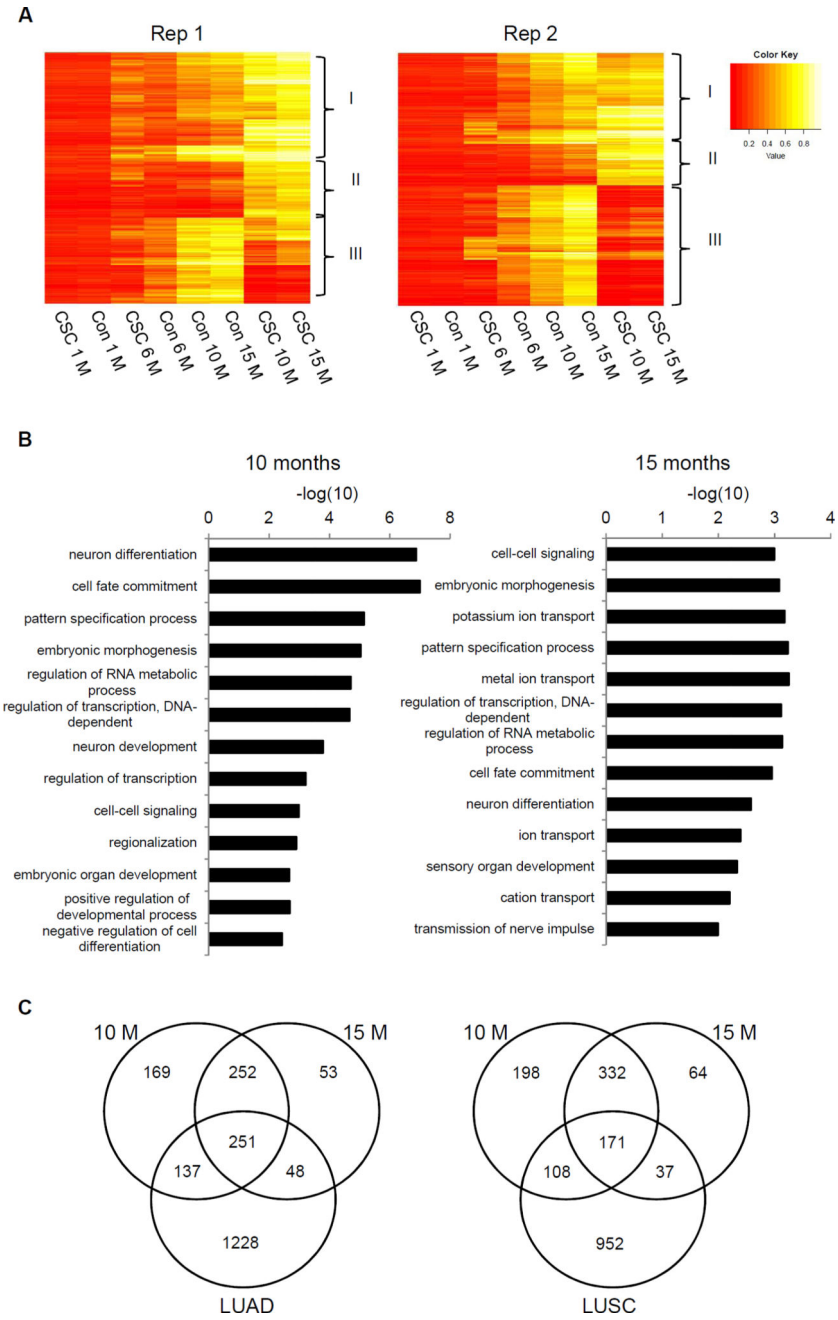


Figure 3. Identification and characterization of genes methylated upon CSC exposure

(A) Heat maps of β values for infinium probes in CpG islands within ± 1500 bp from TSS in control (Con) and 10 μg/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 10 μg/ml CSC treated cells 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. See also Figure S3, Tables S1 and S2. M: months.

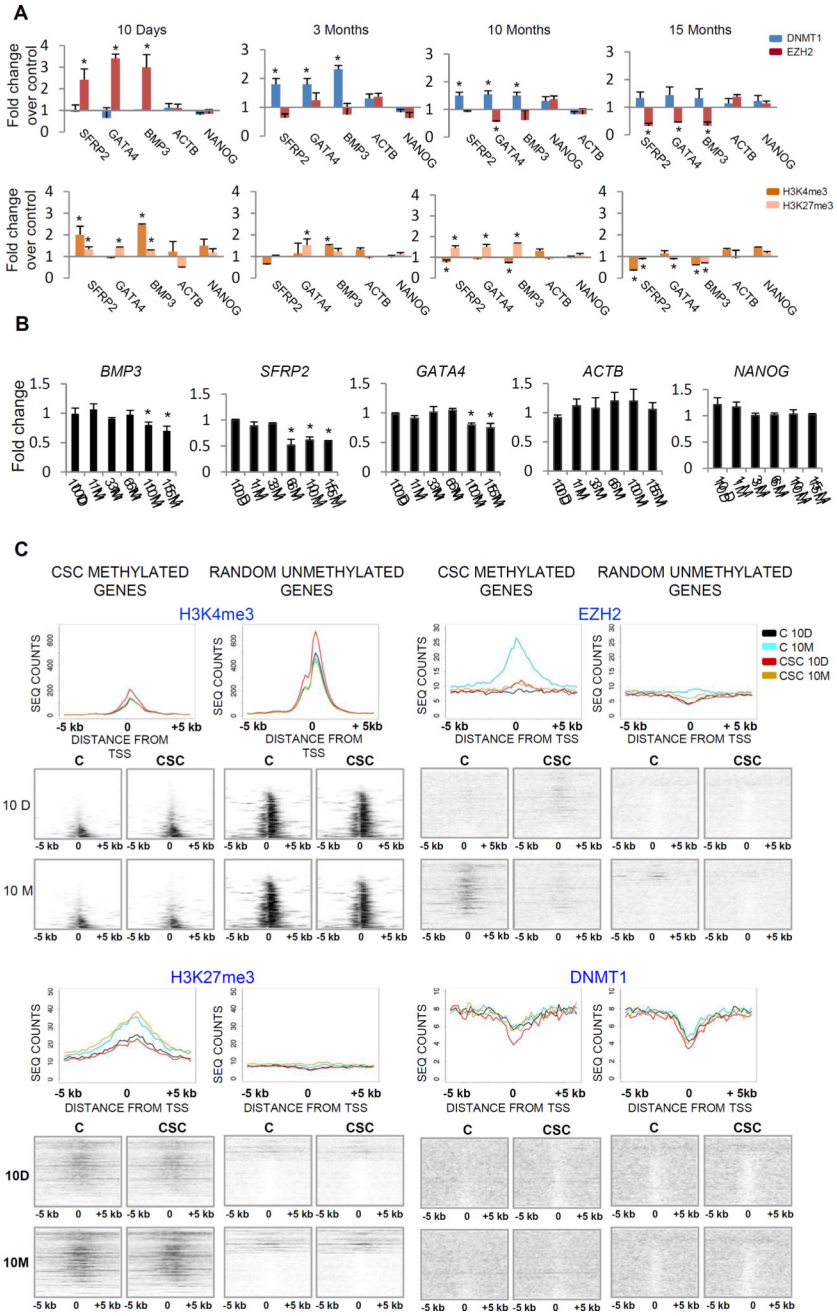


Figure 4. Progressive changes in chromatin marks and binding of chromatin-associated proteins in CSC exposed cells

(A) ChIP studies examining enrichment of H3K4me3, 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). X-axis; 0 = transcription start sites. For each mark analyzed, a graph showing quantitative data depicting the average read counts ± 5 kb from the transcription start site (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 C and CSC cells. See also Figure S4.

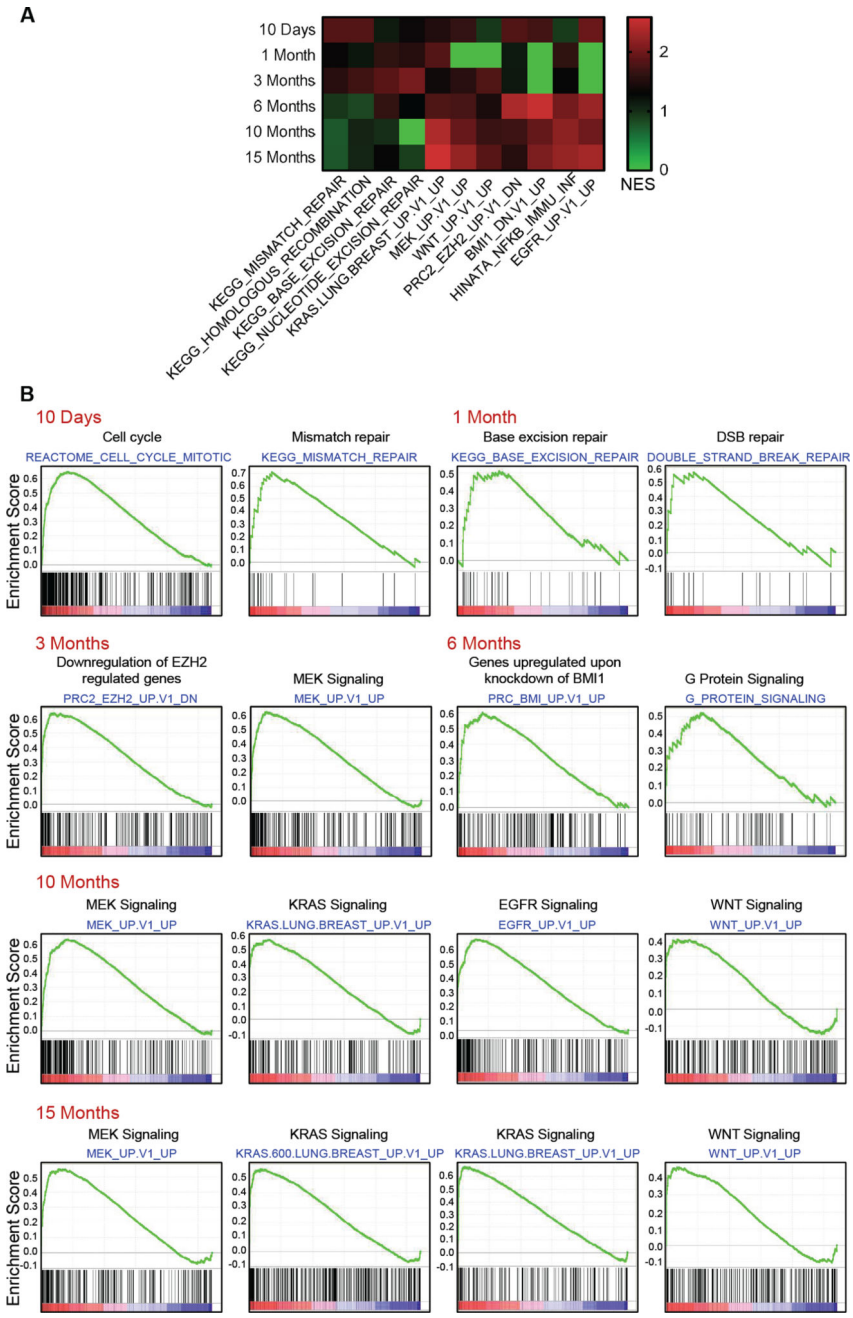
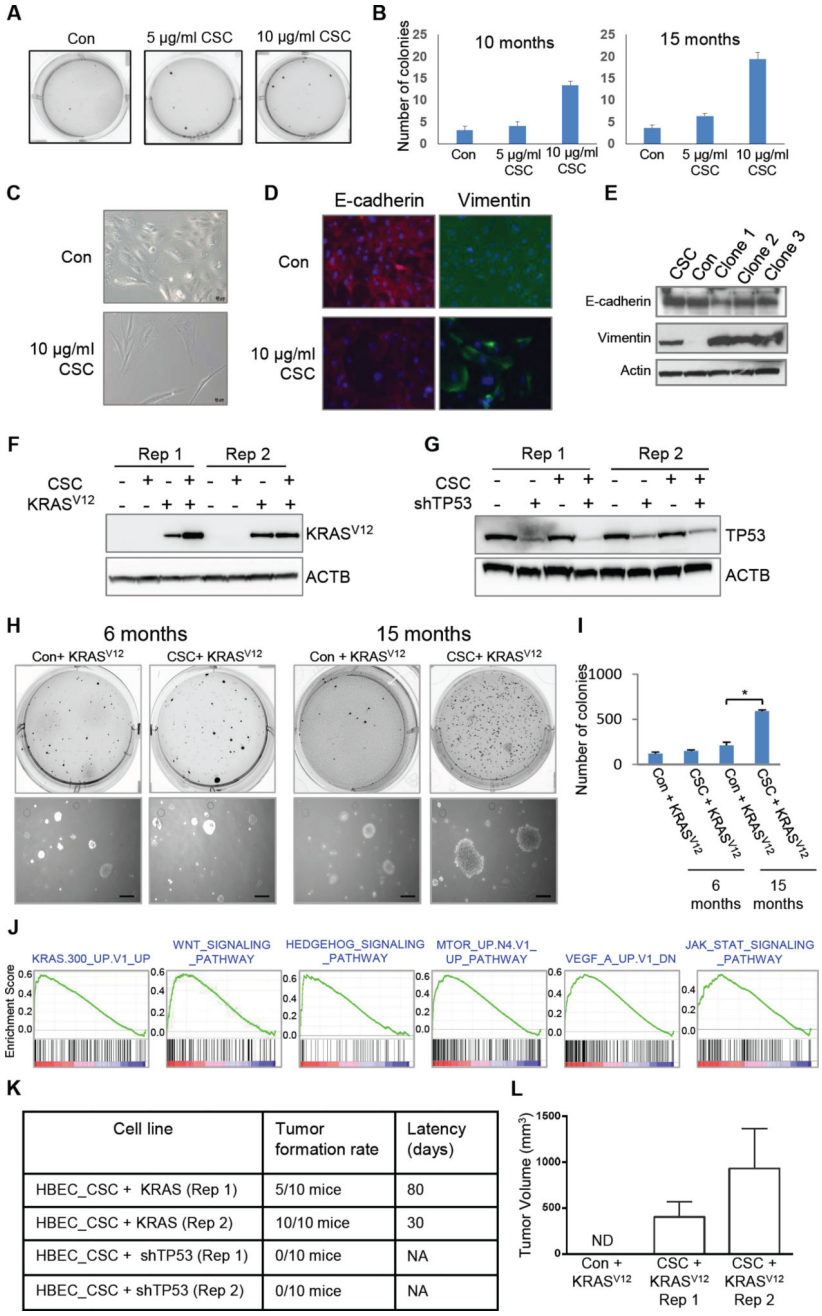


Figure 5. Gene set enrichment analysis (GSEA) of genome-wide gene expression following CSC exposure

(A) Genome wide gene expression analysis of control and 10 μg/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 to matched controls were calculated and analyzed using GSEA. The top 20 enriched Pathways were compared across the time points. Heatmap showing changes in enrichment (based on NES scores) 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 (FDR < 0.25, p value < 0.01). See also Figure S5.



Cell line	Tumor formation rate	Latency (days)
HBEC_CSC + KRAS (Rep 1)	5/10 mice	80
HBEC_CSC + KRAS (Rep 2)	10/10 mice	30
HBEC_CSC + shTP53 (Rep 1)	0/10 mice	NA
HBEC_CSC + shTP53 (Rep 2)	0/10 mice	NA

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 in panel A. Values indicate average number of colonies ± SD from replicate 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 bar = 50 μm. (D) Immunofluorescent staining of E-cadherin and Vimentin in cells grown in serum-containing medium for 96 hr. Scale bar = 100 μm. (E) Immunoblot for EMT markers in cells (CSC and 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 two 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 bar = 100 μm (I) Quantitation of colonies obtained in H. Data are presented as average number of colonies ± 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 to 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 ± SD (n = 5 for Rep1 and n = 10 for Rep 2).

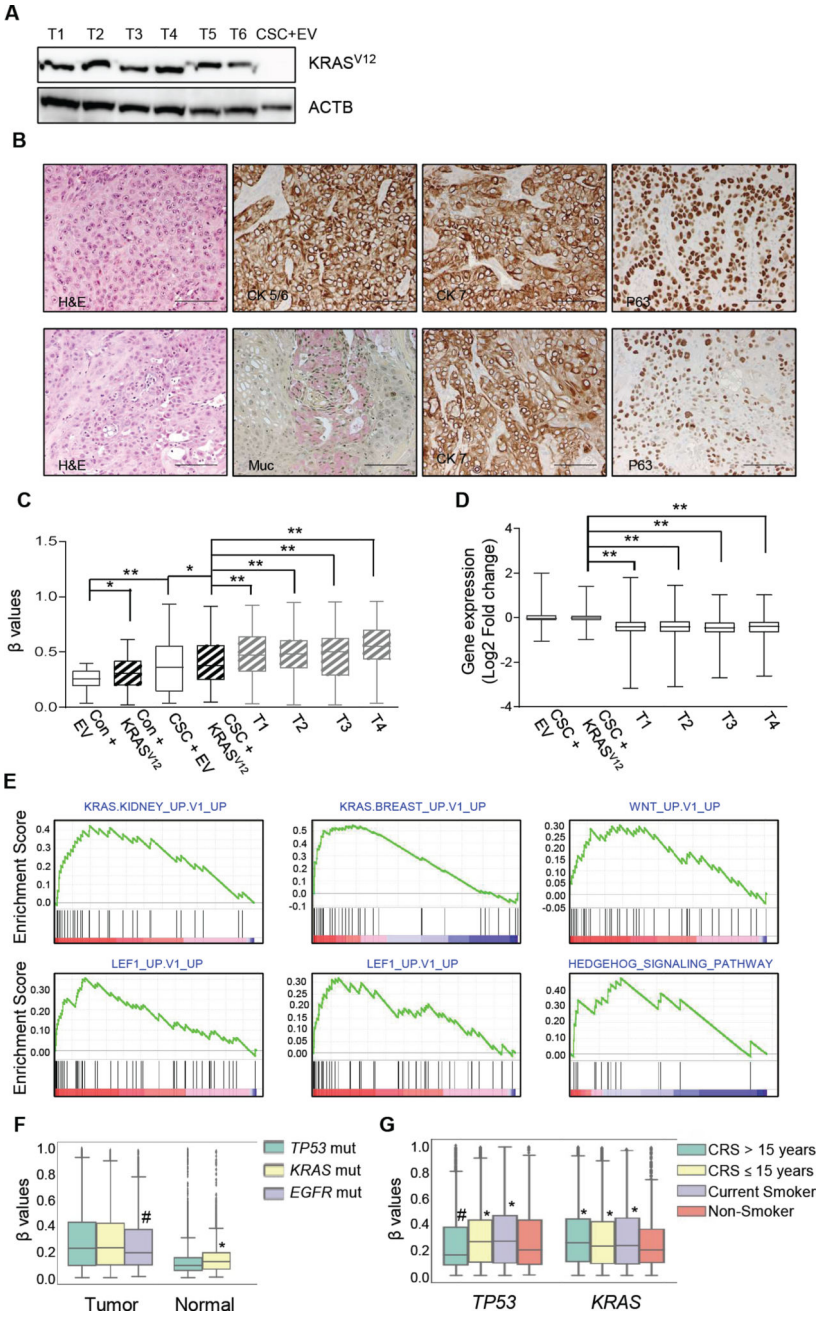


Figure 7. Transformation of CSC exposed cells following expression of mutant KRAS^{V12}

(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). 15 month CSC exposed cells 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} cell four weeks after appearance of the tumor. 5 sections were stained from each Replicate. Representative sections stained with H&E, mucicarmine, CK5/6, CK7, and p63 are shown. Scale bar = 100 μm. (C) Comparison of 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-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 rank test (D) Box plots 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 to 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 rank test (E) Signaling pathways enriched for in xenografts compared to control cells expressing $KRAS^{V12}$ analyzed with GSEA. T: xenografts. (F) Box plots of β values of promoter CpG island probes of CSC methylated genes in tumors 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 to 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) Box plots of β values of promoter CpG island probes of CSC methylated genes in tumors with *KRAS* and *TP53* mutations categorized by smoking status. * $p < 0.05$ by Welch's t-test indicates significant increase compared to non-smoker tumor within mutation group. # $p < 0.05$ by Welch's t-test indicates significant decrease compared to 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 and n = 14 current smokers, 20 CRS < 15 years, 12 CRS > 15 years, 3 non smokers for *TP53* tumors). Top and bottom of Box plots 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's of the lower and upper quartile. Observations that fall outside this range are displayed independently. See also Figure S6.

Table 1

Candidate genes with a driver function

Gene Name	Function
BMP3	Member of the transforming growth factor-beta (TGFβ) 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 regulate genes involved in embryogenesis. Involved in GPCR signaling pathway and is regulated by MAPK/ERK signaling.
SLIT2	Attenuated during lung cancer progression. Deregulates beta-catenin and E-cadherin and is associated with poor prognosis.
GALR1	Methylated in lung cancer. Identified as a tumor suppressor.

Table 1, See also Table S2: Candidate CSC-methylated genes with a driver function for the signaling events modulated by CSC treatment and which are frequently methylated in lung cancer.