The Sulfotransferase SULT1C2 is Epigenetically Activated and Transcriptionally Induced by Tobacco Exposure and is Associated with Patient Outcome in Lung Adenocarcinoma

Candace Johnson

University of Southern California

Daniel J. Mullen

University of Southern California

Suhaida A. Selamat

University of Southern California

Mihaela Campan

University of Southern California

Ite A. Offringa

University of Southern California

Crystal N. Marconett (✉ Crystal.Marconett@med.usc.edu)

University of Southern California

Research Article

Keywords: Lung cancer, lung adenocarcinoma (LUAD), cigarette smoke condensate (CSC), sulfotransferase SULT1C2, tobacco

Posted Date: August 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-795539/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at International Journal of Environmental Research and Public Health on December 31st, 2021. See the published version at <https://doi.org/10.3390/ijerph19010416>.

Abstract

Lung cancer is the leading cause of cancer-related death. Tobacco exposure is associated with 80–90% of lung cancer cases. The *SULT1C2* sulfotransferase modifies xenobiotic compounds to enhance secretion but can also render these compounds carcinogenic. To determine if *SULT1C2* contributes to tobacco-related carcinogenesis in the lung, we analyzed the expression and epigenetic state of *SULT1C2* in human patients in relation to smoking history as well as lung adenocarcinoma (LUAD) cell lines exposed to cigarette smoke condensate (CSC). *SULT1C2* expression was significantly positively correlated to overall LUAD patient survival in smokers, was elevated in LUAD tumors, and was significantly correlated with levels of patient exposure to tobacco smoke. *SULT1C2* promoter DNA methylation was inversely correlated with expression in LUAD and hypomethylation of the *SULT1C2* promoter was observed in Asian patients, as compared to Caucasians. *In vitro* analysis of LUAD cell lines indicates that CSC stimulates expression of *SULT1C2* in a dose-dependent and cell-line specific manner. *In vitro* methylation of the *SULT1C2* promoter significantly decreased transcriptional activity of a reporter plasmid and *SULT1C2* expression was activated by the DNA demethylating agent 5-Aza-2'-deoxycytidine in a cell line in which the *SULT1C2* promoter was hypermethylated. An aryl hydrocarbon receptor (AHR) binding site was detected spanning critical methylation sites upstream of *SULT1C2*. CSC exposure significantly increased AHR binding to this predicted binding site in the *SULT1C2* promoter in multiple lung cell lines. Our data suggests that CSC exposure leads to activation of the AHR transcription factor, increased binding to the *SULT1C2* promoter, and upregulation of *SULT1C2* expression, and that this process is inhibited by DNA methylation at the *SULT1C2* locus. Additionally, our results suggest that the level of *SULT1C2* promoter methylation and gene expression in normal lung varies depending on the race of the patient, which could in part reflect the molecular mechanisms of racial disparities seen in lung cellular responses to cigarette smoke exposure.

Background

Lung cancer is the most frequent cause of cancer-related death worldwide(1), and is the leading cause of cancer deaths of both men and women in the United States(2). Lung adenocarcinoma (LUAD) is the largest histological subtype of lung cancer, and the predominant form found among Asian and never-smoker patients(3, 4). In North America, 90% of the men and 75% of the women with lung cancer are current or former smokers. However, in Taiwan, only 7% of the female patients with lung adenocarcinoma are smokers(5). Several studies have carefully investigated possible mechanisms that account for differences in LUAD occurrence between Asian and Caucasian patients(6, 7).

Depending on social, cultural, and genetic differences, exposure to environmental toxins, including environmental tobacco smoke (ETS), varies among different racial groups(8, 9). Complex enzymatic systems have evolved to solubilize and secrete harmful toxins such as those found in cigarette smoke, however it is largely unknown how these vary across racial groups. *SULT1C2* is a critical component of the environmental detoxification pathway; a sulfotransferase which transfers a sulfur group to various substrates such as xenobiotics like drugs and chemical carcinogens(10). Of the eleven *SULT* family

members, *SULT1C2* showed the strongest enzymatic activity toward cigarette smoke and was the only *SULT* family member that did not affect endogenous chemicals, such as estrogen(11). In most cases, sulfonation of xenobiotics and small endogenous substrates detoxifies the body by increasing water solubility so the compound is cleared via urine or bile. However, in the case of certain environmental toxins like chemical present in cigarette smoke, *SULTs* can metabolically activate substrates into electrophiles that can be both carcinogenic and mutagenic(12–14). It has been established that *SULT1C2* is expressed in human stomach, kidney, fetal liver(15), and recently expression was also observed in the colorectal adenocarcinoma cell line LS180(16). Smoking-associated cancers occur in all of these tissues(17–19). However, the effect of cigarette smoke exposure on *SULT1C2* in lung is unknown(14).

Xenobiotic exposure can induce expression of detoxifying enzymes through activation of the aryl hydrocarbon receptor (AHR). Chemicals found in cigarette smoke are known ligands of AHR, which when liganded translocates to the nucleus where it partners with aryl hydrocarbon receptor nuclear translocator (ARNT) and acts as a transcription factor to upregulate expression of phase I detoxifying enzymes such as CYP1B1(20). The AHR recognition sequence, also known as the xenobiotic response element (XRE) is 5'-G/T.N.G.C.G.T.G.A/C.G/C.A-3', which contains a CpG dinucleotide demonstrated to disrupt AHR binding in certain biological contexts(21, 22). Transcriptional silencing by methylation of DNA at CpG dinucleotides is found at or near promoters or enhancers clustered in regions dubbed “CpG islands”(23). However, 45% of tissue-specific promoters are nearly devoid of CpGs(24, 25). Controversy exists about the role non-CpG island DNA methylation events play in gene regulation. Specific genes with CpG-poor promoters are expressed when methylated(26, 27), while others show an inverse correlation between DNA methylation and gene expression similar to that observed with CpG island promoters(28, 29). *SULT1C2* has a non-CpG island promoter that is normally methylated and silenced in adult lung tissue, however the role of DNA methylation in regulation of the CpG-island poor promoter of *SULT1C2* when exposed to environmental chemicals such as cigarette smoke has not been previously characterized.

In this article, we analyzed *SULT1C2* expression in LUAD and lung cell lines. Bioinformatic analysis showed that *SULT1C2* expression is correlated to LUAD patient survival, but only in patients with cigarette smoke exposure, while methylation of the *SULT1C2* promoter was inversely correlated with overall patient survival. We investigated the effects of cigarette smoke condensate (CSC) exposure on the lung using the immortalized BEAS-2B lung squamous cell epithelial cell line, and two lung adenocarcinoma cell lines, H2347, derived from a Caucasian 54-year old female patient, and PC3 (herein called PC3_LUAD, not to be confused with the PC3 prostate cancer cell line), derived from a non-smoking Asian patient. DNA methylation was able to inactivate downstream *SULT1C2* transcription. In addition, we found that the normal lungs of Asian patients have elevated levels of *SULT1C2* and lower methylation of the *SULT1C2* promoter. CSC-activated AHR was bound to the *SULT1C2* promoter specifically in the unmethylated state, which in turn activated downstream transcription of *SULT1C2*. Taken together, our results suggest a mechanism by which methylation of the *SULT1C2* promoter and subsequent disruption of binding of the aryl hydrocarbon receptor disrupts transcriptional activation of the *SULT1C2* detoxification enzyme and thereby the cell's ability to respond to cigarette smoke exposure. This may be the basis for the molecular

differences that contribute to the racial disparities observed in the etiology of LUAD occurrence between Asian and Caucasian patients.

Methods

Reagents and antibodies

Culture grade DMSO and rabbit polyclonal antibody for SULT1C2 (SC-130274) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). The actin antibody (AAN01) was purchased from Cytoskeleton Inc. (Denver, CO). Cigarette smoke condensate (CSC; #NC9028647) was purchased from Murty Pharmaceuticals Inc. (Lexington, KY). Trypsin-EDTA was obtained from USC Cell Culture Core Facility (Los Angeles, CA).

Cell Culture

Bronchial epithelial cell line BEAS-2B (ATCC#CRL9609) was obtained from American Type Culture Collection (Manassas, VA). PC3 (herein called PC3_LUAD) lung adenocarcinoma cell line (JCRB0077) was obtained from the Japanese Cancer Research Resources Bank (Osaka, Japan). H2347 lung adenocarcinoma cell line was a kind gift from Dr. Eric Haura. Cancer cell lines were maintained in RPMI-1640 from Mediatech (10-040-CV, Manassas, VA) and the BEAS-2B cell line was maintained in modified Eagle's medium (USC Cell Culture Core Facility). All media was supplemented with 10% fetal bovine serum, Genesee Scientific (25-514, El Cajon, CA) and 100U penicillin/streptomycin and grown in a humidified chamber with 5% CO₂ at 37° C. CSC (40mg/mL) was diluted in DMSO from Corning (Manassas, VA) and diluted 1:1000 in media prior to cell exposure. DMSO was used as the vehicle control for all experiments.

RNA Isolation and Quantitative Real Time-PCR (qRT-PCR)

Total RNA from BEAS-2B, PC3_LUAD, and H2347 cells treated with CSC or 5-Aza-CdR was isolated with Qiagen AllPrep DNA/RNA/Protein Kit (Valencia, CA) according to manufacturer's protocol. RNA was quantified using the Implen Nanophotometer Pearl (Westlake Village, CA). Total RNA (500ng) was converted to cDNA using iScript cDNA Synthesis Kit (Hercules, CA). The cDNA reaction product was amplified with primers listed in Table 1. PCR products were analyzed using the BioRad Real-Time System (Hercules, CA) measuring Sybr Green (BioRad).

Table 1

Primers

<i>ACTB</i> Forward	5'-GTTGAGAACCGTGTACCATGT-3'
<i>ACTB</i> Reverse	5'-TTCCCACAATTTGGCAAGAGC-3'
<i>SULT1C2</i> Forward	5'-CAGCCTGCAACTGTGGACAA-3'
<i>SULT1C2</i> Reverse	5'-GATGGCGGTGTTGGATGATG-3'
<i>CYP1B1</i> Forward	5'-CTGCACTCGAGTCTGCACAT-3'
<i>CYP1B1</i> Reverse	5'-TATCACTGACATCTTCGGCG-3'
<i>SULT1C2</i> promoter Forward	5'-aaaaaaactagtCATCCCAGTTCATCCTCCACAAA-3'
<i>SULT1C2</i> promoter Reverse	5'-aaaaaatcatgaTTTGAATAAATGCATCTGTAAAGCCA-3'
MethyLight <i>SULT1C2</i> Forward	5'-GGGTATGGTGGCGTACGTT-3'
MethyLight <i>SULT1C2</i> Reverse	5'-AATCTTAACCTCACTACAACCTCCG-3'
MethyLight <i>SULT1C2</i> Probe	5'-/6FAM-CTCCCGAATTCAAACGATTCTCCTATCTCA-BHQ-3/-3
MethyLight <i>ALU</i> Forward	5'-AGGTCGAGGTCGGCGG-3'
MethyLight <i>ALU</i> Reverse	5'-CCACGCCCGACTAATTTTATATCTT-3'
MethyLight <i>ALU</i> Probe	5'-/6FAM-CAAATAATCTCAAACCTCCGACCTCAAACGA-BHQ-1/-3'
ChIP <i>SULT1C2</i> Forward	5'-CCGTCTCTACTAAAAATACGAA-3'
ChIP <i>SULT1C2</i> Reverse	5'-AGCGATTCTCCTGTCTCAGCC-3'
ChIP <i>CYP1B1</i> Forward	5'-ATATGACTGGACCGACTTTCC-3'
ChIP <i>CYP1B1</i> Reverse	5'-GGCGAACTTTATCGGGTTGA-3'

Next Generation Sequencing Analysis

RNAseq from BEAS-2B (sh-Control rep1) was downloaded from GEO record GSE55215(30). RNAseq from PC3_LUAD and H2347 was downloaded from GEO record GSE110024(31). FASTQ files were cleaned using fastp(30) to retain only those reads for which > 90% of the reads have a quality score > 30. Reads were trimmed to remove degenerate sequence at the 5' and 3' ends, and the resulting cleaned reads underwent alignment to the hg19 genome and featureCount quantitation using RNA STAR(32). BAM alignments were visualized using the Integrated Genomics Viewer (IGVv2.8.2)(33). Expression correlations between *AHR* and either *SULT1C2* or *CYP1B1* were generated using TIMER2.0(34), which was sourced from expression data generated by the TCGA(35). For methylation analysis, Control_24hr_1, Control_24hr_2, and Control_24hr_3 files of BEAS-2B reduced representation bisulfite sequencing (RRBS) were downloaded from GEO record GSE155615(36). Control (untreated) sample files were merged and then aligned to the masked hg19 genome using bwameth(37). Shotgun whole genome bisulfite sequencing (sWGBS) was downloaded from DBTSS (<https://dbtss.hgc.jp/>)(38) and aligned to the

masked hg19 genome using BSMAP(39). BAM-level alignment files were visualized using IGV in bisulfite mode(33).

Survival Curves

Overall survival curves for expression of *SULT1C2* were generated using KMplot(40) on lung cancer, split by upper quartile expression and subsequently split by smoking status (ever-smokers n = 820; never-smokers n = 205). Overall survival curves for methylation at cg13968390 were generated using TCGA data obtained from TCGAbiolinks(41) on lung adenocarcinoma (LUAD) split by upper and lower quartile.

Microarray Analysis

Expression data generated by the Early Detection Research Network (EDRN) was originally run on the Illumina WG-6 microarray platform. Preprocessed and normalized data was downloaded from GEO (GSE32867) and used for all subsequent analyses. Methylation data generated on the Illumina Infinium 27K array by the EDRN and on previously published samples from the Ontario Tumor Bank (OTB) were downloaded from GEO (GSE32861 and GSE32866, respectively)(42). Beta values were calculated by dividing the methylation value at a given CpG by the sum of both methylated (M) and unmethylated signal (U) for that probe $[M/(U+M)]$ (43).

SULT1C2 Promoter Construction

The *SULT1C2* promoter from -1271 from the TSS through +535 in the 5'UTR was PCR amplified from genomic DNA from lung cancer cells using Phusion High Fidelity DNA Polymerase from New England Biolabs (Ipswich, MA) using primers listed in Table 1. The gel fragment was purified using Qiaquick Gel Extraction kit from Qiagen (Valencia, CA). The fragment was then inserted into the CpG-less vector using Instant Sticky-end Ligase Master Mix from New England Biolabs (NEB) (Ipswich, MA). The plasmid was transformed into Invitrogen *E. coli* PIR1 cells (Grand Island, NY). The promoter sequence was verified using Genewiz (La Jolla, CA).

In vitro methylation

pCpGL-*SULT1C2* vector was incubated with *SssI* (2.5 U/ug) in the presence of 160 uM S-Adenosylmethionine (SAM), both from New England Biolabs (Ipswich, MA) overnight. This was repeated for a second overnight treatment. Methylation of plasmids was confirmed by digestion of methylation-sensitive restriction enzymes *HpaII* and *HhaI*, and methylation insensitive *MspI* from New England Biolabs (Ipswich, MA).

Transfection and Luciferase Assay

PC3_LUAD cells were transfected with 800 ng of methylated and unmethylated pCpGL-*SULT1C2* luciferase reporter vector and 200ng of Renilla luciferase using Invitrogen Lipofectamine 2000 in low-serum Optimem media (Grand Island, NY). H2347 and BEAS-2B cells were transfected using FuGene HD transfection reagent from Promega (Madison, WI) using the same vector amounts. 24 hours post-transfection, medium was removed and replaced with medium containing 10% FBS, 20 ug/ml of CSC or DMSO in the absence of antibiotics for 24 hours. Cells were washed with PBS and lysed with Promega passive lysis buffer (Madison, WI). Cell lysates were freeze/thawed and assayed for firefly and renilla luciferase activity using Promega Dual-Luciferase Reporter Assay System on a Promega Glomax Luminometer. pCpGL and CMV vectors were used for negative and positive controls, respectively. Experiments were performed as technical triplicates in three independent experiments.

5-Aza-CdR treatment

BEAS-2B, PC3_LUAD, and H2347 cell lines were plated 24 hours prior to treatment. Cells were treated with the indicated concentrations of 5-Aza-CdR from Sigma Chemical Co. (St Louis, MO) for 24 hours. Cells were allowed to recover and replicate for 72 hours post drug removal. On the third day, cells were washed with cold PBS and harvested for RNA and DNA.

MethyLight Assay

1 ug of DNA was bisulfite treated using the EZ DNA Methylation kit from Zymo Research (D5002 ,Irvine, CA). Bisulfite-treated DNA was probed with MethyLight primers and probe as described in Table 1. Control and treated DNA was incubated with Taq Man enzyme from Applied BioSystems (Carlsbad, CA), primers, and probe in 30 ul reactions and analyzed with Alu repeats used to normalize input DNA as previously published(44).

Chromatin immunoprecipitation (ChIP) Assay

PC3_LUAD, H2347, and BEAS-2B cells were grown to 85% confluency in 150 mm plates and were treated with DMSO or 20 ug/ml CSC for the indicated times. Cells were crosslinked using 1% formaldehyde for 10 minutes followed by quenching with 125 mM Glycine (Sigma). Cells were lysed in 800 ul of cell lysis buffer with protease inhibitors from Sigma-Aldrich (#P2714, St. Louis, MO) and incubated on ice for 20 minutes, then DNA was sonicated to lengths between 200 bp-1000 bp. Input was 1% of the total lysate. Chip-grade AHR antibody (ab84833) from Abcam (Cambridge, MA) and control IgG (sc-66931) from Santa Cruz Biotechnology were used to precipitate DNA. Purified ChIP DNA was then purified by phenol-chloroform extraction and quantified using Implen Nanophotometer Pearl (Westlake Village, CA).

Results

Expression of *SULT1C2* is correlated with overall LUAD patient survival and cigarette smoke exposure, and cigarette smoke condensate can induce *SULT1C2* in lung cell lines

To determine what role *SULT1C2* may play in lung adenocarcinoma (LUAD) we first utilized large-scale publicly available datasets of gene expression levels in LUAD tumors to determine if *SULT1C2* expression was related overall patient survival (OS). Because of the known involvement of *SULT1C2* as a xenobiotic metabolism enzyme, we first split patients based on smoke-exposure status, then plotted survival as a function of *SULT1C2* expression using KMplot(45) (Fig. 1A). The effect of *SULT1C2* on LUAD patient OS appeared to be dependent on cigarette-smoke exposure; never-smoker patients showed no effect of *SULT1C2* on OS, whereas patients with smoke exposure showed significantly better survival with *SULT1C2* expression. To further characterize the relationship between *SULT1C2* and LUAD, we analyzed publicly available data from The Cancer Genome Atlas (TCGA) on LUAD expression in tumor and unmatched adjacent tumor normal (AdjNTL). This indicated that *SULT1C2* expression levels were significantly elevated in LUAD tumors relative to AdjNTL (Fig. 1B). To determine the effect of the subjects' cigarette smoke exposure on *SULT1C2* mRNA levels, LUAD tumor samples were split based on patients' smoke exposure status into 4 major TCGA-annotated categories: never smokers (category 1), former smokers who quit more than 15 years prior (category 2), former smokers who quit less than 15 years prior (category 3), and current smokers (category 4). Plotting expression of *SULT1C2* relative to smoking status indicated that *SULT1C2* expression levels were significantly inversely correlated to duration of smoke exposure ($p = 9.09 \times 10^{-6}$, Fig. 1B, Table 2). To mechanistically investigate the mechanism by which *SULT1C2* expression levels are affected by smoke exposure, we carried out *in vitro* experiments.

We selected three cell lines with varying levels of *SULT1C2* endogenous expression: immortalized non-cancerous lung epithelial cell line BEAS-2B and two lung adenocarcinoma cell lines, H2347 and PC3_LUAD. Publicly available RNAseq profiles of all three were downloaded from the Gene Expression Omnibus(31, 46) and aligned to the hg19 genome prior to determine reads-per-kilobase of gene per millions mapped (RPKM) levels. BEAS-2B did not express detectable levels of *SULT1C2* (Fig. 1C). H2347 robustly expressed *SULT1C2* while PC3_LUAD expressed barely detectable levels of *SULT1C2* levels as measured by RNAseq (Fig. 1C). To determine if cigarette smoke was able to affect transcriptional levels of *SULT1C2*, each of these cell lines was treated with cigarette smoke condensate (CSC). *SULT1C2* RNA levels were measured alongside the positive control gene for CSC exposure, *CYP1B1*. Untreated cells were maintained in a separate incubator so they would not be affected by secondary aerosolized CSC. All cell lines were treated with 10, 20, 40, and 80 ug/ml of CSC for 24 hours. We used 10 and 20 ug/ml to simulate second hand smoke exposure and 40 and 80 ug/ml to simulate a smoker and heavy smoker environment, respectively, and to account for potential substrate inhibition of the phase I and phase II xenobiotic metabolizing enzymes (47). We observed a dose-dependent effect on transcription levels of

SULT1C2 in BEAS-2B and H2347 (Fig. 1D). In contrast, PC3_LUAD cells showed no significant dose-dependent response. All lung cell lines tested showed a significant transcriptional response of *CYP1B1* at 24 hours (Figs. 1E). This raised the question of why we observed differential induction of *SULT1C2* in the tested LUAD cell lines and what could account for differential levels of *SULT1C2* expression prior to CSC exposure.

Methylation of *SULT1C2* promoter is altered in human lung

We hypothesized that the differential endogenous expression of *SULT1C2* and response to CSC exposure may be due to differing epigenetic states in the *SULT1C2* promoter-spanning regulatory region. *SULT1C2* is classified as having a CpG-poor promoter, with sparse CpG dinucleotide occurrence. However, multiple studies support a role of CpG-poor promoters in tissue-specific expression(48) (49). To determine the role CpG methylation in the *SULT1C2* promoter plays in LUAD, we first examined DNA methylation profiles of 390 LUAD patients and 26 AdjNLT controls generated by the TCGA on the Illumina Infinium HumanMethylation450 Beadchip(50). One probe on the array, cg13968390, was located within the *SULT1C2* promoter. Methylation at this probe location was used to split LUAD patients into methylation high and methylation low groups, and OS was compared between them. We observed that methylation levels at cg13968390 are significantly inversely correlated with patient overall survival (Fig. 2A). Next, we evaluated if cg13968390 methylation was altered in LUAD tumors compared to AdjNLT. To do so, we again utilized the TCGA LUAD dataset and observed that DNA methylation levels at cg13968390 were significantly lower in LUAD tumors as compared to AdjNLT (two-tailed unpaired T test; $p = 2.6 \times 10^{-6}$, Fig. 2B). To validate these findings, we used a secondary, independent dataset generated by the Early Detection Research Network (EDRN), that profiled 59 LUAD tumors alongside matched AdjNLT(42). We found that cg13968390 was also hypomethylated in LUAD tumors vs. AdjNLT in this dataset (two-tailed paired T test, $p = 5.2 \times 10^{-4}$, Fig. 2B). To further validate this finding, we utilized a third, independent study that profiled 27 LUAD tumors and AdjNLT derived from the Ontario Tumor Bank tissue repository. We performed a one-sided paired T-test and found that cg13968390 was significantly hypomethylated in LUAD tumor tissue from this source as well ($p = 1.9 \times 10^{-3}$, Fig. 2B). Thus, hypomethylation of cg13968390 in tumor compared to non-tumor lung appears to be a common feature of LUAD.

Next, we set out to determine if CpG methylation of the *SULT1C2* promoter was causally related to *SULT1C2* expression. We examined the correlation between expression and methylation of the *SULT1C2* promoter region in samples from the TCGA database for which matched RNA expression and DNA methylation data was available and found a significant inverse correlation between methylation of cg13968390 and *SULT1C2* expression (Fig. 2C). Linear regression was used to determine the significance of that association ($p = 1.53 \times 10^{-19}$). To confirm this observation, a secondary dataset derived from patients in the EDRN collection (51) consisting of 60 tumors and paired adjacent non-tumor lung (AdjNLT) was also examined for methylation state vs. matched *SULT1C2* expression. We again found a statistically significant inverse correlation ($\text{cor} = -0.346$) between cg13968390 DNA methylation and

expression of *SULT1C2* ($p = 8.37 \times 10^{-6}$, Fig. 2D). In sum, multiple LUAD patient cohorts exhibited hypomethylation of cg13968390 and a significant inverse relationship to *SULT1C2* expression.

DNA methylation represses transcription of *SULT1C2* promoter

Having observed a highly significant correlation between *SULT1C2* promoter hypomethylation in LUAD and *SULT1C2* expression levels, we sought to functionally test this relationship. To do so, we used publicly available RNAseq data for BEAS-2B, H2347 and PC3_LUAD cell lines (31, 46) as well as shotgun whole genome bisulfite sequencing (sWGBS) and reduced representation bisulfite sequencing (RRBS) data made available on the Database of Transcriptional Start Sites (DBTSS)(38) and the Gene Expression Omnibus (GEO)(36), respectively. Visualization of the *SULT1C2* promoter using bisulfite mode in IGV revealed that H2347 cells have unmethylated CpGs throughout the *SULT1C2* promoter and robust expression of *SULT1C2* in untreated cells (Fig. 3A), and that PC3_LUAD cells displayed high levels of promoter methylation and low levels of expression, consistent with promoter CpG methylation blocking transcription of the adjacent gene. This may also account for the previously observed inability of CSC to upregulate *SULT1C2* in PC3_LUAD cells (Fig. 1E). In contrast, RRBS data generated on BEAS-2B cells indicated that the CpGs present in the RRBS data were unmethylated, however BEAS-2B cells lack expression of *SULT1C2*. This could be due to a number of factors, such as the requirement for specific transcription factors not expressed in BEAS-2B under untreated conditions, or repression of enhancers whose association are required for basal activation of the *SULT1C2* promoter. In order to test whether *SULT1C2* promoter methylation functionally affects transcriptional activity of the adjacent *SULT1C2* gene, the promoter region from - 1.27 kb to + 535 surrounding the transcriptional start site (TSS) of *SULT1C2* was cloned into a CpG-less vector(52) containing the luciferase reporter gene. The CpG-less vector is devoid of CpG dinucleotides. Therefore, the only CpGs present in the construct are within the *SULT1C2* promoter. The *in vitro* *SssI*-methylated *SULT1C2* promoter plasmid was transfected into all three cell lines with subsequent CSC treatment to determine the effect of the methylation state of the *SULT1C2* promoter on downstream gene expression levels.

In all three lung cell lines tested, the *SULT1C2* promoter showed baseline activity in the unmethylated state, and this transcriptional activity was repressed by *in vitro* *SssI* methylation (Fig. 3B). Our results are therefore in agreement with Han *et al.*, showing that DNA methylation can directly silence CpG-poor promoters(53). We then tested whether the addition of CSC could activate the *SULT1C2* promoter. Consistent with the transcript level data in Fig. 1, the addition of CSC induced transcriptional activity of the unmethylated *SULT1C2* promoter in BEAS-2B cells. However, we did not observe significant induction by CSC the two LUAD cell lines, suggesting that the while cloned promoter can drive baseline expression, it may lack certain regulatory elements (such as enhancers) mediating CSC induction (Fig. 3B).

While the results of the *in vitro* *SssI* methylated *SULT1C2* promoter construct indicated that methylation can affect downstream transcriptional activity, this did not test whether alteration of DNA methylation

levels *in vivo* at the endogenous *SULT1C2* locus could alter transcriptional activity of *SULT1C2*. In order to directly test the effect of DNA methylation on the activity of the endogenous promoter, we used the DNA methylation inhibitor 5-Aza-deoxycytidine (5-aza-CdR) to block DNA methylation and subsequently measured *SULT1C2* transcription. PC3_LUAD cells were the only cell line showing endogenous *SULT1C2* promoter methylation (Fig. 3C) and were therefore the only line used for this experiment. PC3_LUAD cells were exposed to two doses of 5-aza-CdR, 0.15 μM (half the clinical dose) and 0.3 μM (clinical dose)(54), for 24 hours, after which the cells underwent recovery for three days to allow for DNA replication to incorporate the drug into the daughter cells and block DNA methylation (55). Post treatment with 5-aza-CdR, MethyLight(56) was used to determine methylation status of the *SULT1C2* promoter alongside qRT-PCR to evaluate *SULT1C2* expression levels. Treatment of PC3_LUAD cells with 5-aza-CdR resulted in a dose-independent decrease in methylation at the endogenous *SULT1C2* promoter (Fig. 3C). The percent methylated reference (PMR) decreased and *SULT1C2* gene expression increased in a dose-dependent manner and reached significance at the clinical dose of 0.3 μM . Taken together, these results suggest that methylation plays a significant role in the regulation of *SULT1C2* expression in lung cell lines, which agrees with our findings in human samples that showed an inverse correlation between methylation and RNA expression (Fig. 2C).

SULT1C2 expression is elevated in adjacent non-tumor lung of Asians relative to Caucasians and is significantly correlated to SULT1C2 promoter methylation levels.

Now that we had established a direct relationship between *SULT1C2* promoter methylation and *SULT1C2* expression, we wanted to further understand what underlying features of the cell models could contribute to the differential promoter methylation observed between PC3_LUAD and H2347. PC3_LUAD and H2347 cell lines were both derived from female patients, and no significant difference was observed in *SULT1C2* levels based on sex or age of the patient (Table 2).

Table 2

SULT1C2 expression in TCGA LUAD. Multiple linear regression was used to include all listed clinical features into one model. Bolded p values were considered significant.

<i>SULT1C2</i> <i>Expression</i>	Number of patients	Estimate	P-value
Sample type			
<i>Normal</i>	53	0.60538	0.0479
<i>Tumor</i>	429		
Gender			
<i>Male</i>	210	-0.13455	0.4926
<i>Female</i>	272		
Age		<i>By Year</i>	
	482	0.01207	0.3031
Smoking		<i>vs. Never Smoker</i>	
<i>Never Smoker</i>	71	–	–
<i>Former Smoker (> = 15 years)</i>	129	-0.05579	0.8595
<i>Former Smoker (< 15 years)</i>	169	-0.70430	0.0187
<i>Current Smoker</i>	113	-1.03470	0.0016
Race		<i>vs. Caucasian</i>	
<i>Caucasian</i>	422	–	–
<i>Black or African American</i>	52	0.39562	0.2068
<i>Asian</i>	7	1.63644	0.0414
<i>American Indian or Alaskan Native</i>	1	-0.12833	0.9512

However, expression of *SULT1C2* did vary significantly based on the race of the patient in the TCGA cohort. We then tested the cohort to determine if methylation levels in the *SULT1C2* promoter varied significantly based on race as well. We observed that methylation levels were significantly lower in Asian relative to Caucasian patients (Table 3).

Table 3

The effect of race on *SULT1C2* promoter methylation levels in TCGA LUAD dataset. Univariate linear regression was used on the indicated clinical feature. Bolded p value was considered significant.

<i>cg13968390 Methylation</i>			
Race		vs. <i>Caucasian</i>	
<i>Caucasian</i>	360	–	–
<i>Black or African American</i>	51	0.005096	0.8020
<i>Asian</i>	6	-0.130799	0.0197
<i>American Indian or Alaskan Native</i>	0	N/A	N/A

However, the TCGA cohort contained relatively few Asian patients, and the DNA methylation data was derived from LUAD which is subject to a host of molecular alterations during tumor formation. To further validate these findings in a secondary dataset with a larger number of Asian patients as well as data from non-tumor adjacent normal tissue, we used the AdjNLT subset of the EDRN cohort, which contained expression and methylation data from adjacent normal lung of 22 Asian and 37 Caucasian patients. We found that *SULT1C2* expression was significantly elevated in Asian relative to Caucasian patients (Fig. 4A), and that there was a concomitant lower level of DNA methylation of the *SULT1C2* promoter in Asian patients relative to Caucasians (Fig. 4B). Indeed, methylation of cg13968390 within the *SULT1C2* promoter was the most significantly differential methylation event genome-wide between Asians and Caucasians in the EDRN AdjNLT cohort (Fig. 4C). To determine if methylation and expression were inversely related according to the race of the patient, we plotted methylation vs. expression of the EDRN dataset, which resulted in two clusters with minimal overlap, one consisting mainly of Asian patients, the other consisting primarily of Caucasian patients (Fig. 4D).

CSC alters AHR occupancy at the *SULT1C2* promoter

Now that we had established a direct relationship between *SULT1C2* promoter methylation levels and expression as well as CSC-mediated transcriptional activation, we wanted to understand the mechanism by which CSC mediates activation of the *SULT1C2* promoter. To do so, we performed transcription factor binding site analysis on the area surrounding CpG sites within the *SULT1C2* promoter using Biobase(57) (Fig. 5A). We identified the aryl hydrocarbon receptor (AHR) as a likely binding candidate to a site carrying two CpGs within the *SULT1C2* promoter. AHR is a ligand-activated transcription factor(58) whose transcriptional activity is induced by xenobiotic chemicals, among which polycyclic aromatic hydrocarbons (PAH) such as benzo(a)pyrene found in cigarette smoke(59). It is well established that aryl hydrocarbon receptor acts as a transcriptional activator for phase I detoxifying enzymes such as *CYP1B1* when induced by exogenous ligands(60). However, there is very little evidence available implicating AHR-

mediated activation of phase II enzymes. We therefore hypothesized that AHR was the major transcription factor bridging CSC-induced transcriptional activity and *SULT1C2* upregulation in cells with unmethylated *SULT1C2* promoters.

If *SULT1C2* is a transcriptional target of AHR in LUAD, we would expect to see a positive correlation between *AHR* and *SULT1C2* expression in LUAD patient data cohorts. To test this, we utilized RNA expression levels of *AHR* and *SULT1C2* generated by TCGA PanCancer study through the TIMER2.0 portal(34). Indeed, *SULT1C2* expression levels were significantly positively correlated to *AHR* expression in LUAD (cor = 0.173, $p = 8.1 \times 10^{-5}$, Fig. 5B). This trend was also observed in the known AHR target gene *CYP1B1* (cor = 0.359, $p = 2.83 \times 10^{-17}$).

We then sought to test if AHR bound differentially to the *SULT1C2* promoter in the presence of cigarette smoke. In order to do so, we first analyzed RNAseq from BEAS-2B, H2347, and PC3_LUAD for levels of *AHR* expression. Indeed, all three cell lines expressed AHR and its dimerization partner, ARNT (Fig. 5C). We then performed chromatin immunoprecipitation of AHR from cell lines treated with vehicle or 20 ug/ml CSC. CSC treatment resulted in significant enrichment of AHR at the *SULT1C2* promoter in all three cell lines as compared to DMSO (Fig. 5D). In all cases, enrichment of AHR binding to the *SULT1C2* promoter was greater than the enrichment at a previously described AHR binding site near *CYP1B1*(60).

Discussion

SULT1C2 is a phase II detoxifying enzyme known for its ability to metabolize xenobiotics by adding a sulfonate group to the target, facilitating excretion via urine or bile. In this study, we make several observations. First, that CSC was able to induce *SULT1C2* expression in lung cells, and that this activation was cell-line dependent. Sakakibara *et al.* previously observed *SULT1C2* in fetal lung tissue(61), however there is little to no expression of *SULT1C2* in the human adult lung as reported by multiple consortia, including the Human Protein Atlas(62), GTEx(63), and FAMTOM5(64) projects. We detected expression of *SULT1C2* in the AdjNLT of Asian patients, and in one of the LUAD cell lines, H2347, without exposure to CSC. This may be due to carcinogenic processes affecting the AdjNLT samples, as they exist in proximity to the tumor, or to environmental toxins involved in the etiology of the individuals LUAD development. The presence of *SULT1C2* in these normal lungs would predispose affected individuals for conversion of substrates into metabolically activated electrophiles that have the potential to be both carcinogenic and mutagenic(12–14). The amount of cigarette smoke inhaled could also be of great significance. Phase II enzymes can become over-saturated with substrate and experience substrate inhibition. In our study we used many levels of CSC to simulate various smoking conditions, and found that very low levels, mimicking exposure of second-hand smokers, had the greatest increase in expression of *SULT1C2*.

Secondly, we also determined that methylation of the CpG-poor *SULT1C2* promoter regulated downstream transcriptional activity. The paradigm surrounding the function of DNA methylation at CpG-poor promoters is now shifting toward the idea that both CpG-poor and CpG-rich promoters appear to be repressed by DNA methylation(49),(53). We further observed that CSC can stimulate *SULT1C2* promoter

activity, and that this stimulation was lost when the promoter was methylated *in vitro*. It is well known that cigarette smoke can alter DNA methylation(65–67). We now demonstrate that DNA methylation can affect the cigarette smoke response; it can inhibit the ability of CSC to activate gene expression of a key member in the detoxification pathway.

Third, we examined the underlying role of observed differences in DNA methylation levels between different cell line models of LUAD and discovered that race can play a significant role in levels of DNA methylation at the *SULT1C2* promoter as well as overall expression of the *SULT1C2* gene. Indeed, we observed hypomethylation of the *SULT1C2* promoter in AdjNLT from Asian relative to Caucasian patients, with a correlative increase in expression, in data from both TCGA and EDRN consortia. This could indicate a unique underlying predisposition for people of Asian descent to have hypomethylated and expressed *SULT1C2* without the activating event of environmental toxin exposure and may contribute to the well-documented racial disparities in clinical presentation of LUAD in these different patient populations.

Lastly, we determined that the transcription factor AHR has a DNA binding site in the promoter of *SULT1C2* and binding of AHR to the *SULT1C2* promoter is activated by cigarette smoke. Indeed, the fold enrichment of AHR on the *SULT1C2* promoter was significantly greater than that of the previously characterized AHR binding site at *CYP1B1*(60). While we established that AHR was expressed in all 3 cell lines prior to CSC treatment, it is possible that total AHR levels could also be upregulated in response to CSC exposure. Indeed, AHR levels have been previously reported to change in response to cellular xenobiotic exposure(68). Future studies could characterize three-dimensional interactions between transcription factor complexes assembled on the *SULT1C2* promoter to examine any possible interactions between DNA methylation regulatory complexes and AHR DNA binding sites that interact to coordinate cellular responses to xenobiotic compounds by activating sulfotransferases.

Conclusions

Our analyses indicate a complex role for *SULT1C2* in lung adenocarcinoma. On the one hand, higher *SULT1C2* expression in LUAD is associated with improved survival, which would be in line with the detoxification role of the encoded enzyme. On the other hand, we detected higher expression of *SULT1C2* in the AdjNLT of Asian vs. Caucasian patients. Because *SULT1C2* can also activate the carcinogenic potential of xenobiotics, this may point to a potential role in the etiology of LUAD in these patients, who are largely non-smokers, but who could be exposed to secondhand smoke or other environmental agents. Analysis of patient LUAD and AdjNLT samples indicates that DNA methylation is inversely correlated with *SULT1C2* expression, which is supported by our observation that treatment of PC3_LUAD cells with 5-aza-CdR restores *SULT1C2* expression. CSC induces binding of AHR, a factor known to bind and respond to xenobiotic compounds, to the *SULT1C2* promoter. Taken together, our data indicates that the interplay of DNA methylation, tobacco smoke exposure, *SULT1C2* expression could play a pivotal role in the etiology and observed racial disparities observed in LUAD.

Nonstandard Abbreviations

CSC: Cigarette smoke condensate

LUAD: Lung adenocarcinoma

AdjNTL: Adjacent non-tumor lung

XRE: Xenobiotic response element

AHR (Aryl hydrocarbon receptor)

PAH: Polycyclic aromatic hydrocarbons

5-aza-CdR: 5-Aza-deoxycytidine

ETS: Environmental tobacco smoke

TSS: Transcriptional start site

ALU: *Arthrobacter luteus* repetitive element

RLU: Relative light units

Declarations

Ethics Approval and consent to participate

Not applicable

Consent for Publications

Not applicable

Availability of Data and Materials

All next-generation sequencing and microarray data used in this manuscript is publicly available. Microarray data on LUAD expression and methylation generated by the Early Detection Research Network (EDRN) and Ontario Tumor Bank is available at GEO as parts of SuperSeries GSE32867(42) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32867>). Shotgun whole genome bisulfite sequencing (sWGBS) on PC3_LUAD and H2347 cells is available for download from the DataBank of Transcriptional Start Sites (DBTSS) website (<https://dbtss.hgc.jp/>)(38). Enhanced reduced representation bisulfite sequencing (eRRBS) on BEAS-2B is available from GEO (GSE155615)

(36) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155615>). Data from The Cancer Genome Atlas on LUAD expression, methylation, and smoking status is available through the Genomics Data Portal (<https://portal.gdc.cancer.gov/>) and accessible via the TCGAbiolinks package in R(41). Bulk RNAseq is available at GEO for BEAS-2B (GSE55215)

(46) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55215>) as well as H2347 and PC3_LUAD (GSE110024)(31) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110024>).

Competing Interests

The authors declare that they have no competing interests.

Funding

Major funding for this project was provided by the Tobacco-Related Disease Research Program grant to IAO #26IR-0019. C. Marconett was additionally supported by the Donald and Delia Baxter Foundation, STOP Cancer, the Department of Surgery, the American Cancer Society – Research Scholar Grant [RSG-20-135-01-RMC] and the Department of Defense Lung Cancer Research Program IDEA award [W81XWH-21-1-0231].

Authors Contributions

Candace Johnson, Suhaida A. Selamat, Mihaela Campan, and Ite A. Offringa designed the experiments; Candace Johnson performed experiments; Candace Johnson, Daniel J. Mullen, and Crystal N. Marconett analyzed data; Candace Johnson and Crystal N. Marconett wrote the paper.

Acknowledgements

We thank Dr. T. Ryan Stueve for discussion of methodology and toxicology and Dr. Kimberly D. Siegmund for assistance interpreting the statistical significance of findings from the TCGA dataset.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2011;61(2):69–90.
2. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64(1):9–29.
3. Lu TP, Tsai MH, Lee JM, Hsu CP, Chen PC, Lin CW, et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev*. 2010;19(10):2590–7.

4. Ko YC, Lee CH, Chen MJ, Huang CC, Chang WY, Lin HJ, et al. Risk factors for primary lung cancer among non-smoking women in Taiwan. *Int J Epidemiol.* 1997;26(1):24–31.
5. Giovino GA, Biener L, Hartman AM, Marcus SE, Schooley MW, Pechacek TF, et al. Monitoring the tobacco use epidemic I. Overview: Optimizing measurement to facilitate change. *Prev Med.* 2009;48(1 Suppl):S4-10.
6. Subramanian J, Govindan R. Lung cancer in 'Never-smokers': a unique entity. *Oncology (Williston Park).* 2010;24(1):29–35.
7. Ko YC, Cheng LS, Lee CH, Huang JJ, Huang MS, Kao EL, et al. Chinese food cooking and lung cancer in women nonsmokers. *Am J Epidemiol.* 2000;151(2):140–7.
8. Herzog TA, Pokhrel P. Ethnic differences in smoking rate, nicotine dependence, and cessation-related variables among adult smokers in Hawaii. *J Community Health.* 2012;37(6):1226–33.
9. Lei F, Lee E, Zheng Y. Trajectory of smoking behavior change among Chinese immigrant smokers. *PLoS One.* 2021;16(2):e0246280.
10. Gamage N, Barnett A, Hempel N, Duggleby RG, Windmill KF, Martin JL, et al. Human sulfotransferases and their role in chemical metabolism. *Toxicol Sci.* 2006;90(1):5–22.
11. Allali-Hassani A, Pan PW, Dombrovski L, Najmanovich R, Tempel W, Dong A, et al. Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biol.* 2007;5(5):e97.
12. Weinshilboum RM, Otterness DM, Aksoy IA, Wood TC, Her C, Raftogianis RB. Sulfation and sulfotransferases 1: Sulfotransferase molecular biology: cDNAs and genes. *FASEB J.* 1997;11(1):3–14.
13. Falany CN. Enzymology of human cytosolic sulfotransferases. *FASEB J.* 1997;11(4):206–16.
14. Yasuda S, Idell S, Fu J, Carter G, Snow R, Liu MC. Cigarette smoke toxicants as substrates and inhibitors for human cytosolic SULTs. *Toxicol Appl Pharmacol.* 2007;221(1):13–20.
15. Her C, Kaur GP, Athwal RS, Weinshilboum RM. Human sulfotransferase SULT1C1: cDNA cloning, tissue-specific expression, and chromosomal localization. *Genomics.* 1997;41(3):467–70.
16. Rondini EA, Fang H, Runge-Morris M, Kocarek TA. Regulation of human cytosolic sulfotransferases 1C2 and 1C3 by nuclear signaling pathways in LS180 colorectal adenocarcinoma cells. *Drug Metab Dispos.* 2014;42(3):361–8.
17. Hecht SS. Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. *Lancet Oncol.* 2002;3(8):461–9.
18. Chao A, Thun MJ, Jacobs EJ, Henley SJ, Rodriguez C, Calle EE. Cigarette smoking and colorectal cancer mortality in the cancer prevention study II. *J Natl Cancer Inst.* 2000;92(23):1888–96.
19. Doll R. Cancers weakly related to smoking. *Br Med Bull.* 1996;52(1):35–49.
20. Port JL, Yamaguchi K, Du B, De Lorenzo M, Chang M, Heerdt PM, et al. Tobacco smoke induces CYP1B1 in the aerodigestive tract. *Carcinogenesis.* 2004;25(11):2275–81.
21. Papoutsis AJ, Selmin OI, Borg JL, Romagnolo DF. Gestational exposure to the AhR agonist 2,3,7,8-tetrachlorodibenzo-p-dioxin induces BRCA-1 promoter hypermethylation and reduces BRCA-1

- expression in mammary tissue of rat offspring: preventive effects of resveratrol. *Mol Carcinog.* 2015;54(4):261–9.
22. Vorrink SU, Hudachek DR, Domann FE. Epigenetic determinants of CYP1A1 induction by the aryl hydrocarbon receptor agonist 3,3',4,4',5-pentachlorobiphenyl (PCB 126). *Int J Mol Sci.* 2014;15(8):13916–31.
 23. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet.* 2010;11(3):191–203.
 24. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A.* 2002;99(6):3740–5.
 25. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A.* 2006;103(5):1412–7.
 26. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet.* 2007;39(4):457–66.
 27. Li R, Mav D, Grimm SA, Jothi R, Shah R, Wade PA. Fine-tuning of epigenetic regulation with respect to promoter CpG content in a cell type-specific manner. *Epigenetics.* 2014;9(5):747–59.
 28. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet.* 2006;38(12):1378–85.
 29. Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, et al. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc Natl Acad Sci U S A.* 2008;105(35):12979–84.
 30. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884-i90.
 31. Shahabi S, Kumaran V, Castillo J, Cong Z, Nandagopal G, Mullen DJ, et al. LINC00261 is an epigenetically-regulated tumor suppressor that is essential for activation of the DNA damage response. *Cancer Res.* 2019.
 32. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics.* 2015;51:11.4.1-9.
 33. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29(1):24–6.
 34. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 2020;48(W1):W509-W14.
 35. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell.* 2018;173(2):400 – 16.e11.
 36. Huang SK, Tripathi P, Koneva LA, Cavalcante RG, Craig N, Scruggs AM, et al. Effect of concentration and duration of particulate matter exposure on the transcriptome and DNA methylome of bronchial

- epithelial cells. *Environ Epigenet.* 2021;7(1):dvaa022.
37. S B, KE P, S D, IV Y, DA S. Fast and accurate alignment of long bisulfite-seq reads. *Arxiv*2014. p. 1129.
 38. Suzuki A, Wakaguri H, Yamashita R, Kawano S, Tsuchihara K, Sugano S, et al. DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. *Nucleic Acids Res.* 2015;43(Database issue):D87-91.
 39. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics.* 2009;10:232.
 40. Nagy Á, Munkácsy G, Györfy B. Pancancer survival analysis of cancer hallmark genes. *Sci Rep.* 2021;11(1):6047.
 41. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 2016;44(8):e71.
 42. Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M, et al. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res.* 2012;22(7):1197–211.
 43. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, et al. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* 2006;16(3):383–93.
 44. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet.* 2006;38(7):787–93.
 45. Györfy B, Surowiak P, Budczies J, Lánczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One.* 2013;8(12):e82241.
 46. Zong FY, Fu X, Wei WJ, Luo YG, Heiner M, Cao LJ, et al. The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS Genet.* 2014;10(4):e1004289.
 47. Wu B. Substrate inhibition kinetics in drug metabolism reactions. *Drug Metab Rev.* 2011;43(4):440–56.
 48. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature.* 2010;466(7310):1129–33.
 49. Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics.* 2014;9(10):1382–96.
 50. Network CGAR. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* 2014;511(7511):543–50.
 51. Srivastava S. The early detection research network: 10-year outlook. *Clin Chem.* 2013;59(1):60–7.
 52. Klug M, Rehli M. Functional analysis of promoter CpG methylation using a CpG-free luciferase reporter vector. *Epigenetics.* 2006;1(3):127–30.

53. Han H, Cortez CC, Yang X, Nichols PW, Jones PA, Liang G. DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Hum Mol Genet.* 2011;20(22):4299–310.
54. Jabbour E, Short NJ, Montalban-Bravo G, Huang X, Bueso-Ramos C, Qiao W, et al. Randomized phase 2 study of low-dose decitabine vs low-dose azacitidine in lower-risk MDS and MDS/MPN. *Blood.* 2017;130(13):1514–22.
55. Tsai HC, Li H, Van Neste L, Cai Y, Robert C, Rassool FV, et al. Transient low doses of DNA-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. *Cancer Cell.* 2012;21(3):430–46.
56. Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D, et al. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res.* 2000;28(8):E32.
57. 57.. Available from: <https://portal.biobase-international.com>.
58. Burbach KM, Poland A, Bradfield CA. Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. *Proc Natl Acad Sci U S A.* 1992;89(17):8185–9.
59. Poland A, Knutson JC. 2,3,7,8-tetrachlorodibenzo-p-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity. *Annu Rev Pharmacol Toxicol.* 1982;22:517–54.
60. Yang X, Solomon S, Fraser LR, Trombino AF, Liu D, Sonenshein GE, et al. Constitutive regulation of CYP1B1 by the aryl hydrocarbon receptor (AhR) in pre-malignant and malignant mammary tissue. *J Cell Biochem.* 2008;104(2):402–17.
61. Sakakibara Y, Yanagisawa K, Katafuchi J, Ringer DP, Takami Y, Nakayama T, et al. Molecular cloning, expression, and characterization of novel human SULT1C sulfotransferases that catalyze the sulfonation of N-hydroxy-2-acetylaminofluorene. *J Biol Chem.* 1998;273(51):33929–35.
62. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
63. Consortium G. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580–5.
64. Abugessaisa I, Ramilowski JA, Lizio M, Severin J, Hasegawa A, Harshbarger J, et al. FANTOM enters 20th year: expansion of transcriptomic atlases and functional annotation of non-coding RNAs. *Nucleic Acids Res.* 2021;49(D1):D892-D8.
65. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am J Hum Genet.* 2016;98(4):680–96.
66. Bakulski KM, Dou J, Lin N, London SJ, Colacino JA. DNA methylation signature of smoking in lung cancer is enriched for exposure signatures in newborn and adult blood. *Sci Rep.* 2019;9(1):4576.
67. de Vries M, van der Plaat DA, Nedeljkovic I, Verkaik-Schakel RN, Kooistra W, Amin N, et al. From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function. *Respir Res.* 2018;19(1):212.

68. Vogel CFA, Van Winkle LS, Esser C, Haarmann-Stemmann T. The aryl hydrocarbon receptor as a target of environmental stressors - Implications for pollution mediated stress and inflammatory responses. Redox Biol. 2020;34:101530.

Figures

Figure 1

Expression of SULT1C2 affects overall LUAD patient survival and is activated by cigarette smoke. A) Survival plots of LUAD patients stratified by SULT1C2 expression. Upper panel included patients that never smoked, lower panel includes only patients with documented smoke exposure. KMplot (45) was used to stratify samples by upper quartile. B) Left: Distribution of SULT1C2 in human LUAD patient tumors and adjacent non-tumor lung from TCGA. Samples are not paired. (***) $p = 7.2e-4$. Right: TCGA LUAD tumors split by smoking history. FPKM-UQ = Fragments per kilobase of gene per millions mapped, upper quartile normalized. C) Expression of SULT1C2 in three lung cell lines. RPKM = reads per kilobase of gene per millions mapped. D-E) LUAD cell lines were treated with DMSO or cigarette smoke condensate (CSC) at 10, 20, 40, and 80 ug/ml. Cells were treated for 24 hours and mRNA expression levels were measured. White = Untreated, grey = DMSO, brown = CSC treatment at indicated doses. D) SULT1C2 levels in BEAS-2B cells, H2347, and PC3_LUAD cells E) CYP1B1 levels in BEAS-2B, H2347, and PC3_LUAD cells. N=3. All qRT-PCR data are normalized to actin levels in matched cell line. N=3. (*) = $p \leq 0.05$.

Figure 2

Methylation of *SULT1C2* promoter is altered in human lung. A) Overall patient survival (OS) for samples from the TCGA dataset were split by upper quartile based on methylation at cg13968390 using TCGAbiolinks(41). B) Boxplots of methylation at cg13968390 derived from TCGA (left), EDRN (middle) and OTB (right) datasets. N = number of samples in collection included in calculations. An unpaired T-test was used on TCGA data as the AdjNTL samples are derived from non-matched individuals in that

dataset. Paired t-tests were used for EDRN and OTB datasets as their AdjNTL was derived from the same patient as the LUAD tumor tissues. Yellow = AdjNTL, purple = LUAD. C) Correlation between expression and methylation in TCGA LUAD tumors (correlation = -0.431, $p = 1.53 \times 10^{-19}$). D) Correlation between expression (probe ILMN_1772148) and methylation (cg13968390) from the EDRN AdjNTL dataset.

Figure 3

DNA methylation represses transcription of SULT1C2 promoter. A) IGV browser image of SULT1C2 promoter region. RNAseq (green) was downloaded from publicly available resources(30), (31). BAM alignment files to hg19 genome displayed. DNA includes RRBS (BEAS-2B) and sWGBS (H2347, PC3_LUAD); blue = unmethylated, red = methylated, grey = non-CpG sequence. B) Cells were transfected with in vitro SssI methylated or unmethylated SULT1C2-CpGless vector in the indicated cell lines. Cells were subsequently treated with 20ug/mL CSC for 24 hours. Samples were background subtracted and normalized to CMV-luciferase positive controls as a measure of transfection efficiency. RLUs are expressed as a ratio relative to unmethylated, untreated controls in the indicated cell line. Blue = unmethylated SULT1C2 promoter, red = methylated SULT1C2 promoter. Black outlines indicate CSC-treated samples. N=3. A two-tailed paired T test was used to calculate significance and Bonferroni correction applied for multiple tests; (*) = $p \leq 0.05$, (**) = $p < 0.01$, (***) $p < 0.001$. C) PC3_LUAD cells were treated with 5-aza-CdR at half the clinical dose (0.15 uM) and the clinical dose (0.30 uM). Pyrosequencing was used to determine differential methylation by percent methylated reference (PMR) to ALU. SULT1C2 expression was normalized to ACTB loading controls. N=3.

Figure 4

SULT1C2 expression is elevated in Asian adjacent non-tumor lung relative to Caucasians and is significantly correlated to SULT1C2 promoter methylation levels. A) Boxplots of SULT1C2 expression (ILMN_1772148) and B) SULT1C2 promoter methylation (cg13968390) derived from the EDRN AdjNTL dataset. N = number of samples in collection included in calculations. An unpaired t-test was used to calculate significance as the tissues were derived from different patients. Turquoise = Asian, sandy brown

= White. C) Genome-wide analysis of significant differences in methylation between Asian and Caucasian in AdjNTL from patients in EDNR. Q values were used for genome-wide FDR correction of significance. Methylation beta value differences are expressed as changes in Asian methylation levels relative to Caucasian. Red circle = cg13968390. D) Scatterplot of EDNR AdjNTL methylation (cg13968390) versus expression (ILMN_1772148). Turquoise = Asian, sandy brown = White.

Figure 5

CSC alters AHR occupancy at the SULT1C2 promoter. A) Diagram of the SULT1C2 promoter including CpG at cg13968390, including the predicted AHR binding site. Open circles indicate CpG dinucleotides within the 1.8kb promoter region. Coordinates listed are relative to the transcriptional start site of SULT1C2. B) Scatterplot of expression for indicated genes. Data included was derived from the TCGA LUAD dataset and visualized using TIMER2.0(34). Cor = correlation between expression of the indicated genes. C) IGV browser image of AHR and ARNT genomic regions. RNAseq (blue) was downloaded from publicly available resources(30), (31). D) The indicated cell lines were treated with 20ug/mL CSC for 72 hrs. ChIP was performed with AHR antibody. Grey = AHR bound to promoter when DMSO treated, Brown = AHR enrichment on SULT1C2 promoter when treated with 20 µg/ml CSC for 24 hours. Normal rabbit IgG was used to correct for background binding levels.