Check for updates

SETDB1 Overexpression Sets an Intertumoral Transcriptomic Divergence in Non-small Cell Lung Carcinoma

Yong-Kook Kang[1,2]* and Byungkuk Min[1]

[1] Development and Differentiation Research Center, Korea Research Institute of Bioscience Biotechnology, Daejeon, South Korea, [2] Department of Functional Genomics, Korea University of Science and Technology, Daejeon, South Korea

An increasing volume of evidence suggests that SETDB1 plays a role in the tumorigenesis of various cancers, classifying SETDB1 as an oncoprotein. However, owing to its numerous protein partners and their global-scale effects, the molecular mechanism underlying SETDB1-involved oncogenesis remains ambiguous. In this study, using public transcriptome data of lung adenocarcinoma (ADC) and squamous-cell carcinoma (SCC), we compared tumors with high-level SETDB1 (SH) and those with low-level SETDB1 (comparable with normal samples; SL). The results of principal component analysis revealed a transcriptomic distinction and divergence between the SH and SL samples in both ADCs and SCCs. The results of gene set enrichment analysis indicated that genes involved in the "epithelial–mesenchymal transition," "innate immune response," and "autoimmunity" collections were significantly depleted in SH tumors, whereas those involved in "RNA interference" collections were enriched. Chromatin-modifying genes were highly expressed in SH tumors, and the variance in their expression was incomparably high in SCC-SH, which suggested greater heterogeneity within SCC tumors. DNA methyltransferase genes were also overrepresented in SH samples, and most differentially methylated CpGs (SH/SL) were undermethylated in a highly biased manner in ADCs. We identified interesting molecular signatures associated with the possible roles of SETDB1 in lung cancer. We expect these SETDB1-associated molecular signatures to facilitate the development of biologically relevant targeted therapies for particular types of lung cancer.

Keywords: lung cancer, SETDB1, intertumor heterogeneity, epithelial-mesenchymal transition, RNA interference

Abbreviations: ADC, adenocarcinoma; SCC, squamous-cell carcinoma; SCLC, small-cell lung carcinoma; NSCLC, non-small-cell lung carcinoma; SH and SL, SETDB1-high and -low; TCGA, The Cancer Genome Atlas; CNA, copy number alteration; GSEA, gene set enrichment analysis; fGSEA, fast pre-ranked GSEA; DEGs, differentially expressed genes; ES, enrichment score; NES normalized ES; EMT, epithelial–mesenchymal transition; RNAi, RNA interference; DE, differential enrichment; CpGs, CpG dinucleotides; DMCs, differentially methylated CpGs; CGI, CpG island; ERV, endogenous retrovirus; TE, transposable element.

INTRODUCTION

Lung cancer, which is the leading cause of cancer-related morbidity and mortality worldwide (Jemal et al., 2009), can be categorized into two main clinicopathological categories, namely, small-cell lung carcinoma (SCLC) and non-SCLC (NSCLC). NSCLC accounts for 85% of all lung cancers and is further classified into the following three subtypes: major adenocarcinoma (ADC, ∼40% of all lung cancers), squamous-cell carcinoma (SCC, ∼30% of all lung cancers), and large-cell carcinoma, the occurrence of which is relatively low (Sher et al., 2008). SCC arises from squamous cells in the airway epithelium of the bronchial tubes from the center of the lung, whereas ADC arises from small airway epithelial cells that secrete mucus and other substances (Noguchi et al., 1995). All major histological subtypes of NSCLC are associated with smoking; the occurrence of SCC is strongly associated with smoking compared with ADC, and the latter is the most common histology in never smokers (Sun et al., 2007).

Lung cancer is a molecularly heterogeneous disease, and grasping the underlying biology is essential to develop effective therapies against it. This is especially the case if the relatively higher mutation rates in NSCLC are considered; the mutation frequency has been shown to be 3.5 and 3.9 per megabase (Mb) in ADC and SCC, respectively, and these values are approximately twofold higher than the mean rate of 1.8 per Mb across all tumor types (Kan et al., 2010). A high mutation frequency indicates a broad diversity and heterogeneity within that tumor. The different levels of heterogeneity in cancer are as follows: intertumor heterogeneity, which refers to the diversity between the primary tumor and its metastases, and intratumor heterogeneity, which refers to the subclonal diversity within a single tumor (Burrell et al., 2013; Rich, 2016). Lines of evidence support the fact that lung cancer is composed of subsets of cells and clones with distinct molecular features, even within the same histological subtype (reviewed in Inamura, 2017; Herbst et al., 2018). This tumor heterogeneity, as a source of concern in different tumor responses to treatments, has an impact on the characterization of actionable targets, treatment planning, and drug resistance (Swanton, 2012; Zhang et al., 2014). Therefore, research on tumor heterogeneity should be extended from molecular profiling to epigenetic, phenotypic, and transcriptomic assessment through regional DNA methylation, chromatin state, and RNA and/or protein expression studies over time and during treatment to fully comprehend the impact of tumor heterogeneity on the biology of this cancer type and its impact on the clinical phenotype of patients with cancer (Zhang et al., 2014).

Human cancer genomics studies have recognized that point mutations, translocations, deletions, and gene amplification events frequently occur in genes encoding histone-modifying enzymes such as histone methyltransferases, demethylases, acetyltransferases, and deacetylases (Esteller, 2006; Rodriguez-Paredes and Esteller, 2011). One of the best examples of histone methyltransferases is SETDB1 (SET domain, bifurcated 1). It catalyzes the synthesis of trimethylated histone H3 lysine 9 (H3K9me3), which is a repressive mark to which heterochromatin protein 1 (HP1) is recruited and deposited,

thus inducing heterochromatin formation in that region (Schultz et al., 2002). An increasing volume of recent evidence supports that SETDB1 plays a crucial role in the tumorigenesis of various cancers, validating the classification of SETDB1 as an oncoprotein (reviewed in Kang, 2018). SETDB1 gene copy number amplification (CNA) has been frequently observed in various cancers, including melanoma (Ceol et al., 2011), lung SCLC, and NSCLC (Rodriguez-Paredes et al., 2014), hepatocellular carcinoma (Wong et al., 2016), and breast cancer (Regina et al., 2016); a strong correlation between SETDB1 overexpression and cancer development has been detected in various cancers (Kang, 2018). In a study on lung cancer, SETDB1 overexpression correlated with the clonogenicity and tumor size in a xenograft model (Rodriguez-Paredes et al., 2014). In another study (Lafuente-Sanchis et al., 2016), SETDB1 overexpression was proposed to be a prognostic marker to predict tumor recurrence in patients with early stage NSCLC.

Although the findings of studies on SETDB1 have consistently indicated its oncogenic role in various cancers, the broad ranges of genomic targets and protein partners of SETDB1 (usually, transcription factors and associated epidrivers), along with their global effects, make it challenging to prove an immediate connection of SETDB1 with certain tumors. Therefore, the molecular mechanism of SETDB1-involved oncogenesis remains largely unknown. In this study, based on the SETDB1 levels, we attempted subtyping of the lung ADC and SCC samples to measure how SETDB1 overexpression quantitatively and qualitatively alters the global gene expression profiles of ADC and SCC samples. For this, we utilized public lung transcriptome datasets available on The Cancer Genome Atlas (TCGA) site and analyzed the gene sets involved in the cellular processes that are suggested to be associated with SETDB1 function. We identified interesting molecular signatures associated with the roles of SETDB1 in lung cancer and hope that these findings facilitate the development of biologically relevant targeted therapies for particular lung cancer types.

MATERIALS AND METHODS

To investigate the effect of SETDB1 overexpression in non-small cell lung cancers (NSCLC), we utilized the genomic datasets of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) deposited in The Cancer Genome Atlas (TCGA, 2018). All plots were generated under R-environment or by MS-EXCEL.

SETDB1 Gene Copy-Number Alteration and Expression Analysis

To examine the relationship between the SETDB1 gene copy number and the expression level, the copy-number alteration (CNA) data were download from cBioPortal, and TCGA RNA-seq HTseq-count datasets of LUAD (534 tumor and 59 normal samples) and LUSC carcinoma (503 tumor and 49 normal samples) were downloaded using "TCGAbiolinks" (Colaprico et al., 2016; Mounir et al., 2019). The RNA-seq count data were normalized using DESeq2 (Love et al., 2014), and the SETDB1 expression levels were extracted. The CNA data and

the RNA-seq data were joined together by the sample names, and then the samples were grouped by the types of CNA (amplification, gain, diploid, or loss) and histology (LUAD or LUSC) to generate box plots for the *SETDB1* gene CNA vs. the expression levels. Correlations between SETDB1 CNA and expression in ADC and SCC were calculated using "cor" function with "method = pearson."

The tumor samples were ranked by *SETDB1* levels and the top 20% ($n \cong 100$) and the bottom 20% ($n \cong 100$) of them were arbitrarily chosen as SETDB1-high (SH) and SETDB1-low (SL) group, respectively, for transcriptomic analysis. In ADC samples, the mean expression levels of the selected SH and SL groups were 4,211.15 (ranging from 16,371.35 to 2,983.01) and 1,562.89 (1,723.82–873.64) in normalized counts, respectively, while in SCC samples they were 3,691.52 (13,308.90–3,002.55) and 1,468.37 (1,857.16–1,018.27), respectively. Regarding normal samples, we used all normal samples ($n = 108$ in total) provided in the TCGA ADC and SCC datasets weather they were paired with tumor samples or not. Of 108 samples, 59 were from the TCGA ADC expression datasets and 49 were from the SCC datasets, with 33 and 17 of them paired with the tumor samples.

Survival Probablility Analysis

For survival analysis, clinical data for TCGA LUAD (ADC) and LUSC (SCC) patients were downloaded using "GDCquery," "GDCdowload," and "GDCprepare_clinic" functions in "TCGAbiolinks," and datasets for SETEDB1 high and low patients were extracted and labeled as "Top100" and "Bot100," respectively. Survival analysis was performed, and result was plotted using "TCGAanalyze_survival" function in "TCGAbiolinks."

Cell Culture, Transfection, and Real-Time PCR

The human NSCLC cell line A549 was maintained in RPMI 1640 medium (Gibco, United States) supplemented with 10% heat-inactivated fetal bovine serum (Hyclone, United States), 2 mM glutamine and antibiotics (100 U/ml penicillin and 100 μg/ml streptomycin) at 37°C in a humidified atmosphere of 5% CO_2. Cells were transfected with 4 μg of a CMV promoter-driven RFP-SETDB1 expression plasmid per well using Lipofectamine 3000 (Invitrogen) and were observed for RFP signal in fluorescence miscroscope immediately before harvest 48 h after transfection.

Total RNAs were extracted from A549 cells and reverse transcription was performed by incubating 1 μg of DNase I−pretreated RNA with Superscript III enzyme (Invitrogen), 20 μM oligo−dT primers (Invitrogen), and 50 ng random hexamers (Invitrogen) at 50°C for 1 h. Five nanograms of the synthesized cDNAs was used for real−time quantitative PCR (QuantStudio3 Real−Time PCR system, ABI) with the specific primer sets for individual target genes (**Supplementary Table S1**). PCR was performed with following program: 5 min of pre-denaturation at 95°C followed by 40 cycles of 95°C for 15 s, 55°C for 15 s, and 72°C for 1 min. Finally, relative expression levels of target genes against *GAPDH* gene were calculated using QuantStudio Design and Analysis Software (Thermo).

Differential Expression Analysis and GSEA

To explore the *SETDB1* expression dependent transcriptomic divergences in lung cancer samples, the TCGA LUAD, or LUSC samples were ordered based on the expression levels of *SETDB1*, and top-ranked and bottom-ranked 100 samples were selected; they were named as SH and SL groups, respectively. For t-distributed stochastic neighbor embedding (*t*-SNE), "Rtsne"[1], a R wrapper for *t*-SNE, was used with "perplexity = 30" and "pca = true." Then, transcriptomic data of SH and SL samples were extracted from the earlier DESeq2-normalized expression dataset for the differential expression analysis. For gene set enrichment analysis (GSEA), gene collections for "HALLMARK," "KEGG pathways," and "GO Biological Process" from MSigDB v7.0 (Liberzon et al., 2011, 2015) were obtained, and gene sets enriched in either SH or SL group were identified using fGSEA (Korotkevich et al., 2019).

In addition to the fGSEA that was generated based on group means, single sample GSEA (ssGSEA) was performed using gene set variation analysis (GSVA) (Hanzelmann et al., 2013), a R package for gene set variation analysis among individual samples. First, the normalized count data from "DESeq2" were converted to an input matrix for "GSVA" by a home-brew R code, and enrichment scores (ES) for individual samples were calculated by the "gsva" function with "method = gsva." Then, using "limma" (Ritchie et al., 2015), a R package for differential expression analysis, gene sets with significantly altered activations (FDR $< 1 \times 10^{-5}$) were identified, and the results were visualized on volcano plots and heatmaps generated by "plot" and "ggplot" functions in R.

DNA Methylome Analysis

To compare the methylation states of SETDB1-high and SETDB1-low lung cancer samples, Illumina Human Methylation 450 data (level 3) of TCGA of LUAD and LUSC samples were downloaded using "TCGAbiolinks," and the pre-processed and scaled methylation beta-values for SH and SL samples were extracted. Next, beta-values were converted to *M*-values for variance stabilization, and differentially methylated CpG sites (DMCs) were identified using "limma" ($p_{adj} < 0.01$, $\log_2 FC > 0.5$) and annotated with genomic features (e.g., genes and CGIs, hg38 from UCSC genome browser) using "bedtools" (Quinlan and Hall, 2010). Finally, differential methylation between SH and SL samples was examined by surveying the methylation fold-change distributions of CpGs on annotated CGIs and genes.

RESULTS

Enhanced SETDB1 Expression Sets a Transcriptomic Distinction Among Lung Tumors

In this study, we investigated the effect of increased *SETDB1* transcript levels in the two most frequent histologic subtypes
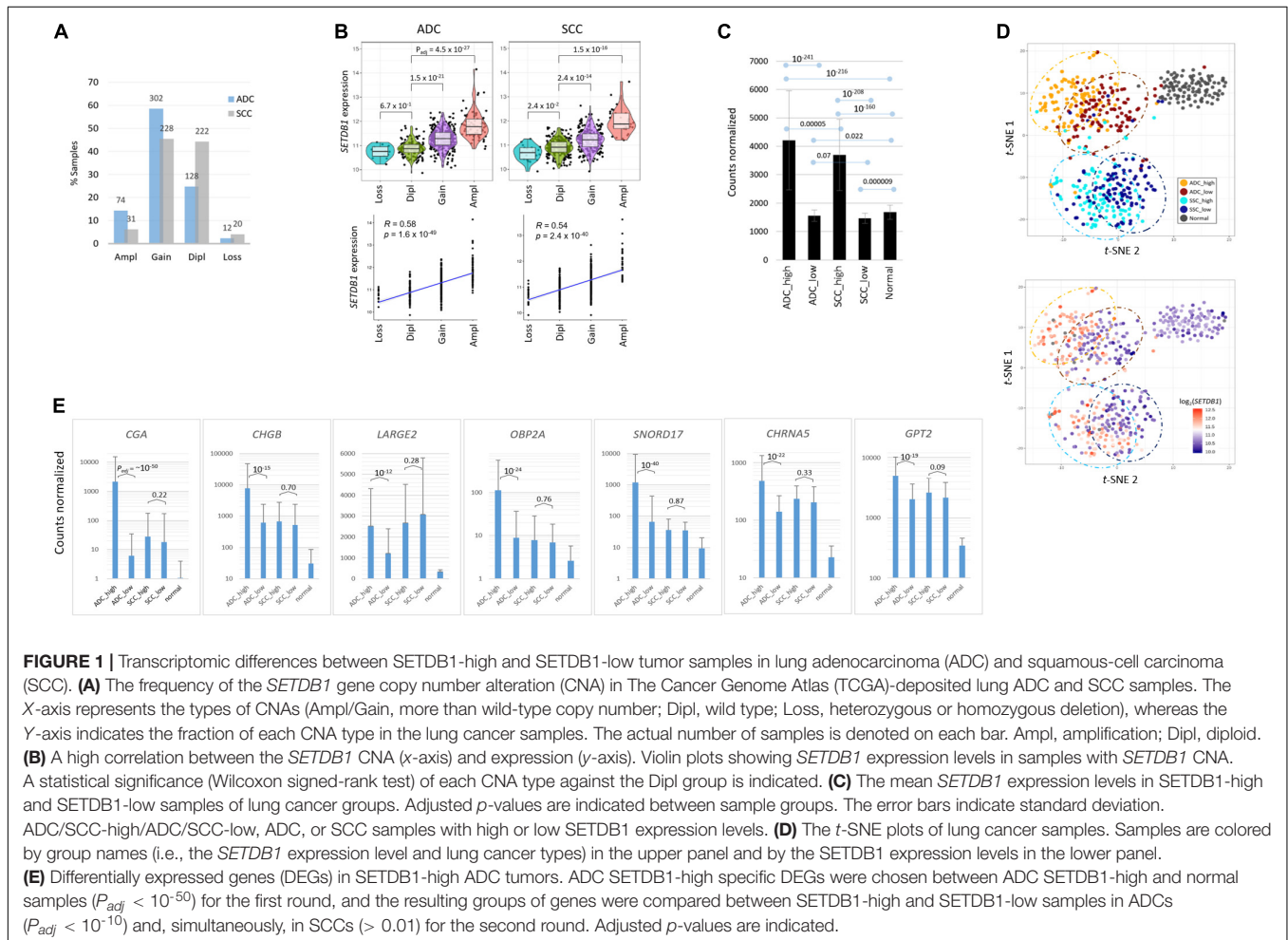
---

[1]https://github.com/jkrijthe/Rtsne

of NSCLC, ADC, and SCC. We inspected the genome and transcriptome data that are publicly available from the cBioPortal and The Cancer Genome Atlas (TCGA). The *SETDB1* gene amplification was more frequent in ADC samples than in SCC ones [14.3% (74/516) vs. 6.2% (31/501); **Figure 1A**]. We observed a correlation between the *SETDB1* copy number and the *SETDB1* expression level in both types of cancer ($p < 2.4 \times 10^{-40}$; **Figure 1B**). We sorted the sample data based on the *SETDB1* levels and, then, classified them into the following two groups: a top-ranked (*SETDB1*-high, SH; $n = 100$) group and a bottom-ranked (*SETDB1*-low, SL; $n = 100$) group. The mean fold difference between the SH and SL groups was 2.69 and 2.53 in the ADCs and SCCs, respectively (**Figure 1C**). The *SETDB1* level significantly differed between SH group samples and the normal samples, whereas the levels in the SL group samples and the normal samples were close. *t*-Distributed Stochastic Neighbor Embedding (t-SNE) analysis revealed a divergence between the SH and SL samples within each cancer type, along with a vivid distinction between the cancer samples and normal samples or between the ADCs and SCCs (**Figure 1D**). We examined the effect of SETDB1 overexpression on the survival rates of lung cancer patients and found no significant

difference in the 5 years survival rates between SH and SL tumors (**Supplementary Figure S1**).

To select marker genes that are specific for the SH samples against the SL samples, we first chose genes that are differentially expressed in the ADC-SH and SCC-SH samples compared with the normal samples ($P_{adj} < 1 \times 10^{-50}$), and then, among the chosen genes, we selected the ADC-specific ($P_{adj} < 1 \times 10^{-10}$ for ADCs and > 0.01 for SCCs in the comparison between SH and SL samples) or the SCC-specific SH marker genes ($P_{adj} < 1 \times 10^{-10}$ for SCCs and > 0.01 for ADCs). The ADC markers, including *CGA, SNORD17, SPINK4, OBP2A, BPIFA2, CHGB,* and *SMKR1,* are representatively shown in **Figure 1E** (see **Supplementary Table S2** for the entire gene list).

For validation, we examined another dataset (GSE41271; Sato et al., 2013; $n = 20$ for the SH and SL groups each; fold change = 2.28) from the lung ADC samples publicly available (**Supplementary Figure S2A**). The dataset was obtained using a microarray platform and thus largely differed in terms of the numbers and types of transcripts from RNA-seq datasets; nevertheless, approximately 40% of the DEGs (21/52; $P_{adj} < 0.00001$) were found to be shared with the present dataset (**Supplementary Figure S2B**). In addition, we transiently

FIGURE 1 | Transcriptomic differences between SETDB1-high and SETDB1-low tumor samples in lung adenocarcinoma (ADC) and squamous-cell carcinoma (SCC). **(A)** The frequency of the *SETDB1* gene copy number alteration (CNA) in The Cancer Genome Atlas (TCGA)-deposited lung ADC and SCC samples. The *X*-axis represents the types of CNAs (Ampl/Gain, more than wild-type copy number; Dipl, wild type; Loss, heterozygous or homozygous deletion), whereas the *Y*-axis indicates the fraction of each CNA type in the lung cancer samples. The actual number of samples is denoted on each bar. Ampl, amplification; Dipl, diploid. **(B)** A high correlation between the *SETDB1* CNA (*x*-axis) and expression (*y*-axis). Violin plots showing *SETDB1* expression levels in samples with *SETDB1* CNA. A statistical significance (Wilcoxon signed-rank test) of each CNA type against the Dipl group is indicated. **(C)** The mean *SETDB1* expression levels in SETDB1-high and SETDB1-low samples of lung cancer groups. Adjusted *p*-values are indicated between sample groups. The error bars indicate standard deviation. ADC/SCC-high/ADC/SCC-low, ADC, or SCC samples with high or low SETDB1 expression levels. **(D)** The *t*-SNE plots of lung cancer samples. Samples are colored by group names (i.e., the *SETDB1* expression level and lung cancer types) in the upper panel and by the SETDB1 expression levels in the lower panel. **(E)** Differentially expressed genes (DEGs) in SETDB1-high ADC tumors. ADC SETDB1-high specific DEGs were chosen between ADC SETDB1-high and normal samples ($P_{adj} < 10^{-50}$) for the first round, and the resulting groups of genes were compared between SETDB1-high and SETDB1-low samples in ADCs ($P_{adj} < 10^{-10}$) and, simultaneously, in SCCs (> 0.01) for the second round. Adjusted *p*-values are indicated.

transfected A549 LUAD cells with *GFP-SETDB1* expression plasmid, and found from a quantitative real-time PCR that the DEGs that were overexpressed in the SH tumor samples were similarly overrepresented in the GFP-SETDB1 expressing A549 cells compared with the control A549 cells (**Supplementary Figure S3**). This finding highlights the fidelity of the markers that we detected and suggests that these marker genes serve to discriminate SH tumors from SL tumors, ADCs from SCCs, and lung cancer samples from normal samples in multiple ways.
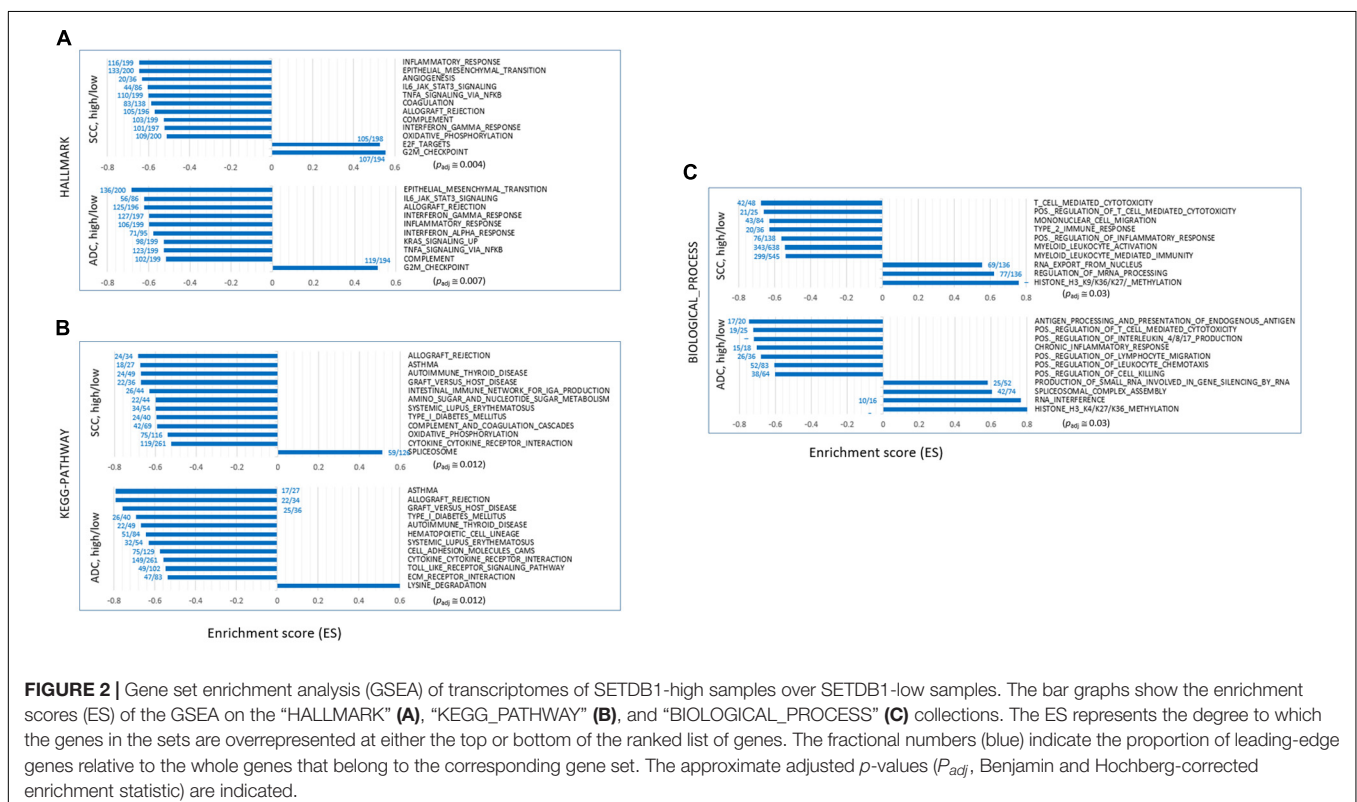
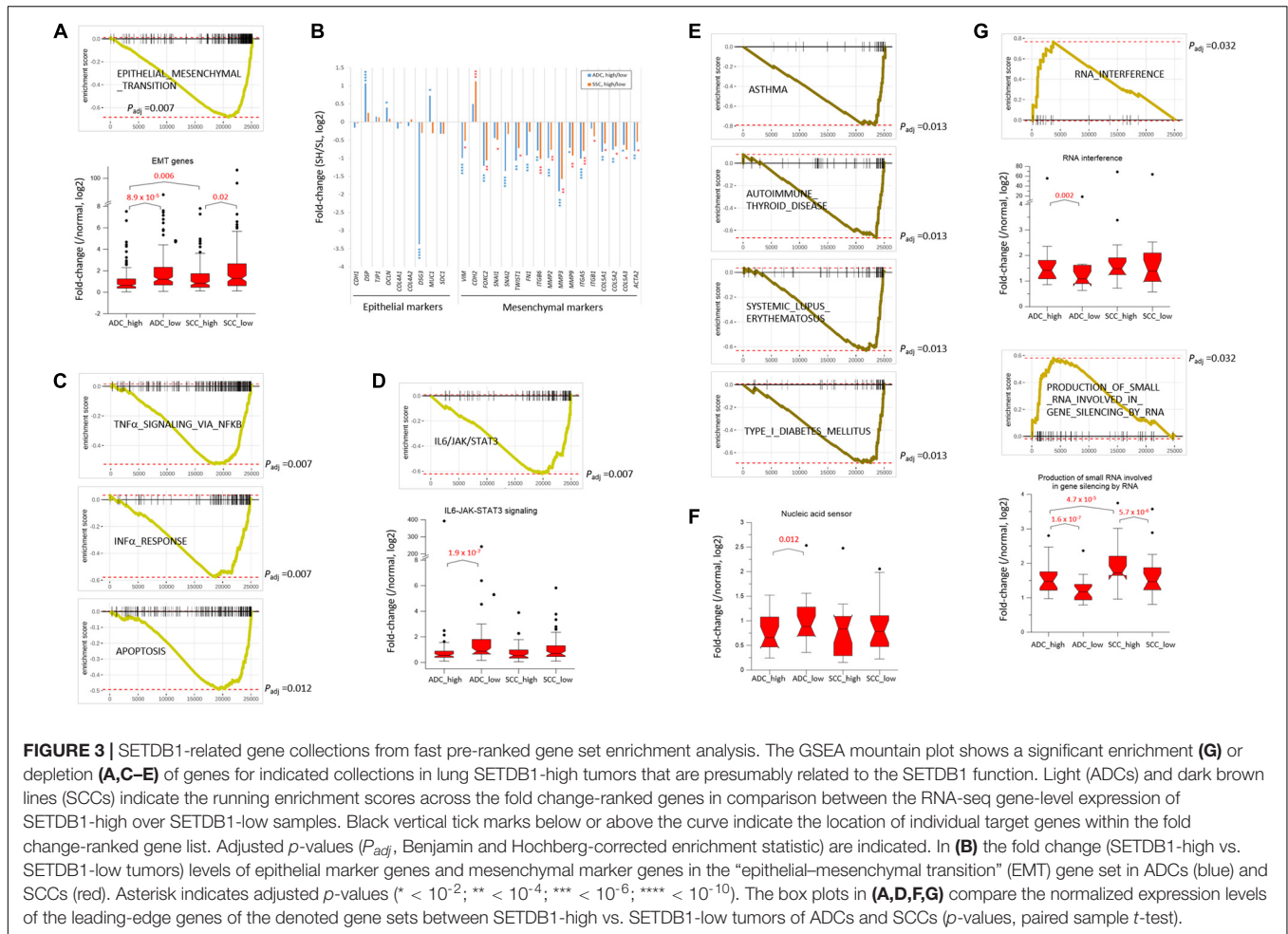## Transcriptomic Features of SETDB1-Overexpressing Lung Cancer Samples

We performed GSEA on the "HALLMARK," "KEGG_PATHWAY," and "BIOLOGICAL_PROCESS" collections using the fast pre-ranked GSEA (fGSEA) package to identify and interpret coordinate changes in the transcriptome of SH samples over the SL samples. When sorted by the enrichment scores [ES; absolute ES value ($|ES|$) > 0.5] and normalized ES (NES; $|NES|$ > 2.0) that estimate the degree to which the genes in the terms are overrepresented at either the top or the bottom of the ranked list of genes, the selected terms mostly showed negative ES/NES values. It indicated that those genes in the designated collections were mostly underrepresented in the SH samples in both the ADC and SCC groups (**Figures 2A–C** and see **Supplementary Table S3**). Overall, the genes in the inflammation- and immune response-related terms including "TNFA_SIGNALING_VIA_NFKB,"

"INTERLEUKIN_PRODUCTION," "INTERFERON_GAMMA_ RESPONSE," "ALLOGRAFT_REJECTION," "COMPLEMENT," and "ANTIGENE_PROCESSING_AND_PRESENTATION" were significantly underrepresented in the SH samples, whereas those in the "G2M_CHECKPOINT," "DNA_REPLICATION_CHECKPOINT," and "E2F TARGETS" terms were overrepresented. Some terms are notable because they were previously shown or hypothesized to be a function of SETDB1, which we have summarized below.

Depletion of Genes Related to Epithelial–Mesenchymal Transition (EMT)

NSCLC mainly arises from the bronchial and alveolar epithelium; thus, it is speculated that EMT plays an important role in the biological behavior of NSCLC and serves as a key mechanism in cancer cell metastasis (Welch and Hurst, 2019). The genes appertained to the "EMT" collection were markedly underrepresented in the SH samples of both ADC and SCC groups (**Figure 3A**). The mean expression levels of the leading-edge genes (the core of a gene set accounting for the enrichment signal) of the "EMT" collection significantly differed between the SH and SL samples, particularly in the ADCs. Although the transcript levels of the epithelial markers in the SH samples relative to those in the SL samples varied largely, all the mesenchymal markers, except *CDH2/NCAD,* were significantly underrepresented in the SH samples in both ADCs and SCCs (**Figure 3B**). In line with this, the genes in the "IL6_JAK_STAT3_SIGNALING" and "TGF_BETA_SIGNALING" collections that are known to induce

FIGURE 2 | Gene set enrichment analysis (GSEA) of transcriptomes of SETDB1-high samples over SETDB1-low samples. The bar graphs show the enrichment scores (ES) of the GSEA on the "HALLMARK" **(A)**, "KEGG_PATHWAY" **(B)**, and "BIOLOGICAL_PROCESS" **(C)** collections. The ES represents the degree to which the genes in the sets are overrepresented at either the top or bottom of the ranked list of genes. The fractional numbers (blue) indicate the proportion of leading-edge genes relative to the whole genes that belong to the corresponding gene set. The approximate adjusted *p*-values (P_{adj}, Benjamin and Hochberg-corrected enrichment statistic) are indicated.

FIGURE 3 | SETDB1-related gene collections from fast pre-ranked gene set enrichment analysis. The GSEA mountain plot shows a significant enrichment **(G)** or depletion **(A,C–E)** of genes for indicated collections in lung SETDB1-high tumors that are presumably related to the SETDB1 function. Light (ADCs) and dark brown lines (SCCs) indicate the running enrichment scores across the fold change-ranked genes in comparison between the RNA-seq gene-level expression of SETDB1-high over SETDB1-low samples. Black vertical tick marks below or above the curve indicate the location of individual target genes within the fold change-ranked gene list. Adjusted p-values (P_{adj}, Benjamin and Hochberg-corrected enrichment statistic) are indicated. In **(B)** the fold change (SETDB1-high vs. SETDB1-low tumors) levels of epithelial marker genes and mesenchymal marker genes in the "epithelial–mesenchymal transition" (EMT) gene set in ADCs (blue) and SCCs (red). Asterisk indicates adjusted p-values (* < 10^{-2}; ** < 10^{-4}; *** < 10^{-6}; **** < 10^{-10}). The box plots in **(A,D,G)** compare the normalized expression levels of the leading-edge genes of the denoted gene sets between SETDB1-high vs. SETDB1-low tumors of ADCs and SCCs (p-values, paired sample t-test).

cancer cell metastasis (Chang et al., 2013; Bellomo et al., 2016) were also depleted (**Supplementary Table S3** and see below) in the SH samples. Our results strongly suggest an inverse relationship between the SETDB1 levels and EMT in lung tumors, which is in disagreement with the notion that overexpressed SETDB1 promotes metastasis, as reported in liver (Fei et al., 2015; Wong et al., 2016) and breast cancers (Weigelt et al., 2010; Ryu et al., 2019); this probably indicates that SETDB1 behaves differently in different cancers.

Depletion of Genes Associated With Host Innate Immune Response

The upregulation of *SETDB1* is suggested to be a common mechanism in tumors to avoid the host innate immune response and apoptosis through a tighter surveillance of transposable element (TE) expression (Cuellar et al., 2017; Kang, 2018). This theory harmonizes well with the reduced expressions of genes in the SH samples that belong to the "INTERFERON_ALPHA_RESPONSE," "TNFα_SIGNALING_VIA_NFKB," and "APOPTOSIS" collections (**Figure 3C**). The genes in the "IL6_JAK_STAT3_SIGNALING" set were underrepresented in the SH samples (**Figure 3D**), and the fold-change difference in expression level of leading-edge

genes was remarkable between the SH and SL samples of ADCs. Because the components of the IL6/JAK/STAT3 pathway are aberrantly hyperactivated in various tumors and targeting these components can inhibit tumor growth (Johnson et al., 2018), lung cancer cells in which the levels of SETDB1 are high are probably supposed to have a natural ability to control the IL6/JAK/STAT3 pathway, creating an environment that is unfavorable for tumor growth.

Depletion of Autoimmunity-Related Genes

Genes in the autoimmunity-related collections ("ASTHMA," "AUTOIMMUNE_THYROID_DISEASE," "SYSTEMIC_LU PUS_ERYTHEMATOSUS," "TYPE_I_DIABETES_MELLITUS," etc.) were largely underrepresented in the SH samples (**Figures 2B, 3E**). This suppression of autoimmunity is consistent with a recent hypothesis stating that SETDB1 overexpression tightly seals off the expression of various genomic TEs, thus blocking self-retroelement-derived nucleic acids (RdNAs) to be piled up in the cytoplasm, which can cause autoimmune diseases such as Aicardi-Goutieres syndrome (Volkman and Stetson, 2014). Notably, the genes encoding various nucleic acid sensors involving *TBK1, MAVS, IFI16, STING1, MDA5, RIG1, DDX41,* and *TLR1-TLR10* that patrol the cytoplasm and induce

type-I interferon responses were significantly underrepresented ($p = 0.012$, paired sample t-test) in the SH samples in the ADCs (**Figure 3F**). The underlying mechanism that links SETDB1 overexpression to the downregulation of nucleic acid sensors in ADCs should be elucidated in the future.
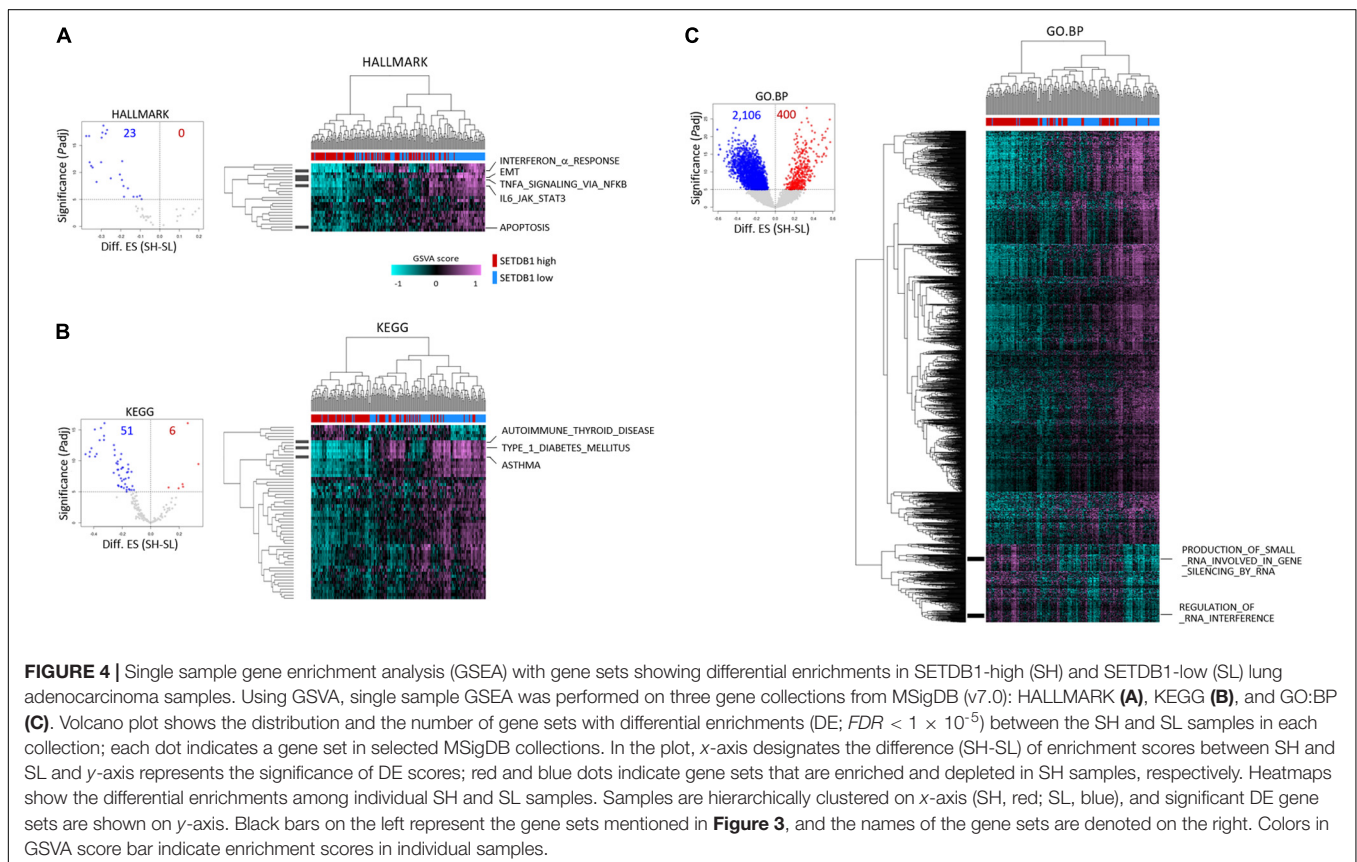
Enrichment of RNA Interference-Related Genes

SETDB1 also participates in the RNA interference pathway, particularly in transcriptional gene silencing (TGS) in association with ARGONAUTE-1 and -2 (Cho et al., 2014). The genes involved in the "RNA_INTERFERENCE (RNAi)" and "PRODUCTION_OF_SMALL_RNA_INVOLVED_IN_GENE_SILENCING_BY_RNA" collections were overrepresented in the SH samples (**Figure 3G**). In line with this, various gene sets related to "small RNA" were enriched ($p < 0.002$) in the ADC-SH samples with positive ES values: "RNA_METABOLIC_PROCESS," "PRE_MIRNA_PROCESSING," "PRODUCTION_OF_SMALL_RNA_INVOLVED_IN_RNA_INTERFERENCE/GENE_SILENCING_BY_RNA," "RNA/NCRNA_EXPORT_FROM_NUCLEUS," etc. (**Supplementary Table S3**: BIOLOGICAL_PROCESS) We speculate that these collections are partly associated with an increased production of endogenous small interfering RNAs and their busy nucleocytoplasmic trafficking to target loci through RNAi pathway in the SH cancer cells (Kang, 2018) although it is unknown how SETDB1 is implicated in the production of small RNAs.

To further test if the ADC and SCC samples could be divided by *SETDB1* level, we performed single sample GSEA using gene set variation analysis (GSVA), which assesses separate enrichment scores for each pairing of a sample and gene set (independent of phenotype labeling) to find out to what extent the genes in a certain gene set are coordinately upregulated or downregulated within a sample (Barbie et al., 2009). The GSVA enrichment scores for the selected gene sets ($FDR < 1 \times 10^{-5}$) showed that, consistent with the result in **Figure 3**, the SH samples showed a depletion of genes in most of the selected gene sets in the ADCs (**Figure 4**) and also in the SCCs (**Supplementary Figure S4**). We found the gene sets which we have noticed from the bulk GSEA result (**Figure 3**) being reproducibly depleted or enriched in the GSVA result.

Expression of the Histone Modification Category Genes in the SETDB1-High Lung Cancer Samples

Because SETDB1 is one of epigenetic players, or epidrivers, that function as a writer, reader, and/or eraser of chromatin marks (Park et al., 2017), the altered expression of SETDB1 may affect other associated epidrivers and their epigenomic signatures. In agreement, the GSEA results revealed that histone lysine (H4K20, H3K27, H3K4, H3K36, and H3K9) methylation-related genes were enriched in the SH samples (**Figure 2C**). We explored the sub-categories of the "HISTONE_MODIFICATION" collection

FIGURE 4 | Single sample gene enrichment analysis (GSEA) with gene sets showing differential enrichments in SETDB1-high (SH) and SETDB1-low (SL) lung adenocarcinoma samples. Using GSVA, single sample GSEA was performed on three gene collections from MSigDB (v7.0): HALLMARK **(A)**, KEGG **(B)**, and GO:BP **(C)**. Volcano plot shows the distribution and the number of gene sets with differential enrichments (DE; $FDR < 1 \times 10^{-5}$) between the SH and SL samples in each collection; each dot indicates a gene set in selected MSigDB collections. In the plot, x-axis designates the difference (SH-SL) of enrichment scores between SH and SL and y-axis represents the significance of DE scores; red and blue dots indicate gene sets that are enriched and depleted in SH samples, respectively. Heatmaps show the differential enrichments among individual SH and SL samples. Samples are hierarchically clustered on x-axis (SH, red; SL, blue), and significant DE gene sets are shown on y-axis. Black bars on the left represent the gene sets mentioned in **Figure 3**, and the names of the gene sets are denoted on the right. Colors in GSVA score bar indicate enrichment scores in individual samples.

for altered gene expression, and the result is summarized in **Table 1**. Collectively, the mean expression levels of the genes in these classes were higher in the SH samples than those in the SL samples in both ADCs and SCCs (**Figure 5A**). The variance of the expression levels of individual histone modification genes was incomparably high in the SCC-SH samples, which indicates an enormous discrepancy and heterogeneity among the SCC samples.

As shown in **Table 1**, the expression levels of the "HISTONE_LYSINE_METHYLATION" and "HISTONE_LYSINE_DEMETHYLATION" category genes differed ($p < 0.02$) between the SH and SL samples in both ADCs and SCCs, and the "HISTONE_UBIQUITINATION" genes differed in the ADCs (**Supplementary Table S4**). We further divided the "HISTONE_METHYLATION" category because of its largest size (122 genes) and the most significant difference ($p < 1.2 \times 10^{-8}$, paired sample t-test) between the SH and SL samples. The genes in the H3K4, H3K9, H3K36, and H4K20 methylation categories significantly differed ($p < 0.01$) in the ADCs, whereas only H3K4 and H3K9 category genes differed in the SCCs (**Figure 5B**). The increase in the expression levels of H3K4 methylation genes was prominent in both the cancer groups, but the upregulation of the H3K4 demethylation category genes was also significant, which could offset the effect of the former on the premise that they share the same sphere of influence over the genome. Conversely, the H3K9 demethylation category genes did not differ between the SH and SL samples, suggesting that both the cancer groups can appreciate the entire impact of the increased expression of the H3K9 methylation genes.

We examined the genes ($n = 154$) encoding effector enzymes and mediators that directly participate in the modification of nucleosomes (Park et al., 2017; Min et al., 2018). **Table 2** presents the differentially expressed epidriver genes between the SH and SL samples in the ADCs and SCCs ($p < 1 \times 10^{-7}$ in either ADCs or SCCs (also see **Supplementary Table S5**). *HR, SETD5, KMT5B, and KDM6B* genes were differentially expressed only in the ADCs, whereas *CBX1, RNF20,* and *RNF38* genes were differentially expressed only in the SCCs; these sets of genes might contribute to the production of characteristic epigenome signatures of ADCs and SCCs, respectively.

Opposite Pattern of DNA Methylation Change in ADC and SCC SETDB1-High Samples
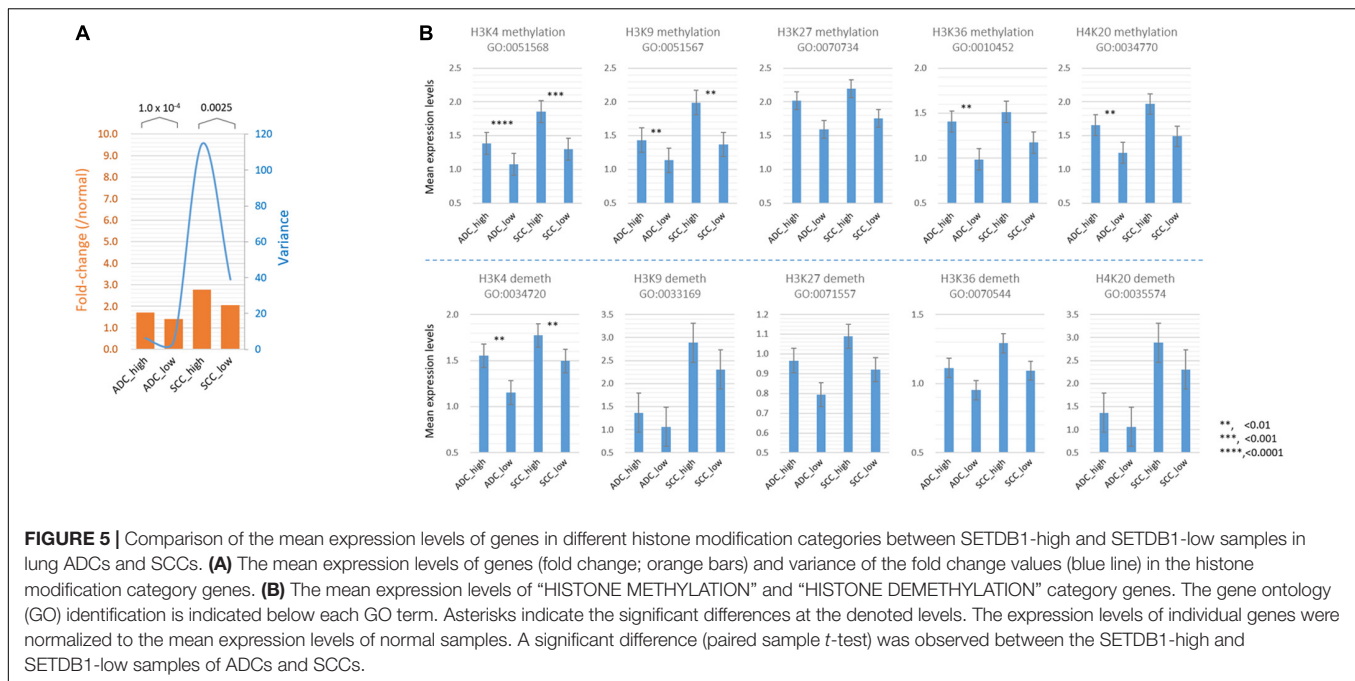
The genes belonging to the DNA methylation category also showed a significant difference between the SH and SL samples (**Table 1**), with the highest root-mean-square deviation (RMSD) value among the "HISTONE_MODIFICATION" sub-categories (**Supplementary Figure S5**), which measures how far off the genes in the cancer samples are from those in the normal samples in terms of the mean expression levels (Kwon et al., 2015; Min et al., 2018). We observed that DNA methyltransferase genes such as *DNMT1, DNMT3A,* and *DNMT3B* were overrepresented in all cancer groups, with a significant difference between the SH and SL samples in both ADCs and SCCs (**Figure 6A** and

TABLE 1 | The mean expression levels of epidriver genes in different "histone modification" categories (GO DAVID).

Histone/DNA modification (GO ID)	No. gene	Mean expression level[a] (SD)				ADC-SH vs. ADC-SL[b]	SCC-SH vs. SCC-SL[b]	ADC-SH vs. SCC-SH[b]	ADC-SL vs. SCC-SL[b]
		ADC-SH	ADC-SL	SCC-SH	SCC-SL				
Acetylation (GO:0016573)	134	1.711 (3.073)	1.627 (3.554)	5.123 (38.413)	4.505 (31.636)	0.6453	0.3187	0.2985	0.2872
Deacetylation (GO:0016575)	71	1.140 (0.461)	1.084 (0.515)	1.228 (0.651)	1.164 (0.599)	0.1747	0.0310	0.0439	0.0660
Lysine methylation (GO:0034770)	122	1.508 (0.960)	1.170 (0.645)	1.921 (1.685)	1.388 (1.038)	2.8E-10	1.2E-08	0.0002	0.0016
Lysine demethylation (GO:0070076)	27	1.276 (0.847)	1.014 (0.365)	2.079 (3.528)	1.711 (2.801)	0.0188	0.0162	0.2671	0.2010
Arginine methylation (GO:0034969)	8	1.868 (1.203)	1.587 (0.835)	1.828 (1.047)	1.928 (1.205)	0.2734	0.4130	0.9294	0.1483
DNA methylation (GO:0044728)	84	2.723 (4.132)	1.666 (1.439)	7.526 (26.109)	4.229 (14.471)	0.0048	0.0207	0.0893	0.0976
Phosphorylation (GO:0016572)	36	2.132 (2.649)	1.998 (2.318)	3.149 (5.130)	3.140 (5.138)	0.3902	0.9445	0.0317	0.0271
Ubiquitination (GO:0016574)	45	1.487 (1.351)	1.289 (1.181)	1.936 (2.815)	1.592 (1.882)	0.0003	0.0285	0.0591	0.0152
Whole (SD)	527	1.720 (2.485)	1.412 (2.071)	2.763 (10.707)	2.067 (6.236)	0.0001	0.0025	0.0198	0.0084

[a]Normalized by the mean levels of genes of normal samples. [b]p-values (paired sample t-test). SH and SL, SETDB1-high and SETDB1-low, respectively. SD, standard deviation.

FIGURE 5 | Comparison of the mean expression levels of genes in different histone modification categories between SETDB1-high and SETDB1-low samples in lung ADCs and SCCs. **(A)** The mean expression levels of genes (fold change; orange bars) and variance of the fold change values (blue line) in the histone modification category genes. **(B)** The mean expression levels of "HISTONE METHYLATION" and "HISTONE DEMETHYLATION" category genes. The gene ontology (GO) identification is indicated below each GO term. Asterisks indicate the significant differences at the denoted levels. The expression levels of individual genes were normalized to the mean expression levels of normal samples. A significant difference (paired sample t-test) was observed between the SETDB1-high and SETDB1-low samples of ADCs and SCCs.

also see **Table 2**), which prompted us to explore the genome for the methylation states. We looked into the public 450 K BeadChip array methylome data obtained from the same tumor samples as those used in the present transcriptome study. The volcano plots showed that the methylation levels at individual CpG sites were mostly reduced in the SH samples in the ADCs, whereas they were increased in the SH samples in the SCCs (**Figure 6B**).

The antithetical pattern of the methylation change was also observed around the CpG island (CGI) regions. In the ADCs, the differentially methylated CpGs identified between the SH and SL samples (DMCs; P_{adj} < 0.01 and fold change > 1.5) were mostly of lower methylation in the SH samples, whereas they were of higher methylation in the SCCs (**Figure 6C**). The pattern was uniform because it was observed at all sub-CGI compartments such as shelfs, shores, and islands as well as in the "open_sea" regions. The DMCs in the genic regions involving distal and proximal promoters, gene body, and 3' untranslated regions (3'-UTRs) were, similar to the CGI DMCs, largely at undermethylated states in the ADC-SH samples (**Figure 6D**). In contrast to the overall consistent and highly biased methylation change in the ADCs, DNA methylation appeared rather randomly altered in the SCCs, considering the similar fractions of DMCs with increasing and decreasing methylation levels. Based on the pattern of methylation change in the lung cancer samples, we assume that SETDB1 overexpression is associated with the uniform change in genomic methylation (i.e., methylation loss) in the ADCs but not in the SCCs, and the underlying mechanism of this is yet to be elucidated. In conclusion, our findings indicate that *SETDB1* overexpression causes changes in the lung cancer transcriptomes and the methylomes, drawing a line between the SH and SL samples and, as a possible result, leads to a within-a-tumor

divergence of cancer cells to a distinctive subpopulation(s), as illustrated in **Figure 6E**.

DISCUSSION

This study presented a "within-a-tumor" subtyping result for ADC and SCC samples. The SH tumor samples were distinguished from the SL tumor and normal samples based on the transcriptomes, and these transcriptomes showed various molecular signatures and interaction networks that are possibly related to SETDB1 functions. Because the SL samples did not greatly differ from the normal samples based on the *SETDB1* level alone, the SH sample group might be more diverged than the SL group among the ADC and SCC sample populations (**Figure 1D**). The SH vs. SL difference was larger in the ADCs than in the SCCs at several points, which are as follows: the fraction of the *SETDB1* gene CNA samples; the mean expression levels of *SETDB1* in the SH samples; the overall ES values and the mean expression levels of leading-edge genes on the GSEA results (**Figures 2, 3**, respectively); the number of SH vs. SL DEGs (**Supplementary Table S2**); and finally, the number of DMCs and the pattern of DNA methylation change (**Figure 6**). Currently, the factors that cause such a difference between the SH samples of ADCs and SCCs are unknown; we assume that ADCs are more sensitive to an altered level of SETDB1, and the histologic difference between the ADCs and SCCs may explain their different sensitivities (Lin et al., 2017). Meanwhile, the SH vs. SL comparison performed in this study is not synonymous with the SH vs. normal comparison because the SL samples, with respect to the transcriptome, were not similar to the normal samples at all and to the SH samples.

The regulation of immunity through the control of TE expression is an emerging research field in cancer, and SETDB1

TABLE 2 | The mean expression levels of differentially expressed epidriver genes in either of ADCs or SCCs.

Category	Gene ID	Expression level					Fold-change (SH/SL)*		P_{adj}-value (SH vs. SL)	
		ADC-SH	ADC-SL	SCC-SH	SCC-SL	Normal	ADC	SCC	ADC	SCC
K meth	SETDB1	4,211.4	1,560.8	3,693.6	1,467.7	1,678.6	1.430	1.331	9.7E-193	1.1E-208
DNA meth	DNMT3A	3,566.2	1,881.1	4,013.7	2,410.1	1,407.5	0.922	0.736	4.8E-28	3.0E-17
K meth	EHMT2	4,885.8	2,705.0	4,936.3	3,392.1	2,303.2	0.851	0.541	8.1E-25	1.5E-11
K demeth	KDM5B	7,773.7	4,439.9	8,035.3	6,011.6	2,906.4	0.808	0.419	1.1E-24	6.1E-08
K meth	ASH1L	6,641.9	3,247.3	5,539.5	2,899.1	3,993.7	1.036	0.934	7.5E-24	2.3E-27
K meth	KMT5B	3,500.9	2,412.2	2,885.9	2,477.1	2,600.2	0.536	0.220	1.2E-16	3.0E-03
K meth	KMT2D	7,272.7	4,065.7	6,393.4	4,556.7	6,288.7	0.839	0.489	2.0E-15	3.2E-06
PRC	CBX2	1,295.1	473.2	2,844.2	1,365.9	180.0	1.448	1.058	1.0E-13	2.4E-09
K meth	SETD6	911.3	602.8	889.4	662.5	448.7	0.596	0.425	9.7E-13	5.6E-07
DNA meth	TET3	2,199.7	1,379.0	4,603.2	3,046.6	1,486.0	0.675	0.595	7.9E-12	2.5E-09
DNA meth	DNMT3B	435.6	207.0	827.7	505.3	88.1	1.073	0.712	8.0E-12	1.1E-05
K meth	SETD5	6,065.4	4,280.1	5,645.4	4,728.0	4,195.0	0.503	0.256	9.1E-12	4.5E-04
PRC	ASXL1	4,446.9	3,265.4	4,486.6	3,621.0	4,054.7	0.443	0.309	4.8E-11	8.9E-05
K meth	EHMT1	3,235.4	2,301.4	3,920.8	2,840.5	2,550.2	0.490	0.465	9.1E-11	1.8E-10
K demeth	HR	68.6	173.7	3,183.4	2,542.4	163.7	−1.339	0.324	3.5E-10	1.9E-01
K demeth	PHF8	2,813.9	1,902.2	3,682.3	2,463.7	1,770.8	0.564	0.580	5.8E-10	1.1E-10
Acetyl	KAT7	2,988.2	2,218.8	3,049.2	2,307.9	2,316.7	0.430	0.402	1.9E-09	2.5E-09
K meth	SETD1B	2,504.0	1,777.3	2,818.1	2,036.4	2,043.2	0.493	0.469	1.9E-09	2.9E-09
Acetyl	KAT2A	4,042.1	2,634.1	4,291.0	3,146.7	1,159.1	0.615	0.447	7.4E-09	3.2E-06
K meth	KMT2B	4,319.8	2,877.6	5,682.8	3,593.2	2,547.8	0.586	0.661	1.0E-08	6.5E-12
R meth	PRMT5-AS1	25.6	14.1	19.5	18.5	8.5	0.871	0.072	2.0E-08	7.2E-01
K demeth	KDM6B	3,318.4	2,259.2	2,954.0	2,887.4	6,097.4	0.554	0.033	8.2E-08	8.5E-01
Acetyl	KAT6B	1,835.0	1,336.1	2,149.9	1,494.4	2,369.4	0.460	0.525	2.4E-06	3.9E-09
K meth	SETD1A	2,409.6	1,796.4	2,810.9	2,043.8	1,773.9	0.422	0.460	4.7E-06	1.1E-08
K demeth	KDM3B	5,409.5	4,359.2	4,703.8	3,374.5	5,018.0	0.309	0.479	8.7E-06	1.0E-14
PRC	CBX5	7,212.4	5,238.4	7,017.1	4,677.5	4,799.1	0.462	0.585	9.4E-06	1.2E-09
K demeth	KDM2B	1,648.9	1,338.5	2,501.5	1,884.1	1,661.2	0.300	0.409	1.2E-05	3.8E-09
Acetyl	KANSL1	3,698.6	2,872.8	3,717.8	2,536.9	2,862.3	0.366	0.551	1.9E-05	2.3E-14
K meth	EZH1	1,829.4	1,472.2	1,657.1	1,194.5	2,514.6	0.312	0.472	5.4E-05	1.0E-09
PRC	CBX1	4,579.3	3,670.7	6,577.2	4,517.0	2,268.0	0.317	0.542	9.0E–04	1.4E-10
Ubiquit	RNF20	2,404.6	2,146.8	3,226.9	2,432.3	2,721.8	0.161	0.408	2.2E-02	4.7E-10
Ubiquit	RNF38	2,095.4	1,907.5	2,493.1	1,749.1	3,353.6	0.136	0.511	1.6E-01	4.3E-09

*Log_2 scale. Meth and demeth, methylation and demethylation; Acetyl and Deacetyl, acetylation and deacetylation; Ubiquit, ubiquitination; R meth, arginine methylation; PRC, Polycomb repressive complex. SH and SL, SETDB1-high and SETDB1-low, respectively.

has been in the spotlight as the key regulator for suppressing innate immunity by limiting the overall abundance of TE transcripts in cancer cells (Guler et al., 2017). Genomic TEs, when derepressed, produce double-stranded RNAs or reverse-transcribed DNAs into the cytoplasm; these atypical nucleic acids should be eliminated by metabolism, and if not removed, they cause an inappropriate activation of nucleic acid-sensing pathways, ultimately inducing the IFNα response and cell death (Zeng et al., 2014) or autoimmunity by their chronic presence (Volkman and Stetson, 2014). The role of nuclear SETDB1 as the primary effector in the repression of TEs, particularly endogenous retrovirus (ERV) and LINE1 copies (Matsui et al., 2010; Liu et al., 2014), can be broadened to its cytoplasmic function for surveillance of these spin-offs of TE transcripts. SETDB1 may indirectly recognize TE-derived nucleic acids and associate with the Argonaute (AGO) protein (Cho et al., 2014) at the start of the RNAi process before

targeting the genomic TEs of interest in the nucleus to re-repress their transcription and eventually block the IFNα response (Kang, 2018). The benefits of this feedback regulation scenario are that SETDB1 is a well-known TE silencer and, by virtue of its operating mechanism, can efficiently scan the chromosomes for complementary sequences of any transcribed TE-derived sequences. The latter is plausible because SETDB1 can simultaneously access and quickly leaf through a large fraction of the genome through the association of promyelocytic leukemia (PML)-nuclear bodies, as suggested (Cho et al., 2011; Kang, 2015). Consistently, an increase in the number and size of PML-NBs was observed after viral infection and IFN treatment (Regad and Chelbi-Alix, 2001). Additionally, double-stranded RNAs of the TE origin have been demonstrated to induce the production of endogenous small interfering RNAs (siRNAs) and subsequently silence TEs by an RNAi-dependent mechanism in human cells (Yang and Kazazian, 2006). This
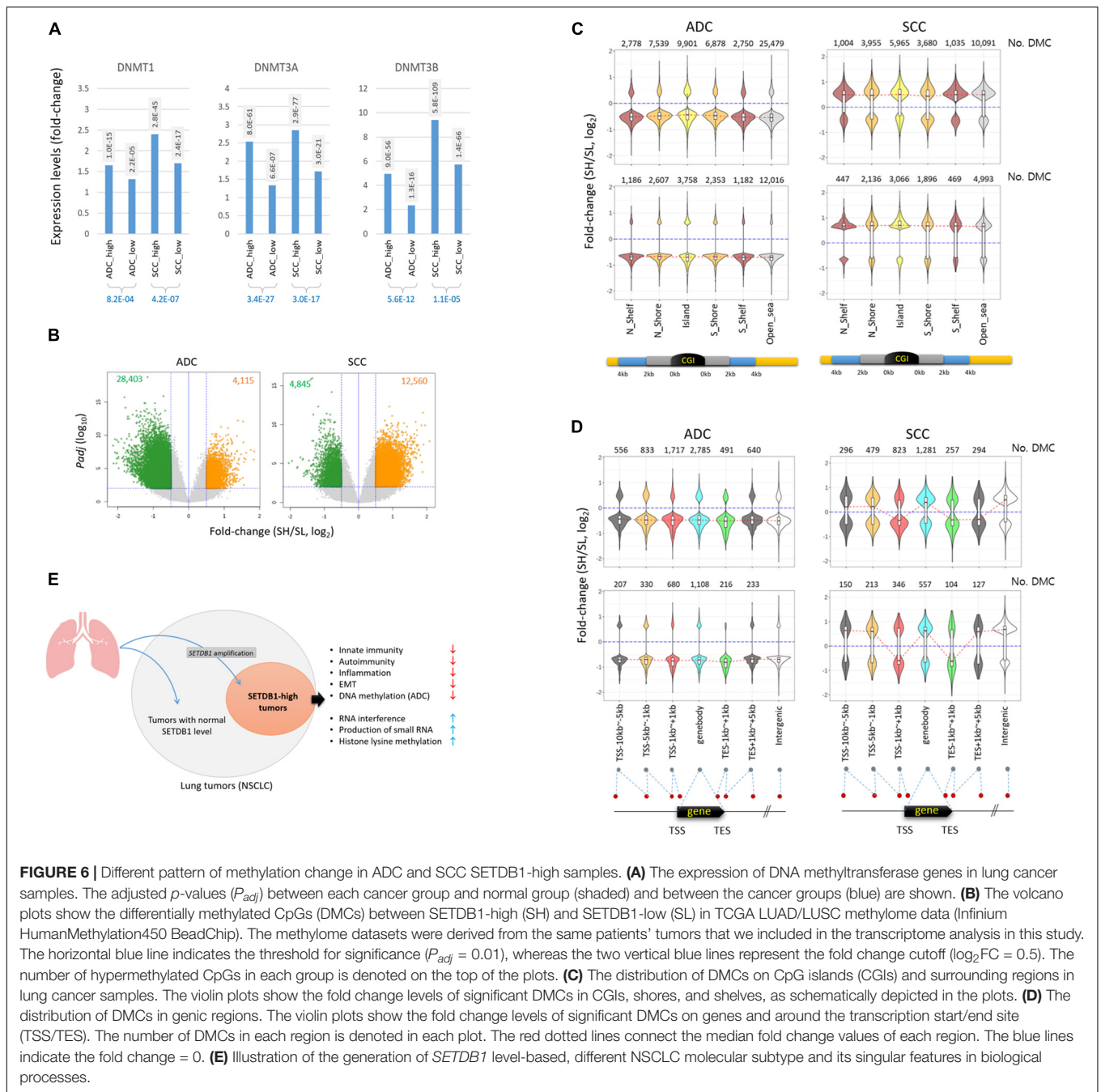
FIGURE 6 | Different pattern of methylation change in ADC and SCC SETDB1-high samples. **(A)** The expression of DNA methyltransferase genes in lung cancer samples. The adjusted p-values (P_{adj}) between each cancer group and normal group (shaded) and between the cancer groups (blue) are shown. **(B)** The volcano plots show the differentially methylated CpGs (DMCs) between SETDB1-high (SH) and SETDB1-low (SL) in TCGA LUAD/LUSC methylome data (Infinium HumanMethylation450 BeadChip). The methylome datasets were derived from the same patients' tumors that we included in the transcriptome analysis in this study. The horizontal blue line indicates the threshold for significance ($P_{adj} = 0.01$), whereas the two vertical blue lines represent the fold change cutoff ($\log_2 FC = 0.5$). The number of hypermethylated CpGs in each group is denoted on the top of the plots. **(C)** The distribution of DMCs on CpG islands (CGIs) and surrounding regions in lung cancer samples. The violin plots show the fold change levels of significant DMCs in CGIs, shores, and shelves, as schematically depicted in the plots. **(D)** The distribution of DMCs in genic regions. The violin plots show the fold change levels of significant DMCs on genes and around the transcription start/end site (TSS/TES). The number of DMCs in each region is denoted in each plot. The red dotted lines connect the median fold change values of each region. The blue lines indicate the fold change = 0. **(E)** Illustration of the generation of *SETDB1* level-based, different NSCLC molecular subtype and its singular features in biological processes.

scenario well harmonizes with our GSEA result, which is as follows: the depletion of genes belonging to the host innate immune response and several autoimmunity-related collections, and the enrichment of genes associated with the RNA*i* and small RNA production in the SETDB1-overrepresented samples (**Figures 2**, **3**). Therefore, results of our meta-analysis of lung SH tumor transcriptomes provided several clues that support the generality of SETDB1 functions for controlling innate immunity and presumably autoimmunity through the surveillance of TE expressions that have been previously proposed in different types of cancer cells or cultured cells.

SETDB1 transcriptionally represses TEs as an effector in the KAP1-built repressive complex (Matsui et al., 2010); therefore, it stands to reason to give the first consideration to these TEs for their expression change in the SH tumor samples. Unfortunately, however, the TE expression data were unavailable in the RNA-seq datasets analyzed in this study; thus, we were unable to investigate the extent to which the genomic TE expression would be affected and altered by SETDB1 overexpression in lung cancers. If the TE data are available, it would be interesting to examine whether the depletion of genes related to innate immunity and autoimmunity-associated collections in the SH samples

(**Figure 3D**) correlates with the transcriptional repression of TEs by SETDB1-involved mechanism and whether the reduced expression of nucleic acid sensors (**Figure 3E**) may be intertwined in this regulatory circuit.

The lack of DNA methylation data on genomic TEs and other repeated sequences is also a limitation. Compared with the SL samples, the SH samples expressed the three DNA methyltransferases *DNMT1, DNMT3A,* and *DNMT3B* at higher levels, but we were unable to completely understand the impact of DNMT overexpression in lung tumors because the methylome datasets used in this study were derived from the Infinium 450 K array platform, which can identify methylation changes with a single CpG resolution at unique genomic loci (Bibikova et al., 2011) but, in exchange for advantage, may be inadequate in scrutinizing entire genomic repeats and unique sequences in the intergenic regions (Min et al., 2019). It might be important to assess methylation changes over the genomic repeats in lung SH tumors, particularly when the targets of DNMT3A and DNMT3B are known to be pericentric heterochromatin (Okano et al., 1999; Xu et al., 1999; Bachman et al., 2001), intracisternal A particles (IAP) (Walsh et al., 1998; Kaneda et al., 2004; Kato et al., 2007), and many other repeated sequences. Genome-wide DNA methylation studies on lung cancer have mostly focused on the genic promoter regions for cancer-specific biomarkers (Kwon et al., 2012; Bjaanaes et al., 2016). In a previous report (Rauch et al., 2008), the combined bisulfite restriction analysis (COBRA) method was used, in which SCC samples exhibited hypomethylation at genomic repeats, including LINE1s and ERVs. This PCR-based method is still of limited use for studying the entire genome. Therefore, to comprehensively understand the lung cancer methylome and the implication of SETDB1 overexpression, we should first determine the methylation change occurring in the genic loci and in the various genomic repeats that occupy almost half of the genome.

The DMCs found in the ADCs were mostly undermethylated in the SH samples compared with the SL samples, whereas in SCCs, it was the opposite (**Figure 6B**). The antithetical pattern of change in DNA methylation was interesting because both ADCs and SCCs commonly overrepresented the *DNMT* gene transcripts in the SH samples (**Figure 6A**). This suggests that although SETDB1 overexpression is related to DNMT overexpression, these proteins may be uncoupled with genomic DNA methylation in the ADCs. The reason for the inverse relationship between DNMT expression and DNA methylation level in the ADCs should be addressed, which, we hope, will be answered when a genome-wide methylation analysis, such as MBD sequencing or whole-genome bisulfite sequencing of lung SH tumor samples, is performed. We hypothesize that abundant transcriptional activators that prevail in the cancer cells preoccupy the must-be-methylated regions in collaboration with other chromatin relaxing epidrivers (such as HATs) and, thus, ward DNMTs off the spots, causing the regions to be passively demethylated. Similarly, the increased expression of *DNMT* genes may be a feedback response to set the dwindled, irrelevant genomic methylation right. In this sense, CGIs, which are frequently hypermethylated in lung (Rauch et al., 2008; Carvalho et al., 2012) and other human cancers,

including particularly those associated with tumor suppression and other genome defense pathways (Esteller, 2002), may be important battlegrounds where both transcriptional activators and repressors compete for occupancy. Therefore, it is interesting to notice that the DMCs around the CGIs were undermethylated in the ADC-SH samples, whereas those in the SCC-SH samples were mostly relatively highly methylated, although they appeared to comply with the overall methylation changes of their own types, as seen in the methylation changes in the "Open_sea" regions (**Figure 6C**).

SETDB1 is increasingly attracting researchers' attention for brain disorders including Huntington's disease (Ryu et al., 2006), schizophrenia (Chase et al., 2013), and autism (Cukier et al., 2012), in addition to cancers. If the questions on the action mechanisms and targets of SETDB1 in diseased cells can be solved, it will be possible to build a therapeutic plan for the targets and at multiple levels in the SETDB1 pathogenic pathway.

DATA AVAILABILITY STATEMENT

Gene expression datasets for TCGA LUAD (https://portal.gdc.cancer.gov/projects/TCGA-LUAD) and lung squamou-cell carcinoma (https://portal.gdc.cancer.gov/projects/TCGA-LUSC) were downloaded from Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/), and copy number alteration data were obtained from cBioPortal (https://www.cbioportal.org/).

AUTHOR CONTRIBUTIONS

Y-KK: conception, design, and administrative support. BM: collection and assembly of data. Y-KK and BM: data analysis and interpretation. Both authors wrote and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.573515/full#supplementary-material

REFERENCES

Bachman, K. E., Rountree, M. R., and Baylin, S. B. (2001). Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem.* 276, 32282–32287. doi: 10.1074/jbc.m104661200

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462, 108–112. doi: 10.1038/nature08460

Bellomo, C., Caja, L., and Moustakas, A. (2016). Transforming growth factor beta as regulator of cancer stemness and metastasis. *Br. J. Cancer* 115, 761–769. doi: 10.1038/bjc.2016.255

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., et al. (2011). High density DNA methylation array with single CpG site resolution. *Genomics* 98, 288–295. doi: 10.1016/j.ygeno.2011.07.007

Bjaanaes, M. M., Fleischer, T., Halvorsen, A. R., Daunay, A., Busato, F., Solberg, S., et al. (2016). Genome-wide DNA methylation analyses in lung adenocarcinomas: Association with EGFR. KRAS and TP53 mutation status, gene expression and prognosis. *Mol. Oncol.* 10, 330–343. doi: 10.1016/j.molonc.2015.10.021

Burrell, R. A., McGranahan, N., Bartek, J., and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501, 338–345. doi: 10.1038/nature12625

Carvalho, R. H., Haberle, V., Hou, J., van Gent, T., Thongjuea, S., van Ijcken, W., et al. (2012). Genome-wide DNA methylation profiling of non-small cell lung carcinomas. *Epigenet. Chromatin.* 5:9. doi: 10.1186/1756-8935-5-9

Ceol, C. J., Houvras, Y., Jane-Valbuena, J., Bilodeau, S., Orlando, D. A., Battisti, V., et al. (2011). The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* 471, 513–517. doi: 10.1038/nature09806

Chang, Q., Bournazou, E., Sansone, P., Berishaj, M., Gao, S. P., Daly, L., et al. (2013). The IL-6/JAK/Stat3 feed-forward loop drives tumorigenesis and metastasis. *Neoplasia* 15, 848–862. doi: 10.1593/neo.13706

Chase, K. A., Gavin, D. P., Guidotti, A., and Sharma, R. P. (2013). Histone methylation at H3K9: evidence for a restrictive epigenome in schizophrenia. *Schizophr. Res.* 149, 15–20. doi: 10.1016/j.schres.2013.06.021

Cho, S., Park, J. S., and Kang, Y. K. (2011). Dual functions of histone-lysine N-methyltransferase Setdb1 protein at promyelocytic leukemia-nuclear body (PML-NB): maintaining PML-NB structure and regulating the expression of its associated genes. *J. Biol. Chem.* 286, 41115–41124. doi: 10.1074/jbc.m111.248534

Cho, S., Park, J. S., and Kang, Y. K. (2014). AGO2 and SETDB1 cooperate in promoter-targeted transcriptional silencing of the androgen receptor gene. *Nucleic Acids Res.* 42, 13545–13556. doi: 10.1093/nar/gku788

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 44:e71. doi: 10.1093/nar/gkv1507

Cuellar, T. L., Herzner, A. M., Zhang, X., Goyal, Y., Watanabe, C., Friedman, B. A., et al. (2017). Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia. *J. Cell Biol.* 216, 3535–3549. doi: 10.1083/jcb.201612160

Cukier, H. N., Lee, J. M., Ma, D., Young, J. I., Mayo, V., Butler, B. L., et al. (2012). The expanding role of MBD genes in autism: identification of a MECP2 duplication and novel alterations in MBD5. MBD6, and SETDB1. *Autism Res.* 5, 385–397. doi: 10.1002/aur.1251

Esteller, M. (2002). CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* 21, 5427–5440. doi: 10.1038/sj.onc.1205600

Esteller, M. (2006). Epigenetics provides a new generation of oncogenes and tumour-suppressor genes. *Br. J. Cancer* 94, 179–183. doi: 10.1038/sj.bjc.6602918

Fei, Q., Shang, K., Zhang, J., Chuai, S., Kong, D., Zhou, T., et al. (2015). Histone methyltransferase SETDB1 regulates liver cancer cell growth through methylation of p53. *Nat. Commun.* 6:8651.

Guler, G. D., Tindell, C. A., Pitti, R., Wilson, C., Nichols, K., KaiWai Cheung, T., et al. (2017). Repression of Stress-Induced LINE-1 Expression Protects Cancer Cell Subpopulations from Lethal Drug Exposure. *Cancer Cell* 32, 221–237e13.

Hanzelmann, S., Castelo, R., and Guinney, J. G. S. V. A. (2013). Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* 14:7. doi: 10.1186/1471-2105-14-7

Herbst, R. S., Morgensztern, D., and Boshoff, C. (2018). The biology and management of non-small cell lung cancer. *Nature* 553, 446–454.

Inamura, K. (2017). Lung Cancer: Understanding Its Molecular Pathology and the 2015 WHO Classification. *Front. Oncol.* 7:193. doi: 10.3389/fonc.2017.00193

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., and Thun, M. J. (2009). Cancer statistics, 2009. *CA Cancer J. Clin.* 59, 225–249. doi: 10.3322/caac.20006

Johnson, D. E., O'Keefe, R. A., and Grandis, J. R. (2018). Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nat. Rev. Clin. Oncol.* 15, 234–248. doi: 10.1038/nrclinonc.2018.8

Kan, Z., Jaiswal, B. S., Stinson, J., Janakiraman, V., Bhatt, D., Stern, H. M., et al. (2010). Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 466, 869–873.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., et al. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429, 900–903. doi: 10.1038/nature02633

Kang, Y. K. (2015). SETDB1 in Early Embryos and Embryonic Stem Cells. *Curr.Issues Mole. Biol.* 17, 1–10. doi: 10.1155/2018/6239245

Kang, Y. K. (2018). Surveillance of Retroelement Expression and Nucleic-Acid Immunity by Histone Methyltransferase SETDB1. *Bioessays* 2018:e1800058.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., et al. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.* 16, 2272–2280. doi: 10.1093/hmg/ddm179

Korotkevich, G., Sukhov, V., and Sergushichev, A. (2019). Fast gene set enrichment analysis. *bioRxiv[Preprint]*.doi: 10.1101/060012

Kwon, S., Park, J. S., Min, B., and Kang, Y. K. (2015). Differences in the gene expression profiles of slow- and fast-forming preinduced pluripotent stem cell colonies. *Stem Cells Int.* 2015:195928.

Kwon, Y. J., Lee, S. J., Koh, J. S., Kim, S. H., Lee, H. W., Kang, M. C., et al. (2012). Genome-wide analysis of DNA methylation and the gene expression change in lung cancer. *J. Thorac. Oncol.* 7, 20–33. doi: 10.1097/jto.0b013e3182307f62

Lafuente-Sanchis, A., Zuniga, A., Galbis, J. M., Cremades, A., Estors, M., Martinez-Hernandez, N. J., et al. (2016). Prognostic value of ERCC1, RRM1, BRCA1 and SETDB1 in early stage of non-small cell lung cancer. *Clin. Transl. Oncol.* 18, 798–804. doi: 10.1007/s12094-015-1440-6

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425. doi: 10.1016/j.cels.2015.12.004

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740. doi: 10.1093/bioinformatics/btr260

Lin, E. W., Karakasheva, T. A., Lee, D. J., Lee, J. S., Long, Q., Bass, A. J., et al. (2017). Comparative transcriptomes of adenocarcinomas and squamous cell carcinomas reveal molecular similarities that span classical anatomic boundaries. *PLoS Genet.* 13:e1006938. doi: 10.1371/journal.pgen.1006938

Liu, S., Brind'Amour, J., Karimi, M. M., Shirane, K., Bogutz, A., Lefebvre, L., et al. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev.* 28, 2041–2055. doi: 10.1101/gad.244848.114

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., et al. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* 464, 927–931. doi: 10.1038/nature08858

Min, B., Jeon, K., Park, J. S., and Kang, Y. K. (2019). Demethylation and derepression of genomic retroelements in the skeletal muscles of aged mice. *Aging Cell* 18:e13042.

Min, B., Park, M., Jeon, K., Park, J. S., Seo, H., Jeong, S., et al. (2018). Age-associated bimodal transcriptional drift reduces intergenic disparities in transcription. *Aging* 10, 789–807. doi: 10.18632/aging.101428

Mounir, M., Lucchetta, M., Silva, T. C., Olsen, C., Bontempi, G., Chen, X., et al. (2019). New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLoS Comput. Biol.* 15:e1006701. doi: 10.1371/journal.pcbi.1006701

Noguchi, M., Morikawa, A., Kawasaki, M., Matsuno, Y., Yamada, T., Hirohashi, S., et al. (1995). Small adenocarcinoma of the lung. *Histolog. Char. Prog. Cancer* 75, 2844–2852.

Okano, M., Bell, D. W., Haber, D. A., and Li, E. D. N. A. (1999). methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257. doi: 10.1016/s0092-8674(00)81656-6

Park, M., Min, B., Jeon, K., Cho, S., Park, J. S., Kim, J., et al. (2017). Age-associated chromatin relaxation is enhanced in Huntington's disease mice. *Aging* 9, 803–822. doi: 10.18632/aging.101193

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rauch, T. A., Zhong, X., Wu, X., Wang, M., Kernstine, K. H., Wang, Z., et al. (2008). High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc. Natl. Acad. Sci. USA* 105, 252–257. doi: 10.1073/pnas.0710735105

Regad, T., and Chelbi-Alix, M. K. (2001). Role and fate of PML nuclear bodies in response to interferon and viral infections. *Oncogene* 20, 7274–7286. doi: 10.1038/sj.onc.1204854

Regina, C., Compagnone, M., Peschiaroli, A., Lena, A., Annicchiarico-Petruzzelli, M., Piro, M. C., et al. (2016). Setdb1, a novel interactor of DeltaNp63, is involved in breast tumorigenesis. *Oncotarget* 7, 28836–28848. doi: 10.18632/oncotarget.7089

Rich, J. N. (2016). Cancer stem cells: understanding tumor hierarchy and heterogeneity. *Medicine* 95(1 Suppl. 1), S2–S7.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007

Rodriguez-Paredes, M., and Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. *Nat. Med.* 17, 330–339. doi: 10.1038/nm.2305

Rodriguez-Paredes, M., Martinez de Paz, A., Simo-Riudalbas, L., Sayols, S., Moutinho, C., and Moran, S. (2014). Gene amplification of the histone methyltransferase SETDB1 contributes to human lung tumorigenesis. *Oncogene* 33, 2807–2813. doi: 10.1038/onc.2013.239

Ryu, H., Lee, J., Hagerty, S. W., Soh, B. Y., McAlpin, S. E., Cormier, K. A., et al. (2006). ESET/SETDB1 gene expression and histone H3 (K9) trimethylation in Huntington's disease. *Proc. Natl. Acad. Sci. USA* 103, 19176–19181. doi: 10.1073/pnas.0606373103

Ryu, T. Y., Kim, K., Kim, S. K., Oh, J. H., Min, J. K., Jung, C. R., et al. (2019). SETDB1 regulates SMAD7 expression for breast cancer metastasis. *BMB Rep.* 52, 139–144. doi: 10.5483/bmbrep.2019.52.2.235

Sato, M., Larsen, J. E., Lee, W., Sun, H., Shames, D. S., Dalvi, M. P., et al. (2013). Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations. *Mol Cancer Res.* 11, 638–650. doi: 10.1158/1541-7786.mcr-12-0634-t

Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G., and Rauscher, F. J. III (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes. Dev.* 16, 919–932. doi: 10.1101/gad.973302

Sher, T., Dy, G. K., and Adjei, A. A. (2008). Small cell lung cancer. *Mayo Clin. Proc.* 83, 355–367.

Sun, S., Schiller, J. H., and Gazdar, A. F. (2007). Lung cancer in never smokers–a different disease. *Nat. Rev. Cancer* 7, 778–790. doi: 10.1038/nrc2190

Swanton, C. (2012). Intratumor heterogeneity: evolution through space and time. *Cancer Res.* 72, 4875–4882. doi: 10.1158/0008-5472.can-12-2217

Volkman, H. E., and Stetson, D. B. (2014). The enemy within: endogenous retroelements and autoimmune disease. *Nat. Immunol.* 15, 415–422.

Walsh, C. P., Chaillet, J. R., and Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* 20, 116–117. doi: 10.1038/2413

Weigelt, B., Mackay, A., A'Hern, R., Natrajan, R., Tan, D. S., Dowsett, M., et al. (2010). Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol.* 11, 339–349. doi: 10.1016/s1470-2045(10)70008-5

Welch, D. R., and Hurst, D. R. (2019). Defining the Hallmarks of Metastasis. *Cancer Res.* 79, 3011–3027. doi: 10.1158/0008-5472.can-19-0458

Wong, C. M., Wei, L., Law, C. T., Ho, D. W., Tsang, F. H., Au, S. L., et al. (2016). Up-regulation of histone methyltransferase SETDB1 by multiple mechanisms in hepatocellular carcinoma promotes cancer metastasis. *Hepatology* 63, 474–487. doi: 10.1002/hep.28304

Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., et al. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 402, 187–191. doi: 10.1038/46052

Yang, N., and Kazazian, H. H. Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.* 13, 763–771. doi: 10.1038/nsmb1141

Zeng, M., Hu, Z., Shi, X., Li, X., Zhan, X., Li, X. D., et al. (2014). MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses. *Science* 346, 1486–1492. doi: 10.1126/science.346.6216.1486

Zhang, J., Fujimoto, J., Zhang, J., Wedge, D. C., Song, X., Zhang, J., et al. (2014). Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 346, 256–259.