Elevated NSD3 Histone Methylation Activity Drives Squamous Cell Lung Cancer

Gang Yuan[1,11], **Natasha M. Flores**[2,11], **Simone Hausmann**[2], **Shane M. Lofgren**[2], **Vladlena Kharchenko**[3], **Maria Angulo-Ibanez**[4,5], **Deepanwita Sengupta**[1], **Xiaoyin Lu**[2], **Iwona Czaban**[3], **Dulat Azhibek**[3], **Silvestre Vicent**[6], **Wolfgang Fischle**[3], **Mariusz Jaremko**[3], **Bingliang Fang**[7], **Ignacio I. Wistuba**[8], **Katrin F. Chua**[4,5], **Jack A. Roth**[7], **John D. Minna**[9], **Ning-Yi Shao**[10,#], **Łukasz Jaremko**[3,#], **Pawel K. Mazur**[2,#], **Or Gozani**[1,#]

[1]Department of Biology, Stanford University, Stanford, CA 94305, USA

[2]Department of Experimental Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA

[3]Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

[4]Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

[5]Geriatric Research, Education, and Clinical Center, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA 94304, USA

[6]University of Navarra, Center for Applied Medical Research, Pamplona, 31008, Spain

[7]Department of Thoracic and Cardiovascular Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA

[8]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA

[9]Hamon Center for Therapeutic Oncology Research and Departments of Internal Medicine and Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA

[10]Faculty of Health Sciences, University of Macau, Macau SAR, China

[11]These authors contributed equally to the work

[#]To whom correspondence should be addressed: ogozani@stanford.edu; pkmazur@mdanderson.org; lukasz.jaremko@kaust.edu.sa; nshao@um.edu.mo.

Abstract

Amplification of chromosomal region 8p11-12 is a frequent genetic alteration implicated in the etiology of lung squamous cell carcinoma (LUSC)[1-3]. FGFR1 (fibroblast growth factor receptor 1) is the main candidate driver within this region[4]. However, clinical trials evaluating FGFR1 inhibition as a targeted therapy have been unsuccessful[5]. Here we identify the H3K36 methyltransferase NSD3 (nuclear receptor binding SET domain protein 3), an 8p11-12-localized gene, as a key regulator of LUSC tumorigenesis. In contrast to other 8p11-12 candidate LUSC drivers, increased *NSD3* expression strongly correlates with its gene amplification. Ablation of *NSD3*, but not *FGFR1*, attenuates tumor growth and extends survival in a potent LUSC mouse model. We identify $NSD3_{T1232A}$ as an LUSC-associated variant that increases H3K36 dimethylation (H3K36me2) catalytic activity *in vitro* and *in vivo*. Structural dynamic analyses reveal that the T1232A substitution elicits localized mobility changes throughout NSD3's catalytic domain to relieve auto-inhibition and increase H3 substrate accessibility. $NSD3_{T1232A}$ expression *in vivo* accelerates tumorigenesis and decreases overall survival in LUSC mouse models. Pathologic H3K36me2 generation by $NSD3_{T1232A}$ rewires the chromatin landscape to promote oncogenic gene expression programming. Further, NSD3's catalytic activity promotes human tracheobronchial cell transformation and xenograft growth of human 8p11-12-amplified LUSC cell lines. NSD3 depletion in patient-derived xenografts (PDXs) from primary LUSC harboring *NSD3* amplification or the $NSD3_{T1232A}$ variant attenuates neoplastic growth. Finally, NSD3-regulated LUSC PDXs are markedly sensitive to bromodomain inhibition (BETi). Together, our work identifies NSD3 as a principal 8p11-12 amplicon-associated oncogenic driver in LUSC and suggests that NSD3-dependency renders LUSC therapeutically vulnerable to BETi.

NSD3, not FGFR1, promotes LUSC in vivo

Many genes present within the 8p11-12 amplicon could potentially contribute to oncogenesis (Extended Data Fig. 1a). However, the minimal region of amplification across many 8p11-12-signature neoplasms centers around *FGFR1* and the neighboring gene and candidate oncogene *NSD3*[3,4,6,7]. Indeed, analysis of LUSC datasets from The Cancer Genome Atlas (TCGA) shows that *NSD3* and the immediate neighboring genes (e.g. *FGFR1*) are the most commonly amplified 8p11-12 genes (Extended Data Fig. 1a). Overall, amplification of the genomic area spanning the *NSD3* and *FGFR1* genes is one of the more common molecular alterations in LUSC (Extended Data Fig. 1b). Notably, for NSD3, gene amplification strongly correlates with increased mRNA expression; in contrast, there is little correlation between FGFR1 gene copy number and mRNA expression (Fig. 1a; Extended Data Figs. 1a-b)[2,3]. Accordingly, sgRNA-mediated depletion of *NSD3*, but not depletion of *FGFR1* or the gene immediately adjacent to *NSD3* (*PLPP5*), in the 8p11-12 amplified (8p11^AMP) H520 LUSC cell line inhibited xenograft tumor growth (Fig. 1b; Extended Data Figs. 2a-e). These data point to NSD3, rather than FGFR1, as the mutational driver within the 8p11-12 amplicon in LUSC.

We next established a robust LUSC mouse model consisting of canonical LUSC alterations co-occurring with 8p11-12 amplification (constitutively active *P*I3K, *S*OX2 overexpression, and *CDKN2A/B* deletion, hereto named *PSC* mice; (see Extended Data Fig. 1b and Methods). *PSC* mice develop with high penetrance lung tumors characterized by LUSC-

defining molecular hallmarks (Extended Data Fig. 2f)[8]. In this model, increased NSD3 expression tracks with tumor progression, consistent with NSD3 overexpression observed in ~60% of LUSC samples (beyond the 20% harboring 8p11^{AMP}) and frequently co-occurring with *PSC* molecular alterations (Extended Data Figs. 2g-i).

Conditional *FGFR1$^{loxP/loxP}$* and *NSD3$^{loxP/loxP}$* mutant mice were generated and are viable, fertile and develop normally (see Methods). In the *PSC* background, targeted deletion in the lung of NSD3 (*PSC$^{NSD3-KO}$*) (Fig. 1c), but not FGFR1 (*PSC$^{FGFR1-KO}$*) (Fig. 1d), significantly attenuated tumor growth and cancer cell proliferation, while increasing apoptosis (Fig. 1e; Extended Data Figs. 2j-l). In survival studies, deletion of *NSD3* extended lifespan of *PSC* mice by ~30%, whereas *FGFR1* knockout had no impact (Fig. 1f). Together, these data support a key *in vivo* role for NSD3 in LUSC tumorigenesis.

NSD3$_{T1232A}$ is a hyperactive variant

NSD3, along with NSD1, NSD2, and ASH1L, comprise the four enzymes in mammals that specifically synthesize the euchromatin-associated H3K36me2 modification (reviewed in [9]). Depending on cell type, NSD1 and NSD2 are the major H3K36me2-generating enzymes, whereas the physiologic context in which NSD3 regulates H3K36me2 is less clear[9]. We observed lower global H3K36me2 levels in NSD3-deleted lung tumor tissue compared to control tumors (Fig. 1c), suggesting an etiological role for NSD3-catalyzed H3K36me2 synthesis in *NSD3* amplified-LUSC. We reasoned this idea could be tested in mice by transgenic expression of an NSD3 hyperactive variant, which would model the consequence of increased NSD3 catalysis due to elevated *NSD3* dosage on tumorigenesis. A recurrent hyperactivating NSD2 mutation is present in different cancers[9], suggesting that a functionally analogous cancer-associated mutation may be present in NSD3. To this end, we tested *in vitro* histone methylation activity for 35 different TCGA-documented mutations within the NSD3 catalytic domain (Extended Data Figs. 3a-c). The variant showing the highest activity was the T1232A substitution (NSD3$_{T1232A}$), a recurrent cancer-associated mutation[10].

The *in vitro* methylation activity on nucleosomes of recombinant NSD3 catalytic domain harboring the T1232A substitution (NSD3$_{SET-T1232A}$), which is stronger than wild-type (NSD3$_{SET}$) enzyme, is abrogated when paired with a catalytic mutant (NSD3$_{SET-T1232A/Y1174A}$) (Figs. 2a-b; Extended Data Fig. 3d). Further, NSD3$_{SET-T1232A}$ was more efficient in selectively generating H3K36me2 relative to wild-type enzyme (Fig. 2b). Despite the higher activity, NSD3$_{SET-T1232A}$ behaved similar to NSD3$_{SET}$ with respect to (i) nucleosomal versus free H3 substrate preference (Fig. 2a), (ii) methylation state specificity, generating H3K36me1/2 but not H3K36me3 (Fig. 2b), and (iii) site-specific selectivity, methylating K36, but not several other histone lysine residues (Fig. 2b). Recombinant full-length NSD3$_{T1232A}$ behaves like the isolated catalytic domain in being hyperactive while retaining the same catalytic selectivity profile (Fig. 2c). Finally, overexpression of NSD3$_{T1232A}$, but not NSD3$_{WT}$ or NSD3$_{T1232A/Y1174A}$, increases H3K36me2 levels in NSD2-depleted HT1080 fibrosarcoma cells (Extended Data Fig. 3e). These findings identify NSD3$_{T1232A}$ as a hyperactive cancer-associated mutant.

The enhanced activity of $NSD3_{SET-T1232A}$ versus $NSD3_{SET}$ is more evident when using nucleosomes with longer linker DNA (187 versus 147 base-pairs) (Extended Data Fig. 3f). The mechanism behind the increased catalysis, irrespective of linker DNA, is not mediated by the T1232A substitution significantly altering NSD3's binding affinity for nucleosomes or the cofactor methyl donor S-adenosyl-L-methionine (SAM) (Extended Data Fig. 4a; Supplementary Table 1). High-resolution nuclear magnetic resonance (NMR) comparison of the protein backbone of $NSD3_{SET-T1232A}$ and $NSD3_{SET}$ in solution shows that the static structure is largely unaffected by the T1232A substitution (Extended Data Figs. 4b-c; Supplementary Table 1). In contrast, chemical shift perturbations (CSPs) of the backbone amide $^1H/^{15}N$ resonances, which serve as sensitive probes for local structural and allosteric effects, were observed throughout the SET domain (Fig. 2d; Extended Data Figs. 4c-d). The highest CSPs are centered around T1232A, with effects extending throughout the substrate-binding cleft and include the conserved NSD3 auto-regulatory loop (Fig. 2d; Extended Data Figs. 4d-e)[11,12]. Consistent with these findings, backbone ^{15}N spin relaxation studies revealed multiple differences in local dynamics between $NSD3_{SET-T1232A}$ and $NSD3_{SET}$ concentrated at the substrate-binding cleft and auto-regulatory loop regions (Fig. 2e; Extended Data Figs. 4g-h; Supplementary Table 1). These data suggest that T1232A increases activity by (1) enhancing the mobility of the auto-regulatory loop, which destabilizes the inhibitory state, acting in concert with (2) the active site being rendered more accessible to the H3 substrate and (3) direct, localized modification of the H3 binding surface. Indeed, the polar hydroxyl group loss with T1232A increases local hydrophobicity, which can alter H3 anchoring dynamics (Figs. 2d-e)[13]. Introduction of a second alanine substitution at V1243, a residue adjacent to A1232 in the structure (T1232A/V1243A), is predicted to restore local surface hydrophobicity[13]. Accordingly, methylation activity of a double alanine mutant ($NSD3_{SET-T1232A/V1243A}$) is similar to wild-type enzyme and lower than activity of the single mutant ($NSD3_{SET-T1232A}$), while the $NSD3_{SET-V1243A}$ shows no activity and decreased thermal stability (Extended Data Fig. 4i; Supplementary Table 1). Thus, T1232A substitution induces local dynamics changes at many residues within three key functional regions of the NSD3 catalytic domain that likely work in concert to increase H3K36 methylation catalysis.

H3K36me2 deregulation accelerates LUSC

$NSD3_{T1232A}$ was next used to model consequences of elevated catalytic activity due to *NSD3* amplification on LUSC tumorigenesis *in vivo*. Murine and human NSD3 are highly similar (94% overall; 98% in the catalytic domain) with murine NSD3 (mNSD3) T1242 corresponding to human T1232 (Extended Data Fig. 5a). We generated Cre-inducible V5-tagged $mNSD3_{T1242A}$ mice and transgene expression was verified in lung fibroblast cells transduced with Ad-Cre *ex vivo* (Extended Data Figs. 5b-e). In the *PSC* background, expression of $mNSD3_{T1242A}$ (hereto named *PSCN*) accelerated tumor growth (Figs. 3a-b) and proliferation (Figs. 3c-d) while decreasing apoptosis (Extended Data Figs. 5f-g) relative to the control *PSC* group. For example, at 90 days after Cre-induction, when *PSC* mice have no detectable tumors, multiple growths are observed in the *PSCN* cohort (Fig. 3a). At 120 days, tumor burden in *PSCN* mice is over an order of magnitude greater than in control *PSC* mice (Figs. 3a-b). Consistent with these data, expression of $mNSD3_{T1242A}$ reduced *PSC*

mice lifespan by ~30% (Fig. 3e). In independent LUSC tumor biopsies, high NSD3 levels (due to $mNSD3_{T1242A}$ expression) result in markedly elevated H3K36me2 (Fig. 3f; Extended Data Fig. 5f). Moreover, $mNSD3_{T1242A}$-driven tumors have higher levels of oncoproteins MYC and BRD4 and phosphorylated 4EBP1 (a proxy for mTOR pathway activation) (Fig. 3f). Together, these data indicate that NSD3's activity significantly impacts the clinical course and outcome of LUSC *in vivo*.

Comparison of cell lines derived from $mNSD3_{T1242A}$ and control *PSC* matched tumors (hereto named PSC_N and PSC_C, respectively) show that H3K36me2 levels in cells, like in tumors, are higher if expressing the $mNSD3_{T1242A}$ transgene (Extended Data Fig. 5h). Further, NSD3 depletion in PSC_N cells results in depletion of H3K36me2 (Fig. 3g). Based on transcriptomic experiments, the impact of $mNSD3_{T1242A}$ expression status on the total number of differentially expressed genes (DEGs) is, as expected, far greater when comparing *PSCN* to *PSC* tumor biopsies than for PSC_N cells ±NSD3 depletion, in which NSD3 was depleted for a relatively short time prior to the analysis (Extended Data Figs. 5i-j; Supplementary Tables 2-3). Nonetheless, the gene set enrichment analysis (GSEA) profiles of the tumor and cellular datasets show highly significant overlap with the same pathways, including strong correlations with MYC targets and mTOR signaling signatures (Fig. 3h; Extended Data Fig. 5k). We note that regulation of these genes (e.g. *Prkaa2, Myc* and *Irgm1*) was observed in independent NSD3 depleted cell lines and dependent on NSD3's catalytic activity (Extended Data Figs. 5l-m).

Next, Cleavage Under Targets and Release Using Nuclease (CUT&RUN)[14] was used to profile the genomic distribution of H3K36me2 in PSC_N cells ±NSD3 depletion. We also mapped H3K27me3 as the deposition of this important chromatin modification is directly antagonized by H3K36 methylation (reviewed in [9]). An averaged genome-wide distribution of H3K36me2 in PSC_N cells across all gene bodies shows that the signal peaks 5' of the transcription start site (TSS), with a second smaller peak 3' of the TSS, followed by a slow decay towards the 3' end of genes (Fig. 3i; Extended Data Fig. 6a). Upon NSD3 depletion, the intensity of the H3K36me2 signal is proportionally lower across the averaged gene unit with little change in the overall pattern (Fig. 3i; Extended Data Fig. 6a). In contrast, the NSD3-dependent distribution of H3K36me2 specifically at the gene bodies of downregulated DEGs (dDEGs), the most likely direct targets of NSD3, diverges from the overall genome-wide pattern in two ways: (i) the baseline signal is 2-3 fold higher and (ii) the signal distribution across the gene is distinct, with the most intense peak being 3' – rather than 5' – proximal to the TSS (Fig. 3i; Extended Data Fig. 6a). Furthermore, quantitation of the difference between H3K36me2 peak signal intensities ±NSD3 depletion reveals that the magnitude of peak change at dDEGs gene bodies is significantly greater than the change observed genome-wide (Fig. 3j). Notably, relative to other genes, dDEGs show both high H3K36me2 chromatin signal and high mRNA transcript levels (Extended Data Fig. 6b). These data suggest that robust transcription may predispose genes to NSD3 deregulation.

For H3K27me3, there is an overall increase in signal intensity upon NSD3 depletion genome-wide and at dDEGs, with no quantitative difference in the magnitude of change between the two gene groups (Figs. 3i-j). The inverse relationship between H3K36me2 and

H3K27me3 in relation to NSD3 status extends to intergenic regions (Extended Data Fig. 6c-d). The pattern of diminished H3K36me2 and increased H3K27me3 upon NSD3 depletion is also observed at the level of individual dDEG tracks (Extended Data Fig. 6e). Finally, using chromatin immunoprecipitation (ChIP) assays, we observe a reduction in occupancy of mNSD3$_{T1242A}$ and H3K36me2 at target genes (*Prkaa2* and *Irgm1*) upon NSD3 depletion, which was reconstituted to control levels by complementation with catalytically active, but not catalytically-dead, NSD3 (Extended Data Figs. 6f-h). Together, these data suggest that aberrant H3K36me2 synthesis by NSD3 directly regulates an LUSC-advancing gene expression program.

NSD3 promotes human LUSC tumorigenesis

To explore the role of NSD3 in human LUSC, we assembled the majority of publicly available human LUSC cell lines, including all five available 8p11^{AMP} cell lines, an 8p11^{AMP}-negative NSD3-overexpressing cell line with a PSC-like genetic background, and six 8p11^{AMP}-negative control cell lines lacking NSD3 overexpression (Fig. 4a; Extended Data Fig. 7a). NSD3 depletion impaired xenograft tumor formation in the 8p11^{AMP}/NSD3-overexpressing human LUSC cell lines but did not affect tumor growth in the six control cell lines (summarized in Fig. 4a; see Extended Data Fig. 4b). The phenotype in the five 8p11^{AMP} LUSC lines is unlikely resulting from non-specific DNA damage due to sgNSD3-targeted cutting within amplified regions[15] as: (i) sgFGFR1 and sgPLPP5 had no effect on xenograft growth in 8p11^{AMP}-positive H520 cells (Extended Data Fig. 2e), (ii) γH2AX levels were not altered by sgNSD3 treatment in 8p11^{AMP} cell lines (Extended Data Figs. 2d and 7b), and (iii) depletion of NSD3 in the 8p11^{AMP} LUSC cell lines by RNAi also attenuated xenograft growth (Fig. 4a; Extended Data Fig. 7c). Finally, complementing NSD3-depleted H520 cells with NSD3$_{WT}$ and NSD3$_{T1232A}$ reconstituted xenograft tumor growth, whereas catalytically-dead NSD3 and a short NSD3 isoform lacking the catalytic domain (NSD3$_{short}$) previously implicated in breast cancer pathogenesis[6] both failed to reconstitute full xenograft growth (Fig. 4b; Extended Data Fig. 7d).

We next investigated the ability of NSD3 to cooperate with SOX2 to transform human tracheobronchial epithelial (AALE) cells[16]. As previously reported, ectopic expression of SOX2 with a second transcription factor (FOXE1) promotes anchorage-independent growth of AALE cells[16]. We found that ectopic expression of SOX2 with either NSD3$_{WT}$ or NSD3$_{T1232A}$ was more efficient in transforming AALE cells than SOX2/FOXE1, with SOX2/NSD3$_{T1232A}$ showing the greatest promotion of anchorage-independent growth *in vitro* (Extended Data Figs. 7e-g) and tumor growth *in vivo* (Fig. 4c; Extended Data Figs. 7e,h). In contrast, catalytically-dead NSD3 and NSD3$_{short}$ failed to promote AALE transformation (Fig. 4c; Extended Data Figs. 7e-h). Together, these data argue that NSD3, via its catalytic activity, promotes tumor growth of human LUSC cells and oncogenic transformation of human AALE cells.

To further investigate roles of the NSD3-H3K36me2 axis in human cancer a collection of 37 LUSC PDX samples was screened for NSD3 alterations. Of these, one was NSD3$_{T1232A}$ positive (PDX^{T1232A}), twelve were *NSD3* amplification positive (PDX^{AMP}), and the others had no NSD3 alteration (PDX^{CTR}) (Extended Data Fig. 8; data not shown). Four PDXs

(PDX^{T1232A}, PDX^{AMP-1}, and two controls (PDX^{CTR})), which apart from the NSD3 alterations have similar mutational profiles (e.g. PSC-signature), were selected for further analysis. In comparison to PDX^{CTR} samples, NSD3 levels are higher in PDX^{T1232A} cells and highest in PDX^{AMP-1} cells (Fig. 4d). H3K36me2 levels in the PDX^{T1232A} and PDX^{AMP-1} cells are comparable to one another and elevated relative to controls (Fig 4d); similar H3K36me2 abundance in PDX^{T1232A} and PDX^{AMP-1} cells, while the former has lower NSD3 levels, is consistent with NSD3$_{T1232A}$ hyperactive. NSD3 depletion, which resulted in H3K36me2 loss, attenuated growth of PDX^{T1232A} and PDX^{AMP-1} xenograft tumors (Figs. 4e-g) but did not impact growth of PDX^{CTR} samples (Extended Data Figs. 9a-b). PDX growth inhibition upon NSD3 depletion was also seen with an independent 8p11^{AMP} PDX sample (Extended Data Fig. 9c). Thus, a subset of PDXs from primary LUSC human tumors are NSD3-regulated and selectively sensitive to NSD3 depletion.

NSD3 sensitizes LUSC to BETi therapy

While at present there is no NSD3 catalytic inhibitor that can be used in a physiologic setting, we speculated that selection for NSD3-dependency may confer adaptive vulnerability of LUSC to drugs targeting known cancer pathways. To test this idea, we screened a library comprising 285 characterized inhibitors covering ~170 cancer targets to identify drugs that have increased efficacy in PSC$_N$ versus PSC$_C$ cells. The four BETi drugs present in the library exhibited the highest differential lethality, consistent with BRD4 interacting directly with NSD3 in LUSC cells (Supplementary Table 4; Extended Data Fig. 9d)[17,18]. Under normal conditions, PSC$_N$ cells proliferate more rapidly than PSC$_C$ cells (Extended Data Fig. 9e). However, treatment with the BETi AZD5153[19], which has a modest effect on PSC$_C$ cell proliferation, inhibits NSD3-target gene expression and proliferation of PSC$_N$ cells (Extended Data Fig. 9e-f). NSD3-dependent BETi sensitivity was also observed in PDXs; actively growing PDX^{T1232A} and PDX^{AMP} tumors stall in response to BETi treatment, a phenotype that is weakened by NSD3 depletion (Figs. 4h-i). In contrast, PDX^{CTR} samples are moderately responsive to BETi therapy independent of NSD3 (Extended Data Figs. 9g-h). Thus, the oncogenic advantage provided by NSD3 to LUSC potentially comes with a cost – BETi-hypersensitivity – a clinically actionable vulnerability.

Summary

LUSC, one of the leading causes of cancer-related mortality worldwide, is characterized by several well-defined driver mutations. However, in contrast to malignancies like lung adenocarcinoma in which targeted therapy approaches have been encouraging, translating similar strategies for LUSC has been challenging. Our study has uncovered a pivotal role for NSD3 in LUSC, a finding with potential prognostic and therapeutic relevance for the estimated 100,000 patients diagnosed worldwide each year with 8p11^{AMP}-LUSC[20]. We propose a model in which NSD3, via H3K36me2, acts as an epigenetic deregulator to facilitate expression of oncogenesis-promoting genes (Extended Data Fig. 10). The genes most acutely impacted by NSD3 are marked by high basal H3K36me2 and are robustly transcribed. These features are consistent with the vulnerability of NSD3-regulated LUSC to BETi, as this drug class targets active transcription[21]. Indeed, NSD3 depletion was recently

identified as a strong sensitizer of acute myeloid leukemia cells to BETi[22]. Thus, while clinical NSD3 inhibitors are presently unavailable, the increased sensitivity of NSD3-regulated LUSC to BETi may expand the narrow therapeutic window for these medicines[21]. Finally, NSD3's oncogenic activity likely extends beyond LUSC, as the 8p11-12 amplicon is a common molecular signature of breast cancer and other malignancies[7].

METHODS

Plasmids.

The cloning of full-length NSD3 and NSD3$_{SET}$ (aa1021-1320) was based on NCBI sequence NM_023034.2. Full length NSD3 was cloned into pCAG vector for transient transfection or into pENTR3C entry vector for subsequent cloning into pLenti-CMV-hygro (w117-1) destination vector for expression in HT1080 cells. Different versions of V5-NSD3 (WT, T1232A, Y1174A, T1232A/Y1174A or Short), SOX2 and V5-FOXE1[23] and AkaLuc (gift from Dr. Michael Z. Lin)[24] were cloned into pENTR3C entry vector for subsequent cloning into pPB-CAG-Dest-pA-pgk-bsd (Addgene) destination vector for reconstitution in H520 cells or into pPB-CAG-Dest-IRES-puro destination vector for transformation in AALE cells [23]. NSD3$_{SET}$ was cloned into pGEX-6P-1 vector and all the different mutant versions of NSD3$_{SET}$ were generated by site-directed mutagenesis. The choice of residue Y1174A as a catalytic mutant of NSD3 was based on its homology to a validated NSD2 catalytic mutant (Y1092A)[25] and independently shown to abrogate NSD3 catalytic activity (data not shown). Wildtype and T1232A mutant versions of NSD3$_{SET}$ (aa1055-1285) were cloned into pQE30 vector for NMR analysis.

Immunoblot analysis and co-immunoprecipitation.

For western blot analysis, cells were lysed in RIPA buffer with 1 mM PMSF and protease inhibitor cocktail. Protein concentration was determined using the Pierce Coomassie Plus Assay. Protein samples were resolved by SDS-PAGE and transferred to a PVDF membrane. The following antibodies were used (at the indicated dilutions): H3K36me2 (Thermo Fisher Scientific #701767, 1:1000), H3K36me3 (Thermo Fisher Scientific #MA5-24687, 1:1000), H3K27me3 (Thermo Fisher Scientific #MA5-33081, 1:1000), V5 (Thermo Fisher Scientific #R960-25, 1:1000), H3K36me1 (Abclonal #A2364, 1:5000); H3K4me2 (CST #9725S, 1:1000), H3K9me2 (CST #4658S, 1:1000), H3K27me2 (CST #9728S, 1:1000), γH2AX (CST ##9718, 1:1000), SOX2 (CST #3579, 1:1000), c-Myc (CST #5605S, 1:1000), Phospho-4E-BP1 (CST #2855S, 1:1000), 4E-BP1 (CST #9644S, 1:1000), FGFR1 (CST #9740S, 1:1000), H4K20me2 (Abcam #ab9052, 1:1000), H3 (Abcam #ab1791, 1:5000), BRD4 (Abcam #ab128874, 1:1000), β-Actin (CST #4970, 1:5000), β-Tubulin (Millipore #05-661, 1:5000), NSD2[25] (1:500), NSD3 antibody (1:1000) was generated using protein fragment corresponding to amino acid residues 1021-1320 of NSD3 (to detect full-length NSD3 but not NSD3$_{short}$) as the epitope to generate antiserum (Genemed Synthesis Inc.) followed by purification with immobilized epitope, NSD3 (CST #92056S, 1:1000) was used for detecting both NSD3 and NSD3$_{Short}$.

For co-immunoprecipitation, nuclear extracts were incubated with 4 μg BRD4 antibody or IgG control overnight at 4°C. Extracts were then incubated with protein A Sepharose beads

for 2 hours at 4°C then washed and eluted in 2× sample buffer at 98 °C. Eluates and nuclear extract (input sample) were resolved by SDS-PAGE, transferred to a PVDF membrane and immunoblotted

Preparation of recombinant proteins, nucleosomes, and DNA.

E.coli BL21-RIL cells transformed with respective expression vectors were cultivated in 2xYT or M9 media at 20°C for 16-20 hours. Cells were lysed using a French press (Constant Systems TS-Series) and lysates were cleared by centrifugation at 14000 rpm for 1 hour and subsequently injected into a HiTrap SP FF 5 ml column followed by elution with a salt gradient. Fractions containing $NSD3_{SET}$/$NSD3_{SET-T1232A}$ expressed from pGEX-6P-1 were collected and injected into a GSTrap HP 5 ml column followed by elution with a gluthatione gradient. Eluate was mixed with GST-PreScission protease in a 100:1 ratio and dialyzed overnight at 4°C against storage buffer (50 mM TRIS/TRIS-HCl, pH = 7.5, 150 mM NaCl, 1 mM TCEP). Final purification was achieved by passing through a GSTrap HP 5 ml column, the flow through was collected and contained proteins were concentrated with an Amicon 3000 MWCO. Proteins expressed from pQE30 were purified with the help of a nickel affinity column HisTrap FF 5 ml, and then dialyzed overnight at 4°C against final NMR buffer. For the high-resolution NMR structural and dynamics studies, U-[^{15}N], U-[^{13}C,^{15}N] $NSD3_{SET}$ and $NSD3_{SET-T1232A}$ versions were prepared from pQE30 vector by expression in M9 media. $^{15}NH_4Cl$ at a concentration of 1 g/L served as sole source of nitrogen, for U-[^{13}C,^{15}N] labeling a ^{13}C-glucose (CIL) at 3 g/L was added. Recombinant nucleosomes were assembled using 601 Widom DNA[26]. The 601 Widom DNA with 20bp overhang (187bp) and without overhang (147bp) were prepared by PCR and purified by phenol/chloroform extraction and isopropanol precipitation. The Cy5-labeled dsDNA (147bp or 187bp) was prepared by annealing oligonucleotides 187-F Cy5-CGAGGCTGTTCAATACATGC, 187-R GGACCCTATACGCGGCC, 147-F Cy5-CTGGAGAATCCCGGTGCCGAGGC, 147-R ACAGGATGTATATATCTGACACGTGCCTG in equimolar ratio at 10 M concentration via heating to 95°C and slow cooling at 0.1°C/5 s in a PCR thermocycler. Human H2A and H2B histones were prepared according to previously described protocol[27]. Histone octamers and nucleosomes were assembled according to methods described previously[28].

Purification of Recombinant Full-Length NSD3.

HEK 293T cells were transfected with pCAG-FLAG-NSD3 (WT or NSD3 mutants) or empty vector. Cells were lysed with buffer containing 10 mM HEPES (pH = 7.9), 1.5 mM $MgCl_2$, 10 mM KCl to deplete cytoplasmic fractions, the pellet was treated with buffer containing 20 mM HEPES (pH = 7.9), 1.5mM $MgCl_2$, 420 mM KCl, 0.2 mM EDTA, 10% glycerol, 0.2 mM PMSF and proteinase inhibitor cocktail (Roche) to get nuclear extract. The nuclear extract was diluted to contain 300 mM KCl, full-length NSD3 was affinity-purified with anti-FLAG(M2)-conjugated agarose (Sigma) from the diluted nuclear extract and eluted with 0.5 mg/ml FLAG peptides for in vitro methylation assays.

In vitro methylation assay.

In vitro methylation assays were performed as previously described[29] using 2 μg of enzyme (unless otherwise noted) and 2 μg of recombinant nucleosomes (EpiCypher, catalog#s:

16-0009 and 16-2004) or recombinant histone H3 (generated as describe in [30]) in the presence of 3H-SAM(S-adenosyl-L-methionine) or non-radiolabeled SAM.

Cell lines, primary cell cultures and cell assays.

HEK-293T (ATCC #CRL-1573), KNS-62 (JCRB #IFO50358), SK-MES-1 (ATCC #HTB-58), HT1080 (ATCC #CCL-121) cells were grown in DMEM medium supplemented with 10% fetal calf serum, 2 mM L-glutamine and penicillin–streptomycin (Life Technologies). H520 (ATCC #HTB-182), H1703 (ATCC #CRL-5889), HCC15 (DSMZ #ACC496), HCC95 (KCLB #70095), VMRC-LCP (JCRB #0103), EBC-1 (RIKEN #RCB1965), LK-2 (RIKEN # RCB1970), RERF-LC-Sq1 (JCRB #1019), LUDLU-1 (ECACC # 92012463) cells were grown in RPMI 1640 medium supplemented with 10% fetal calf serum, 2 mM L-glutamine and penicillin–streptomycin (Life Technologies). H1869 (ATCC #CRL-5900) cells were grown in DMEM/F12 medium supplemented with 10% fetal calf serum, 2 mM L-glutamine and penicillin–streptomycin (Life Technologies). Primary mouse cancer cell lines were prepared from tumor biopsies isolated from indicated mouse models according to the methods previously described [31]. Primary mouse cancer cell lines were cultured in DMEM medium supplemented with 10% fetal calf serum, 2 mM L-glutamine and penicillin–streptomycin (Life Technologies). Primary cultures of mouse lung fibroblast were prepared from indicated mice using standard methods[32]. Briefly, mice were euthanized by CO_2 asphyxiation and perfused with 5 ml of normal saline. Lungs were sterilely removed and cut into small 2-mm slices and allowed to adhere on tissue culture plate. Lung tissue explants were cultured in DMEM medium supplemented with 10% fetal calf serum, 2 mM L-glutamine and penicillin–streptomycin (Life Technologies). Lung fibroblasts were purified by repeat trypsinization and passaging to achieve a homogenous population. Lung fibroblasts were treated with Adenovirus-Cre or vehicle (PBS). Lysates were obtained for Western blot analyses 72 hours post transduction. All cells were cultured at 37°C in a humidified incubator with 5% CO_2. Cell lines were authenticated by short tandem repeat profiling and tested negative for mycoplasma (PromoKine). Cell proliferation assays and drug screen was performed as previously described[33]. Briefly, cells were seeded at 5×10^3 cells/mL in triplicate in 96-well plates. Confluency of cells treated was measured using live cell kinetic imaging system (IncuCyte Zoom) for 120 h with data collection every 4h. Data are represented as mean ± SEM (three independent experiments).

Transfection and Viral Transduction.

Transient expression was performed using TransIT-293 (Mirus Bio) following the manufacturer's protocol. For NSD3 reconstitution, different versions of NSD3 gene were constructed into pPB-CAG-Dest-pA-pgk-bsd (Addgene) and transiently co-transfected with transposase into recipient cells with Lipofectamine 3000 (Invitrogen) following the manufacturer's protocol. After 48 hours of transfection, cells were selected with 5 μg /mL blasticidin for 6 days, then CRISPR mediated knockdown of endogenous NSD3 was performed in the selected cells. For CRISPR-Cas9 or shRNA knockdowns, virus particles were produced by co-transfection of 293T cells with the lentiCRISPRv2/puro (Addgene) construct expressing indicated guide RNAs or pLKO.1/puro (Addgene) construct expressing indicated shRNAs, psPAX2 and pCMV-VSVg in a ratio of 5:3:2 by mass. The following guide RNA or shNRA oligo sequences were utilized: sgControl

CTTCGAAATGTCCGTTCGGT, sgNSD2 ACTCGTTAACAAATTCTCCC (made by Saumya Sankaran), sgNSD3 GGATACTGATTATATGAC, sgFGFR1 TCCCCGACCTTGCCTGAACA, sgPLPP5 CAATAAAACTGATCGTAGGG, shControl CCTAAGGTTAAGTCGCCCTCG, shNSD3 GCAGATTGTTTGGGTCAAATT. After 48 hours of transfection, target cells were transduced with 0.45 μm filtered viral supernatant and 8 μg/mL polybrene. Cells were selected 24 hours after media replacement with 2 μg /mL puromycin. For NSD3 overexpression in HT1080 cells, virus particles were produced by co-transfection of 293T cells with pLenti CMV Hygro DEST (W117-1) (Addgene) or pLKO.1 constructs, psPAX2 and pCMV-VSVg in a ratio of 5:3:2 by mass. After 48 hours of transfection, target cells were transduced with 0.45 μm filtered viral supernatant and 8 μg/mL polybrene. The subsequent selection was carried out with 250 μg/mL hygromycin.

Mice.

$ROSA26^{LSL-Pik3ca(H1047R)}$, $Col1a1^{LSL-Sox2}$, $Cdkn2a/b^{LoxP/LoxP}$, $Fgfr1^{LoxP/LoxP}$ mice have been described before[34-37]. Reporter-tagged insertion with conditional potential $Nsd3^{tm1a(EUCOMM)}$ mouse strain originates from ES clone #HEPD0563-4-G02 obtained from European Mouse Mutant Archive repository[38]. The targeting vector includes the Neo-LacZ cassette flanked by Frt sites and exon 4 sequence flanked by LoxP sites. Founder mice ($Nsd3^{LacZ}$) were confirmed as germline-transmitted *via* crossbreeding with C57BL/6N wild-type animals. Next, $Nsd3^{LacZ}$ mice were crossed with $ROSA26^{FlpO}$ deleter strain[39] to generate conditional allele $Nsd3^{LoxP/LoxP}$. $LSL-Nsd3^{T1242A}$ model was generated by knockin of the CAG-LoxP-Stop-LoxP-V5-mNsd3^{T1242A} cDNA-polyA cassette into intron 1 of $ROSA26$ using methods previously described[40]. Founder animals were identified by PCR followed by sequence analysis and germline transmission confirmed by crossbreeding with C57BL/6N wild-type animals. All mice were maintained in a mixed C57BL/6;129/Sv background, and we systematically used littermates as controls in all the experiments. Immunocompromised NSG mice ($NOD.SCID-IL2Rg^{-/-}$) were utilized for tumor xenograft studies. All experiments were performed on balanced cohorts of male and female mice as our initial data did not indicate significant differences in disease progression or response to treatment between females or males. All animals were numbered and experiments were conducted in a blinded fashion. After data collection, genotypes were revealed and animals assigned to groups for analysis. For treatment experiments mice were randomized. None of the mice with the appropriate genotype were excluded from this study or used in any other experiments. Mice had not undergone prior treatment or procedures. All mice were co-housed with littermates (2–5 per cage) in pathogen-free facility with standard controlled temperature of 72 °F, with a humidity of 30–70%, and a light cycle of 12 h on/12 h off set from 7am to 7pm and with unrestricted access to standard food and water under the supervision of veterinarians, in an AALAC-accredited animal facility at the University of Texas M.D. Anderson Cancer Center (MDACC). Mouse handling and care followed the NIH Guide for Care and Use of Laboratory Animals. All animal procedures followed the guidelines of and were approved by the MDACC Institutional Animal Care and Use Committee (IACUC protocol 00001636, PI: Mazur). Tumor size was measured using a digital caliper and tumor volume was calculated using the formula: Volume = $(width)^2 \times length / 2$ where *length* represents the largest tumor diameter and *width* represents the perpendicular tumor diameter. The endpoint was defined as the time at which a

progressively growing tumor reached 20 mm in its longest dimension as approved by the MDACC IACUC protocol (00001636, PI: Mazur). and in no experiments was this limit exceeded.

Lung Squamous Cell Carcinoma Mouse Models.

To generate tumors in the lungs of $ROSA26^{LSL-Pik3ca(H1047R)}$, $Col1a1^{LSL-Sox2}$ $Cdkn2a/b^{LoxP/LoxP}$ (PSC), $ROSA26^{LSL-Pik3ca(H1047R)}$; $Col1a1^{LSL-Sox2}$; $Cdkn2a/b^{LoxP/LoxP}$; $Fgfr1^{LoxP/LoxP}$ ($PSC^{FGFR1-KO}$), $ROSA26^{LSL-Pik3ca(H1047R)}$; $Col1a1^{LSL-Sox2}$; $Cdkn2a/b^{LoxP/LoxP}$; $Nsd3^{LoxP/LoxP}$ ($PSC^{NSD3-KO}$), $ROSA26^{LSL-Pik3ca(H1047R)}$; $Col1a1^{LSL-Sox2}$; $Cdkn2a/b^{LoxP/LoxP}$; $LSL-Nsd3^{T1242A}$ ($PSCN$) mutant mice, we used replication-deficient adenoviruses expressing Cre-recombinase (Ad-Cre) as previously described[41]. Briefly, 8-week old mice were anesthetized by continuous gaseous infusion of 2% isoflurane for at least 10 min using a veterinary anesthesia system (D19 Vaporizer, Vetland Medical). Ad-Cre was delivered to the lungs by intratracheal intubation. Prior to administration, Ad-Cre was precipitated with calcium phosphate to improve the delivery of Cre by increasing the efficiency of viral infection of the lung epithelium. Mice were treated with one dose of 5×10^6 PFU of Ad-Cre (Baylor College of Medicine, Viral Vector Production Core). Mice were analyzed for tumor formation and progression at indicated times after infection.

Xenograft models.

Patient-derived xenografts (PDXs) were obtained from the NCI Patient-Derived Models Repository (PDMR), NCI-Frederick, Frederick National Laboratory for Cancer Research (997726-040-R, 692585-246-R, 417821-307-R) and the University of Texas M.D. Anderson Cancer Center repository. Briefly, surgically resected tumor specimens were obtained from deidentified patients with histologically confirmed LUSCs. All tumor specimens were collected after written patient consent and in accordance with the institutional review board-approved protocols of the University of Texas M.D. Anderson Cancer Center (PA19-0435, PI: Mazur). Patient-derived xenograft tumors were generated and propagated by transplanting small tumor fragments isolated directly from surgical specimens subcutaneously into NSG mice as we established previously[33]. Genome-wide copy number analysis was performed for each PDX using an OncoScan CNV Assay Kit (Thermo Fisher Scientific) as previously described[42]. BioDiscovery's Nexus Copy Number software was used to call CNAs using the default parameters. All data were analyzed and reported using the February 2009 NCBI human genome build 37.1 (hg19). $NSD3$ mutation analysis was performed by Sanger sequencing of exon 20 using genomic DNA and following primers NSD3ex20-F GACCGTATAATTGATGCCGGC, NSD3ex20-R CTGCAGGAATATCACAGAGAGC. For analysis of NSD3 knockdown growth, collected PDX tumors were minced using a razor blade and digested in collagenase digestion buffer at 37°C for 1 hour. Cells were passed through 100 μm and 40 μm cell strainers and centrifuged for 1200 rpm for 5 min. Cells were incubated in RBC lysis buffer for 2 min and then resuspended in 6 mL of media and spun through 0.5 mL of serum layered on the bottom of the tube to remove cellular debris. Contaminating human or mouse hematopoietic and endothelial cells (CD45, Ter119, CD31) are depleted using biotin conjugated anti-mouse CD45, CD31 and Ter119 antibodies and separated on a MACS LS column using anti biotin microbeads. Next, the cells were transduced with lentivirus expressing sgRNA/Cas9 and

selected with puromycine for 72 hours. Next, the cells were collected, mixed with Matrigel (1:1) and transplanted to the flanks of NSG mice. For xenograft studies of human LUSC lines, transduced cells were cultured in 15- cm dishes trypsinized and singularized. The trypsin was washed with excess growth medium and the cells were counted. The cells were then resuspended in PBS and mixed with matrigel (1:1) at a density of 2×10^7 cells per ml and kept on ice until injection. Next, 100 μl of the cell suspension was injected subcutaneously onto the hind flanks of NSG mice. When tumors became palpable, they were calipered to monitor growth kinetics.

Tumorigenicity Assay *in vitro* and *in vivo*.

Immortalized human tracheobronchial epithelial (AALE) were described before[23]. AALE cells were maintained in Small Airway Epithelial Cell Growth Medium (SAGM) containing SAGM growth supplements (Lonza). AALE cells (1×10^7) were electroporated with PiggyBac transposon plasmids using NEPA21 electroporator (Nepagene) at the following settings: two poring pulses: 175 V, length 5.0 ms, interval 50.0 ms, decay rate 10%, positive polarity and five transfer pulses: 20 V, length 50.0 ms, interval 50.0 ms, decay rate 40%, at reversing polarities. For *in vitro* tumorigenic ability growth in soft agar was determined by plating 5×10^5 cells in triplicate. Colonies were counted 3 weeks after plating. For *in vivo* tumorigenic ability – ALLE cells were transplanted bilaterally under sub-renal capsule into immunocompromised NSG mice using protocol previously described [43]. The cells resuspended in Matrigel (10^6 in 40 μl) were delivered through a blunted 25g needle attached to a silicone catheter. The catheter was advanced through the capsule incision to the cranial pole of the kidney and the cells discharged slowly from the catheter. For in vivo transformation assays AALE cells were electroporated with PiggyBac transposon plasmid expressing AkaLuc which catalyzes oxidation reaction of a substrate AkaLumine producing near-infrared bioluminescence which can penetrate most animal tissues [24,44]. To monitor cell growth mice were injected i.p. with 3 μmol (1.0 mg) of AkaLumine-HCl (Sigma-Aldrich) in 100 μl 0.9% NaCl. Immediately after substrate injection, bioluminescent images were acquired in an AMI HTX bioluminescence imaging system (Spectral Instruments Imaging). Imager settings were: Emission filter, open; field of view, 25 cm; f-stop 1.2; low binning 2 × 2 and exposure time, 30 s. X-ray imaging camera settings were: field of view, 25 cm; low exposure and high resolution. Images were analyzed in the Aura software (Spectral Instruments Imaging) and quantified in radiance units of photons per second per square centimeter per steradian (photons/s/cm^2/sr) and plotted as mean ± s.e.m..

Micro-Computed Tomography.

Micro Computed Tomography (μCT) scans were performed on indicated tumor mice at 90, 120 and 150 days after Ad-Cre induction as previously described[45]. Briefly, mice were anesthetized by continuous gaseous infusion of 2% isoflurane for at least 10 min using a veterinary anesthesia system. The mice were intubated using a 20 gauge × 1 inch catheter and were transferred onto the bed of Explore Locus RS pre-clinical *in vivo* scanner (GE Medical Systems). The mice were mechanically ventilated in a small animal ventilator, and mCT images were captured at 80 kV and 450 microamperes. The X-ray source and CCD-based detector gantry is rotated around the subject in roughly 1.0 degree increments. Each animal's breathing was held at a pressure of 20 cm H_2O during the 20 second acquisition.

The resulting raw data were reconstructed to a final image volume of $875 \times 875 \times 465$ slices at 93 μm^3 voxel dimensions. The images were cropped to display the lung region. The total chest space volume, including the heart, was selected using manual segmentation. An optimal threshold value was automatically determined using the function of the MicroView analysis software. Tumors formed at air space in the lung can be distinguished from other soft-tissue in a reconstructed 3D image of the higher voxels; therefore, the tumor nodule structure was selected using a combination of manual segmentation and semi-automated contouring of the optimal threshold value. These analyses were consistent between two independent operators and performed by a well-trained researcher in a blinded manner.

Histology and immunohistochemistry.

Tissue specimens were fixed in 4% buffered formalin for 24 hours and stored in 70% ethanol until paraffin embedding. 3 μm sections were stained with hematoxylin and eosin (HE) or used for immunohistochemical studies. Human tissue sections were collected in accordance with the institutional review board-approved protocols of the University of Texas M.D. Anderson Cancer Center (PA19-0435, PI: Mazur). Immunohistochemistry (IHC) was performed on formalin-fixed, paraffin embedded mouse and human tissue sections using a biotin-avidin method as described before[41]. The following antibodies were used (at the indicated dilutions): cleaved Caspase 3 (CST #9664, 1:100), H3K36me2 (CST #2901, 1:2000), Ki67 (BD Bioscience #550609, 1:1,000), KRT5 (CST #71536, 1:1000), NSD3 (described above), p63 (Abcam, #ab124762, 1:500) and TTF1 (Abcam #ab76013, 1:500). Sections were developed with DAB and counterstained with hematoxylin. Pictures were taken using a PreciPoint M8 microscope equipped with the PointView software. Analysis of the tumor area and IHC analysis was done using ImageJ software. Quantification of NSD3 IHC chromogen intensity was performed by measuring the reciprocal intensity of the chromogen stain as previously describe[46]. Briefly, standard RGB color images acquired from bright field microscopy have a maximum intensity of value 250 (represented by white, unstained areas) as measured by the standard intensity function in the open source ImageJ Fiji software. We subtracted the intensity of a stained tissue sample from 250, thereby deriving a reciprocal intensity that is directly proportional to the amount of chromogen present.

Microscale thermophoresis (MST).

Binding affinities (K_d) were determined from three independent experiments using Cy5-labeled dsDNA (147bp or 187bp)-nucleosomes at concentration 100 nM mixed in 1:1 volume ratio with appropriate $NSD3_{SET}$ or $NSD3_{SET-T1232A}$ proteins at 16 different concentrations. Buffer conditions were 50 mM TRIS/TRIS-HCl, at pH 7.5, 150 mM NaCl, 1 mM TCEP. Binding was quantified by microscale thermophoresis on a Monolith NT.115Pico instrument with all measurements performed at room temperature.

ITC measurements.

SAM [S-(5′-Adenosyl)-L-methionine chloride dihydrochloride; Merck A7007] was dissolved in 50 mM TRIS/TRIS·HCl pH = 7.5, 150 mM NaCl, 1 mM TCEP, and the pH was checked. SAM was used at 150 μM concentration, $NSD3_{SET}$ or $NSD3_{SET-T1232A}$ domains at 100-125 μM, and ITC measurements were performed in a MicroCal PEAQ-ITC instrument

(Malvern Panalytical) at 20°C with 55 injections of 0.7 μL of SAM. Data were analyzed using ITC software to obtain K_ds and stoichiometry.

Thermal stability essays.

The SET-domain stability in NSD3$_{SET}$, NSD3$_{SET-T1232A}$ or NSD3$_{SET-T1232A/V1243A}$ was studied by protein thermal stability assay. The proteins (at 1mg/ml) were incubated with SYPRO Orange (ThermoFisher, cat. No S6651) for 1min at 25 °C, containing 50 mM TRIS/TRIS·HCl, 150 mM NaCl, pH 7.5, and 1 mM TCEP. The stability essay was repeated for each protein 8 times and resulting temperatures were averaged. Fluorescence measurements (excitation, 450-490 nm; detection, 560-580 nm; BioRad CFX96 Touch Real-Time PCR Detection System) were collected following 30s incubation steps, respectively, with each step including a rise of 0.5 degree starting 15 °C to finally 95 °C.

NMR spectroscopy.

All NMR experiments were carried out at 25°C on NSD3$_{SET}$ or NSD3$_{SET-T1232A}$ samples at 200-370 μM in buffer of 50 mM TRIS/TRIS·HCl, pH = 7.5, 150 mM NaCl, and 1 mM TCEP with 1% / 99% (*v/v*) D_2O/H_2O, and a twofold molar excess of SAM, using Bruker NEO 700 and 950 MHz spectrometers with inverse-detected TCI cryogenic probes. Nearly complete sequence specific backbone resonance assignments (>95% for NSD3$_{SET}$ or NSD3$_{SET-T1232A}$) were achieved on double labeled 200-370 μM U-[$^{13}C,^{15}N$] NSD3$_{SET}$ or NSD3$_{SET-T1232A}$ samples using transverse relaxation optimized spectroscopy, TROSY-based triple resonance three-dimensional (3D) experiments of HNCO, HNcaCO, HNCA, HNcoCA, HNCACB, CBCAcoNH with further support from TROSY-version of 3D ^{15}N-edited NOESY (100 ms mixing time) recorded on 370 μM U-[^{15}N] samples[47-49]. Data were processed using Topspin 4.07 program and analyzed with the program SPARKY (Goddard and Kneller) and CARA (cara.nmr.ch). NSD3$_{SET}$ and NSD3$_{SET-T1232A}$ backbone torsion angles were determined with Talos-N software[50]. The mobility of the protein backbone was studied by NMR ^{15}N spin relaxation measurements using TROSY-based $^{15}N[^{1}H]$ nuclear Overhauser effect (NOE) experiments optimized for high-molecular weight proteins on 370 μM uniformly labeled U-[^{15}N] NSD3$_{SET}$ or NSD3$_{SET-T1232A}$ preparations containing variants[51]. The recycle delay for $^{15}N[^{1}H]$ NOE was 10 s and 4 s of ^{1}H saturation. The errors of $^{15}N[^{1}H]$ NOE were determined based on the following equation: $dNOE_i = |NOE_i| \cdot [(1/SNR_{NOE,i})^2 + (1/SNR_{noNOE,i})^2]^{0.5}$, where SNR is signal to noise ratio for the experiment with (noNOE) and without saturation (NOE) and $|NOE_i|$ the intensity ratio of saturated and non-saturated signal. The chemical shift perturbation of $^{1}H/^{15}N$ resonances was determined with the equation: $CSP_i = [(\delta_{H,i})^2 + 0.25 \cdot (\delta_{N,i})^2]^{0.5}$[52], where $\delta_{H,i}$ and $\delta_{N,i}$ are detected chemical shift changes of proton and nitrogen, respectively.

Docking of H3 peptide to NSD3$_{SET}$ domain and modelling the NSD3$_{SET}$ auto-inhibited state.

The HADDOCK2.4 program[53] was used in order to dock a H3.1 peptide (residues 29-42) into the binding cleft of NSD3$_{SET}$. The starting structure for the NSD3$_{SET}$ domain was the 6CEN X-ray structure[54] with the G^{1267}-N^{1268} residues modelled which are lacking from the experimental electron density in the SwissModel[55]. The structure of NSD3$_{SET}$ (PDB id: 6CEN) is optimal as template for this approach as it is complexed with a peptidic inhibitor

and its backbone is overlapping nearly completely with the analogical residues from H3.3K36M peptide (residues 29-42) in the SETD2$_{SET}$:H3.3K36M complex[56,57]. The starting peptide pose for H3.1 peptide for NSD3$_{SET}$ was taken from SETD2$_{SET}$:H3.3K36M model, with M36 modified to K36 and H3.3 into H3.1 (S^{32} to A^{32}) and used to define ambiguous interaction restraints (AIRs) driving the docking process. Remaining parameters were treated according to standard docking protocol[53]. The NSD3$_{SET}$ with the regulatory loop in the closed conformation (auto-inhibited state) was modelled using SwissModel[55] and the crystal structure of NSD2$_{SET}$ [PDB id: 5LSU:A; fragment 983-1203][58] as the template (sequence identity 79.2%, similarity 87.3%).

RNA-seq and Real-Time PCR Gene Expression Analysis.

RNA samples were extracted from cells using TRIzol reagent (Life Technologies), the RNA-seq libraries were constructed and sequenced in BGI. Or the RNA samples were reverse transcribed into cDNA using the SuperScript IV First Strand Synthesis System (Invitrogen). Quantitative real-time PCR analysis was performed on a Roche LightCycler 480 using SYBR Green Master Mix (Applied Biosystems) following the manufacturer's manual. The expression of each gene was normalized to actin (Actb). The primer sequences are listed below: *Irgm1*-qPCR-F AGCATCTTTGTGACTGGGGA, *Irgm1*-qPCR-R TAAGTCCCACAGCACCACAT, *Prkaa2*-qPCR-F TACATCTGCAAACATGGGCG, *Prkaa2*-qPCR-R TCTTAGCGTTCATCTGGGCA, *Myc*-qPCR-F CCTTCTCTCCTTCCTCGGAC, *Myc*-qPCR-R TGCCTCTTCTCCACAGACAC, *Actb*-qPCR-F ACTGGGACGACATGGAGAAG, *Actb*-qPCR-R GTCTCCGGAGTCCATCACAA.

RNA-seq analysis.

The PSC$_N$ cells ±NSD3 depletion RNA-seq reads were aligned to mouse genome mm10 by HISAT2 (ccb.jhu.edu/software/hisat2)[59]. Then the transcriptomes were annotated to GENCODE M23 (gencodegenes.org)[60] by FeatureCounts (2.0.0) [61]. Differentially expressed genes (DEGs) were detected by DESeq2 package from Bioconductor (bioconductor.org/packages/release/bioc/html/DESeq2)[62] using likelihood ratio test (LRT, adjusted *P*-value < 0.001) or Wald test. The expression levels of genes were normalized and log2-transformed, and the variances-stabling transformation were implemented. Bioconductor package EnhancedVolcano (bioconductor.org/packages/release/bioc/html/EnhancedVolcano, 1.8.0) was used to generate the volcano plot. Hierarchical clustering was implemented to cluster the trends of the DEGs. The functional enrichment analyses were implemented by GSEA (gsea-msigdb.org/gsea, 4.0.0)[63] and the Bioconductor package GeneAnswers (bioconductor.org/packages/release/bioc/html/GeneAnswers, 2.32.0). The comparisons of expression levels of DEGs and all genes were implemented by two-sided robust t-test, and the trim level was 0.1.

TCGA data set analysis.

Lung squamous cell carcinoma patient TCGA data (Project ID: TCGA-LUSC, dbGaP Study Accession: phs000178)[64] was analyzed using cBioPortal [65,66] to identify copy number and mRNA expression. Copy number amplification datasets within the portal were generated by the Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm identifying

significantly altered regions of amplification or deletion across sets of patients. GISTIC-derived copy number analysis indicate copy-number level per gene, in which 1 and 2 indicates amplification. Basal-normalized transcript expression data (z score) used for this analysis were RNA Seq V2 RSEM.

CUT&RUN.

This assay was performed as described previously with modifications[67] and using the H3K36me2 and H3K27me3 antibodies noted above. Buffer exchange was performed by centrifugation at 4 °C, 600g for 5 minutes. After the addition of concanavalin A beads, the buffer exchange was performed with a magnetic stand, allowing at least 5 minutes for the beads to completely bind to the magnet. Cells were harvested into a single cell suspension, counted, washed once with ice-cold PBS and resuspended with ice-cold Nuclear Extraction (NE) buffer to a final concentration of 1×10^7 cells/mL. After 10 minutes on ice, cells were washed with Wash Buffer (WB) and resuspended in WB at a final concentration of 5×10^6 cells/mL. Efficient nuclear extraction was confirmed by Trypan Blue staining. In case of presence of nuclear aggregates, suspension was filtered through a 40 μm strainer before proceeding to next step. Concanavalin A beads were equilibrated in ice-cold binding buffer and mixed with the isolated nuclei (10 μl bead slurry per 5×10^5 cells). The nuclei-beads mixture was incubated for 10 minutes rotating at 4°C. Successful nuclei-bead binding was confirmed with Trypan Blue staining. Beads-coated nuclei were resuspended in WB at a final concentration of 5×10^6 cells/mL and 100 μL were aliquoted into each sample tube. While gently vortexing, 100 μL of 2× primary antibody mix was added to the corresponding sample tube. Tube samples were mixed and incubated in rotation overnight at 4°C. Next day, nuclei were washed twice with 1 mL of ice-cold WB and resuspended in 25 μL of ice-cold WB. While gently vortexing, 25 μL of 2× pAG-MNase mix (EpiCypher, Inc. catalog # 15-1016) was added to each tube and samples were incubated for 1 hour at 4°C. Nuclei were washed twice with WB and resuspended in 150 μL of ice cold WB. While gently vortexing, 3 μL of pre-chilled 100 μM $CaCl_2$ were added and after briefly flicking to mix, tubes were incubated for 30 minutes in a 0 °C water bath. After 30 minutes, the MNase was inactivated by adding 150 μL of 2× STOP buffer. To isolate the digested DNA though diffusion, samples were incubated 20 minutes at 37 °C with no shaking. Samples were centrifuged at 16,000 g for 5 minutes at 4°C and the resulting supernatant was treated with 3 μL of 10% SDS and 2.5 μL of 20 mg/mL proteinase K at 65 °C for 1 hour to overnight without shaking. DNA was purified with Qiagen MinElute PCR purification kit. Sequencing libraries were constructed using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB E7645L) following the manufacturer's manual. The libraries were sequenced at BGI.

CUT&RUN sequencing data analysis.

The CUT&RUN sequencing data were aligned by Cut&RunTools (bitbucket.org/qzhudfci/cutruntools/src/master/, version: 49ddd24)[68]. The reads were trimmed by Trimmomatic[69], then aligned to the mouse genome mm10 by Bowtie2 (bowtie-bio.sourceforge.net/bowtie2)[70]. The duplicate reads at same locations were removed by MarkDuplicates of Picard package (broadinstitute.github.io/picard, 2.23.8). The enrichment density of different samples were normalized to read counts per million mapped reads and same scale of the gene body by ngs.plot (github.com/shenlab-sinai/ngsplot, 2.63)[71]. Enrichment density was

normalized by the spike-ins and then the average profile figures of the gene body regions were generated by ngs.plot. Gene body regions were scanned for DEGs and genome-wide genes, and the peak intensity was extracted for PSC_N cells with control knockdown over NSD3 knockdown (Peaks), the differences between them calculated, and compared to the distribution of Peaks of DEGs and the genome-wide genes with two-sided robust t-test at the trim level 0.1. The intergenic region annotation was generated by Complement of BedTools[72], and then the average profile of the intergenic regions was visualized by ngs.plot.

ChIP-qPCR.

ChIP assays were performed as described[25]. Purified DNA was subject to quantitative real-time PCR analysis. Purified ChIP DNA was analyzed on LightCycler 480 (Roche) using SYBR Green Master Mix (Applied Biosystems) following the manufacturer's manual. All samples were normalized to input as control. The primer sequences are listed below:

Irgm1-p1 F: GTTGGGGTTGGAGTTGTGTC, *Irgm1*-p1 R: TTGCTGGGTCTGGAGGTTAG, *Irgm1*-p2 F: CTGAGCTGGGTGTATAGGCA, *Irgm1*-p2 R: CTTGTCTGTGTACCCCGTCT, *Irgm1*-p3 F: TAAGTCCCACAGCACCACAT, *Irgm1*-p3 R: AGCATCTTTGTGACTGGGGA, *Prkaa2*-p1 F: GTGCCAGTTTCTTAGCCCAG, *Prkaa2*-p1 R: ACACATGGTCAGACTCAGCA, *Prkaa2*-p2 F: TTTGTTTCACCAGCCCCTTG, *Prkaa2*-p2 R: AACGGACTGTAGCCTGGAAA, *Prkaa2*-p3 F: GACACACCTCCAAGCATGTC, *Prkaa2*-p3 R: GCTCTGTTCACATCCTCCCT, *Btg2* F: GAGTGGTATGAAAGGCGCAG, *Btg2* R: GGAGACTGGAGAGGAAACCC, *Gadd45g* F: CTGGGAATCTTTACCTGCGC, *Gadd45g* R: AGACCACTACCAGACAAGGC.

Statistics and Reproducibility.

All statistics used in this study are described when used in the appropriate section. For all the Western blot and autoradiography results, the reported data were reproduced three independent times (or more as indicated) and representative data are shown.

Extended Data
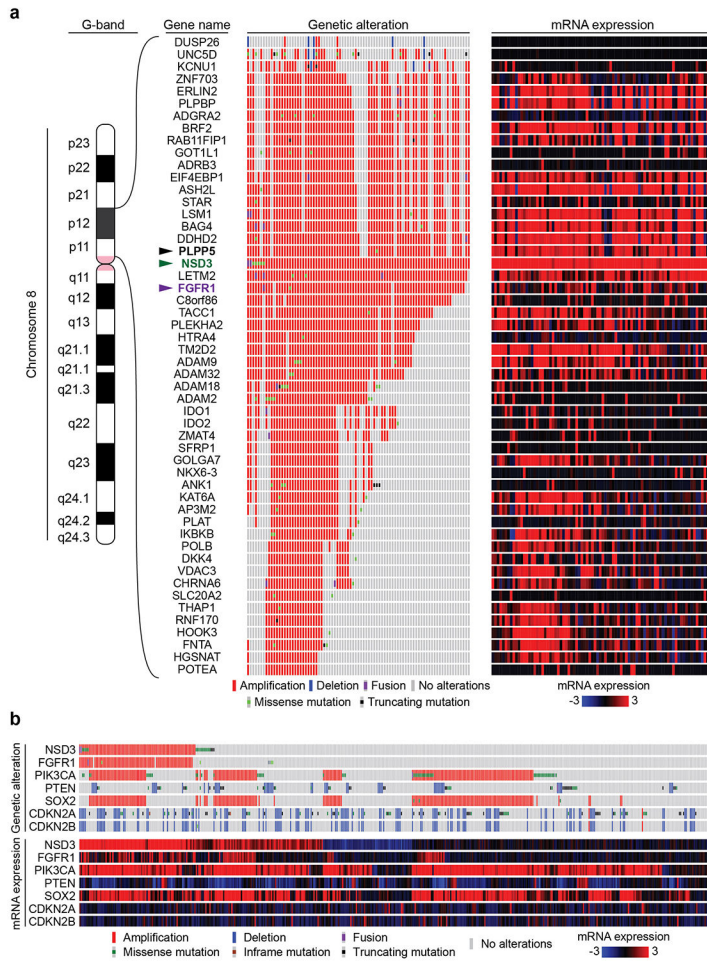
Extended Figure 1. Characterization of genetic drivers of LUSC tumorigenesis.

a, Genetic alterations and mRNA expression of genes in the 8p11-12 amplification genomic region of chromosome 8 from the 85 8p11-12 amplified human LUSC specimens in the TCGA dataset. **b**, Genetic alterations and mRNA expression of *NSD3, FGFR1, PIK3CA, PTEN, SOX2, CDKN2A, CDKN2B* from all 464 of the human LUSC specimens in the TCGA dataset (ID: phs000178) analyzed using cBioPortal.
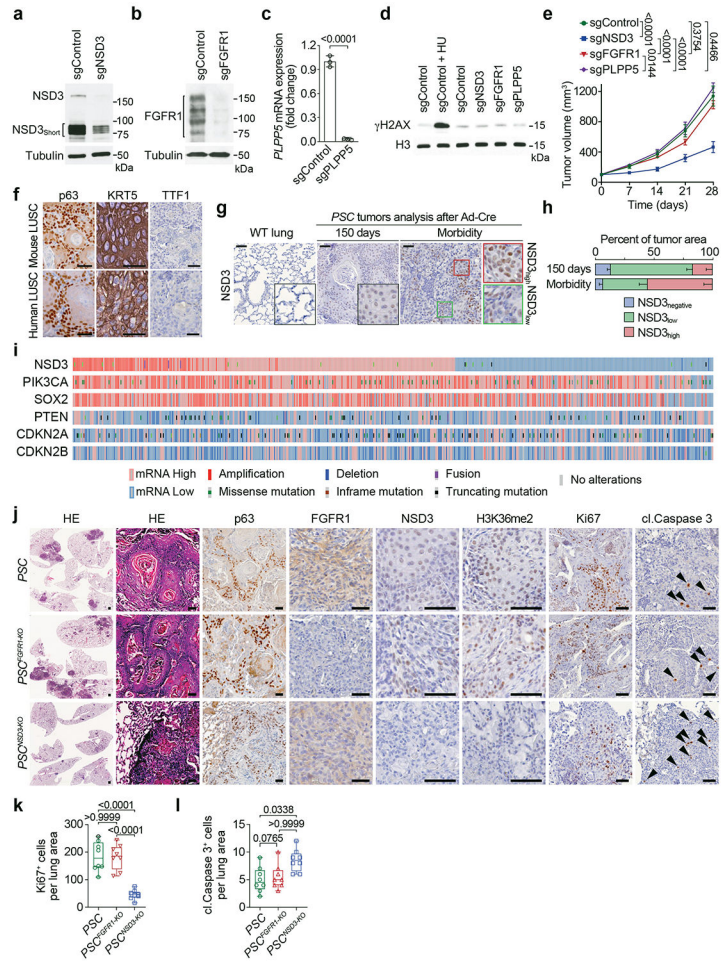
Extended Figure 2. NSD3 but not FGFR1 depletion inhibits LUSC tumorigenesis *in vivo*.
a, Western blot analysis with the indicated antibodies of whole-cell lysates from H520 cells expressing sgControl or sgNSD3. The bands representing NSD3 and NSD3$_{Short}$ are indicated. Tubulin was used as a loading control. **b**, Western blot analysis with the indicated antibodies of whole-cell lysates from H520 cells expressing sgControl or sgFGFR1. Tubulin was used as a loading control. **c**, Real-time (RT) qPCR analysis of *PLPP5* mRNA expression in H520 cells expressing sgControl or sgPLPP5. We note that several commercial antibodies against PLPP5 were tested but we could not detect any reproducible band close to the correct size (data not shown). The RT-qPCR data were normalized to *Actb* and presented as fold change relative to the sgControl sample. Error bars represent mean ± s.d. from three independent experiments. *P* value determined by two-tailed unpaired t-test. **d**, Treatment of the 8p11^{AMP}-positive H520 LUSC cell line with sgRNAs targeting the 8p11-amplified region does not cause an increase in phosphorylated H2AX (γH2AX). Western blot analysis with the indicated antibodies of H520 cells expressing Cas9 and the indicated sgRNAs. DNA damage-induced by hydroxyurea (HU) treatment was used as a positive control, H3 was used as a loading control. **e**, Depletion of NSD3, but not depletion of FGFR1 or PLPP5, attenuates xenograft growth of the 8p11^{AMP} H520 LUSC cells. Tumor volume quantification of the indicated H520 xenografts generated in immunocompromised mice (n

= 5 mice, for each treatment group). *P* values indicated determined by two-way ANOVA with Tukey's post hoc test. Data are represented as mean ± s.e.m. Data showing sgControl, sgNSD3 and sgFGFR1 are the same as from Figure 1b. **f**, Representative HE-stained sections and IHC staining of lung tissue from *PSC* mouse model lung tumor showing key diagnostic histological features of LUSC including positive staining for keratin and p63 and negative staining for TTF1 (representative of n = 8 samples for each group). Human LUSC samples are shown for comparison (representative of n = 8 samples for each group). Scale bars, 50 μm. **g**, Increased expression of NSD3 in lungs tracks with tumor progression. IHC analysis of NSD3 levels in wild type lung and tumor biopsies collected from the *PSC* model at 150 days and at the clinical endpoint when animals develop significant morbidity (representative of n = 8 samples for each group). Scale bars, 50 μm. **h**, Quantification of NSD3 levels in tumor biopsies as in (**g**) within tumor areas that have high-, low-intensity or negative staining for chromogen. Data are represented as mean ± s.e.m. **i**, NSD3 mRNA overexpression is common (~60%) and not limited to 8p11-12 amplified tumors. Genetic alterations and mRNA expression of the indicated genes in human LUSC patient samples from the TCGA dataset showing frequency of overexpression and alternations in NSD3 and known driver mutations *PIK3CA, PTEN, SOX2, CDKN2A, CDKN2B*. **j**, Representative HE-stained sections and IHC staining of lung tissue from *PSC* (control), *PSC*$^{FGFR1-KO}$ and *PSC*$^{NSD3-KO}$ mutant mice, (representative of n = 8 mice for each experimental group). Scale bars, 50 μm; arrowheads, positive cleaved Caspase 3 cells. **k**, **l**, Quantification of Ki67 (Ki67$^+$ positive cells) a marker of proliferation (**k**) and cleaved Caspase 3 (cl. Caspase 3$^+$ positive cells) a marker of apoptosis (**l**) in samples as in (**j**). In **k**, **l** and in all subsequent box plots, the line indicates the median, the box marks the 75th and 25th percentiles and the whiskers indicate the minimum and maximum values. All data points are shown. *P* values indicated determined by one-way ANOVA with Tukey's post hoc test (**k**, **l**).
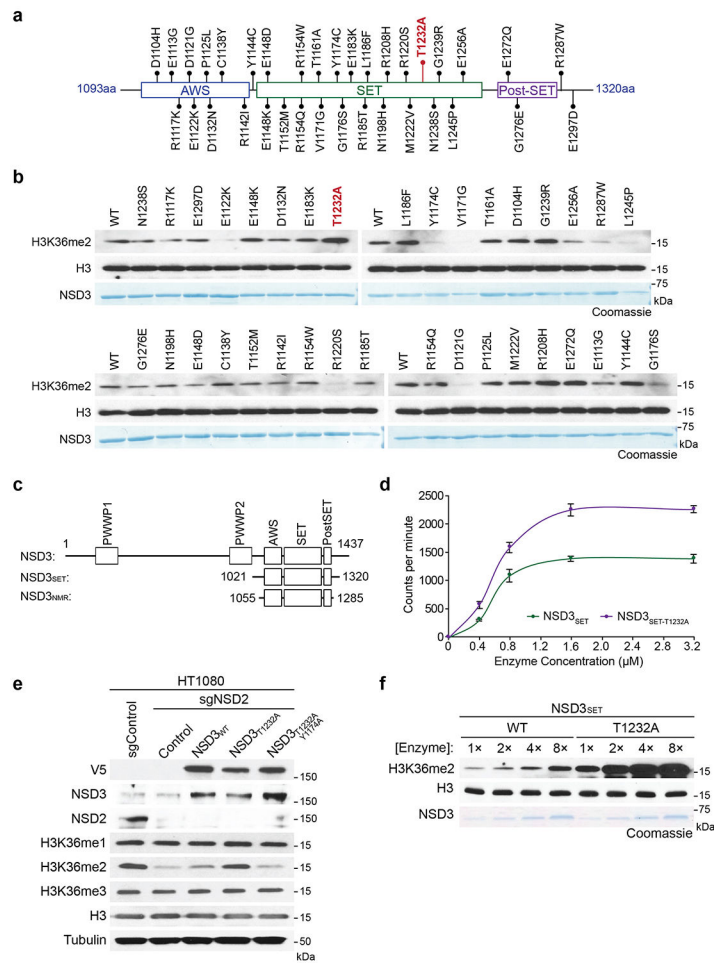
Extended Figure 3. Identification of the cancer-associated NSD3$_{T1232A}$ variant as a gain of function variant.

a, Schematic diagram of the distribution of the different NSD3 candidate mutants tested in (**b**). **b**, *In vitro* methylation assay as in Figure 2b on rNuc with recombinant wild-type NSD3$_{SET}$ or the indicated mutants (1-35) with non-radiolabeled SAM (the T1232A mutant is indicated in red). For each screening panel set, the top of the panel shows the Western blot and the bottom panel, Coomassie blue stain of the NSD3 proteins. **c**, Schematic diagram of NSD3 fragments used in this study. **d**, *In vitro* methylation reactions with rNucs and different concentrations of recombinant wild-type NSD3$_{SET}$ or NSD3$_{SET-T1232A}$ with ^{3}H-SAM. Methylation intensity was measured by liquid scintillation counting. Error bars represent mean ± s.d. from three independent experiments. **e**, Western blot analysis with the indicated antibodies of whole cell extracts from NSD2-deficient HT1080 cells (to deplete H3K36me2 compared to control HT1080 cells (lanes 1 and 2)) overexpressing the indicated NSD3 variants. Total H3 was used as a loading control. **f**, Methylation assays as in Figure 2b with rNuc with linker DNA (187 bp) and increasing concentrations of enzyme as indicated. Western blot shows H3K36me2 generation, H3 is shown as a loading control. Bottom panel, Coomassie blue stain of NSD3 proteins.
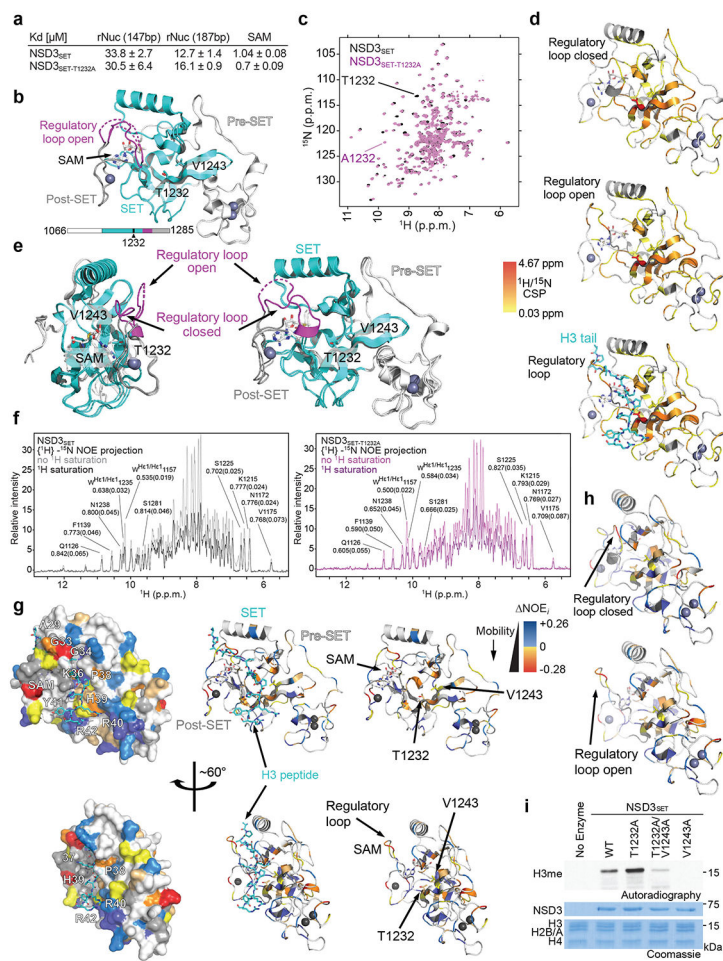
Extended Figure 4. Biophysical characterization of T1232A substitution on the NSD3$_{SET}$ domain.

a, Kd values with corresponding errors being s.d. are shown for binding studies of NSD3$_{SET}$ and NSD3$_{SET-T1232A}$ to recombinant nucleosomes (rNuc) reconstituted on 147bp and 187bp 601 Widom DNA as indicated (determined by microscale thermophoresis (MST)) and the cofactor SAM (determined by ITC) (see Supplementary Table 1; see Methods). Error bars represent mean ± s.d. from three independent experiments. b, The topology of NSD3SET domain segments of Pre-SET (white-gray), SET (cyan), Post-SET (dark gray) and regulatory loop (magenta) marked onto the ribbon representations of existing NSD3SET domain crystal structures (PDB 5UPD, 6CEN). The missing residues in the tip of the regulatory loop, not modelled into the electron densities, are denoted with dotted lines. The zinc(II) ions are depicted as gray spheres, SAM cofactor with sticks. Residues T1232 and V1243 are marked with their side-chains. Bottom panel: Schematic representation of the primary sequence with indicated domains and the location of the T1232A substitution. **c**, The overlay of the 2D [^{1}H, ^{15}N] TROSY-HSQC (Transverse Relaxation Optimized SpectroscopY - Heteronuclear Single Quantum Coherence) spectra of 250 μM U-[^{15}N] NSD3SET wild-type (dark gray) and NSD3$_{SET-T1232A}$ (magenta) at pH 7.5 and 25°C. The ^{1}H/^{15}N backbone amide cross-peaks of T1232 in NSD3$_{SET}$ wild-type and A1232 of NSD3$_{SETT1232A}$ mutant are marked. **d**, The backbone amide chemical shift perturbations (CSP$_i$ = [($\delta_{H,i}$)^2+0.25·($\delta_{N,i}$)^2]$^{0.5}$)

between NSD3$_{SET}$ and NSD3$_{SET-T1232A}$ are mapped onto the ribbon representations of protein structures; left: the NSD3$_{SET}$ with regulatory loop in closed (PDB 5UPD), center: in open (based on the NSD2$_{SET}$ crystal structure (PDB 5LSU) modelled with SwissModel) and right: the open conformation with the H3.1 A^{29}-R^{42} substrate docked (see Methods). The T1232 and V1243 residues are marked with their side-chains. Residues are colored as follows: 0.03 CSP$_i$ < 0.05 (yellow), 0.05 CSP$_i$ < 0.11 (light orange), 0.11 CSP$_i$ < 0.75 ppm (orange), CSP$_i$ > 4.60 (red), prolines and residues missing amide assignments are marked gray. **e**, The overlay of NSD3$_{SET}$ domains with the regulatory loop in open and closed conformations. The open state is the crystal structure of NSD3$_{SET}$, while the closed state is based on the NSD2$_{SET}$ crystal structure (PDB 5LSU) modelled with SwissModel. The protein sequence graph and coloring as in 4h. **f**, The projections of 2D TROSY-type heteronuclear {^{1}H}-^{15}N nuclear Overhauser effect (NOE) experiments collected for 370 μM U-[^{15}N] NSD3SET wild-type (left panel) and NSD3$_{SET-T1232A}$ (right panel) at pH 7.5 and 25°C with several residues marked. (Experimental details in Methods). The higher NOE ratio value the more rigid the N-H vector, meaning less dynamic motion, whereas the lower value means more dynamic, indicative of less restricted motions. The tryptophan imidazole side-chain ^{1}H/^{15}N correlations are also detected in this experiment, H^{e1}/N^{e1} from W1235 and W1157. **g**, The observed changes in {^{1}H}-^{15}N NOE values (NOE$_i$) reporting on ps-ns polypeptide main-chain mobility differences between NSD3$_{SET}$ and NSD3$_{SET-T1232A}$ are plotted on the 3D static structures of NSD3SET (PDB: 6CEN, left column represented as protein surface, middle column ribbon with H3 peptide, right column ribbon with no substrate). Key features are indicated. The light blue to dark blue indicates the decreased fast dynamics after T1232A substitution, and yellow to red increased dynamics on ps-ns time scale. **h**, The T1232A substitution enhances the mobility of the regulatory loop. The observed changes in heteronuclear {^{1}H}-^{15}N nuclear Overhauser effect (NOE) mapped on the ribbon representations of the open and closed states of the regulatory loop within the NSD3$_{SET}$ and in Figure 2e. The structures are the same as in (e). **i**, Methylation assays as in Figure 2a with rNuc(187bp) and the indicated NSD3 variants.
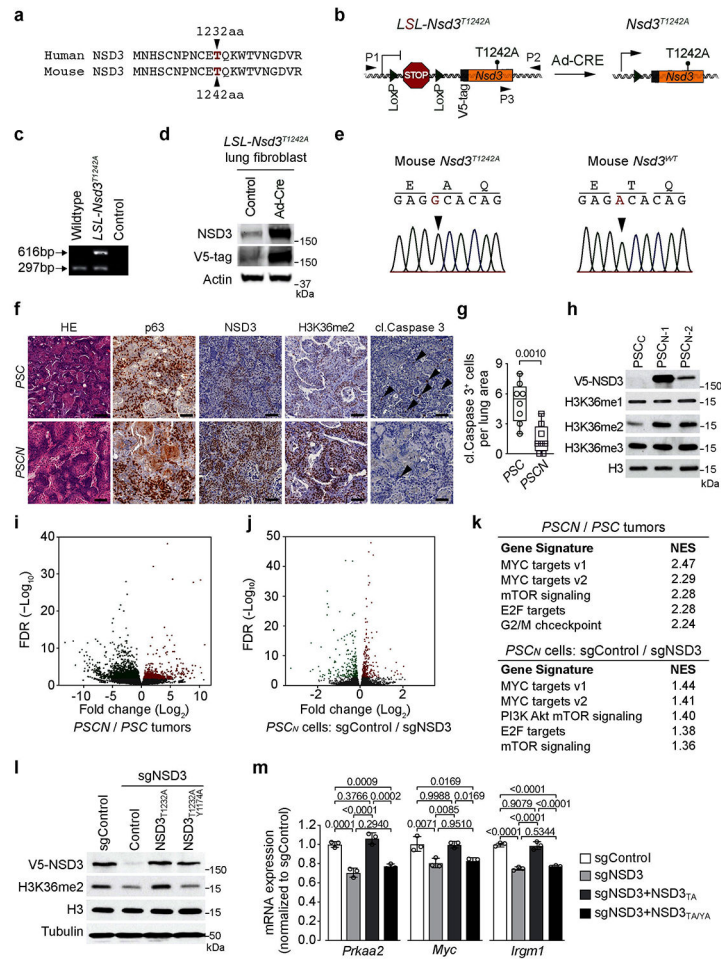
Extended Figure 5. Generation of *PSCN* LUSC mouse model and NSD3 coordination of an oncogenic gene expression program in LUSC.

a, Alignment of NSD3 human and mouse residues spanning human T1232 indicates that human NSD3 T1232 corresponds with murine NSD3 T1242. **b**, Schematic of the *LSL-V5-Nsd3^{T1242A}* conditional allele. In the presence of Cre recombinase, a translational stop cassette flanked by *LoxP* recombination sites is deleted to enable *V5-Nsd3^{T1242A}* expression. P1 and P2 indicate location of genotyping primers. **c**, Confirmation of successful generation of *LSL-Nsd3^{T1242A}* allele by PCR (P1+P2+P3 primers) on DNA isolated from mouse tail biopsies from indicated mouse genotypes, expected products sizes are marked (upper panel). **d**, Western blots with the indicated antibodies (V5 antibody detects V5-tagged NSD3$_{T1242A}$) of lysates from *LSL-V5-Nsd3^{T1242A}* lung fibroblast transduced *ex vivo* with Ad-Cre or vehicle (control). Actin is shown as a loading control. **e**, Left panel: Sanger sequencing confirmation *Nsd3^{T1242A}* mutation is present in generated conditional *LSL-V5-Nsd3^{T1242A}* mutant mouse. Wild-type NSD3 sequence from control animals is shown on the right panel. **f**, Representative HE-stained and IHC staining with indicated antibodies of lung tissue sections from *PSC* and *PSCN* mutant mice, (representative of n = 8 mice for each group). Scale bars, 50 μm; arrowheads, positive cleaved Caspase 3 cells. **g**, Quantification of cleaved Caspase 3 (cl. Caspase 3^{+} positive cells) a marker of apoptosis in samples as in (**f**). *P* values indicated determined by two-tailed unpaired t-test. **h**, Western blots with indicated

antibodies of whole-cell lysates from PSC_C or two different PSC_N cell line. Total H3 was used as a loading control. **i**, Volcano plot of RNA-seq comparison between *PSCN* and *PSC* tumor biopsies (three independent biological replicates for each condition) results in higher expression of 891 genes shown in red (fold change log2 −0.5 and *P* < 0.05 by Wald test) and decreasing expression of 1839 genes shown in green (fold change log2 0.5 and *P* < 0.05 by Wald test). False discovery rate (FDR) values are provided. **j**, Volcano plot of RNA-seq comparison of PSC_N cells ± sgNSD3 (three biological replicates for each condition) causes decreasing expression of 234 genes shown in red (adjusted *P* < 0.001 by likelihood ratio test) and increasing expression of 229 genes shown in green (adjusted *P* < 0.001 by likelihood ratio test). False discovery rate (FDR) values are provided. **k**, Top hallmark gene sets identified in the GSEA analysis of datasets from (**i-j**) shows high overlap. Normalized enrichment scores (NES) provided (detailed statistics description in Methods). **l**, Reconstitution of NSD3-deficient PSC_N cells with NSD3 and derivatives. Western blot analysis with the indicated antibodies of PSC_N whole cell lysates ± sgNSD3 and complemented with the indicated CRISPR-resistant NSD3 derivatives. H3 and tubulin were used as loading controls. **m**, Quantitative real-time (RT)-PCR analysis of the transcript levels of the indicated genes in the cells as described in (**l**). The RT-qPCR data for each gene were normalized to *Actb* and presented as fold change relative to the sgControl sample. Error bars represent mean ± s.d. from three independent experiments. *P* values indicated determined by one-way ANOVA with Tukey's post hoc test.
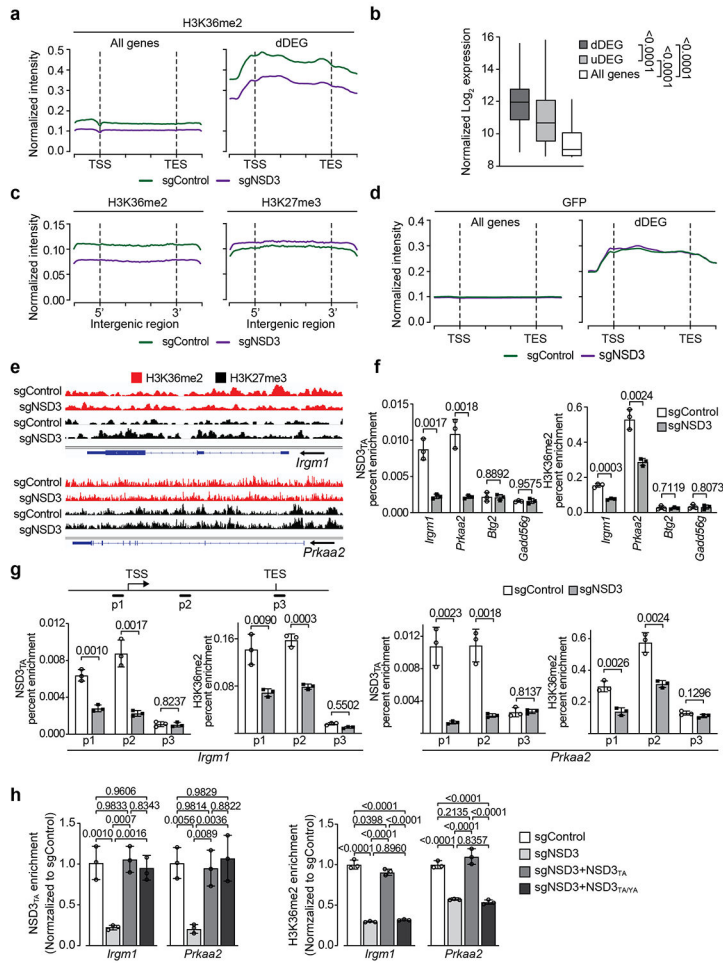
Extended Figure 6. NSD3$_{T1242A}$-mediated H3K36me2 in LUSC oncogenic reprogramming.
a, Independent CUT&RUN replicate for H3K36me2 as in Figure 3i. **b**, Normalized gene expression levels from the indicated groups from PSC$_N$ cell line RNA-seq datasets. dDEG (downregulated DEGs) and uDEG (upregulated DEGs) from Extended Data Fig. 5j, and all genes. P values indicated determined by two-tailed robust t-test (detailed statistics description in Methods). dDEG $n = 234$, uDEG $n = 229$, all genes $n = 16091$. Boxes: 25th to 75th percentile, whiskers: min. to max., center: median. **c**, CUT&RUN profile for H3K36me2 and H3K27me3 in PSC$_N$ cells ± sgNSD3 over average of intergenic regions on a genome-wide scale. **d**, CUT&RUN profile for a monoclonal IgG (against GFP) as in Figure 3i for all genes and dDEG as indicated in PSCN cells ±sgNSD3 is shown as a negative control. **e**, Genome browser view of CUT&RUN signals for H3K36me2 and H3K27me3 on the indicated genes and conditions. Arrow indicates the direction of gene transcription. **f**, Loss of occupancy of NSD3$_{T1242A}$ and H3K36me2 at the indicated NSD3-target genes (*Irgm1* and *Prkaa2*) upon NSD3 depletion. *Btg2* and *Gadd45g* are not NSD3-target genes and shown as control regions that do not change in response to NSD3 depletion. ChIP-qPCR analysis of V5-NSD3$_{T1242A}$ (top panel) and H3K36me2 (bottom panel) in the body of the indicated genes. The data were plotted as percent enrichment relative to input. Error bars represent mean ± s.d. from three independent experiments. P values indicated

determined by two-tailed unpaired t-test. **g**, $NSD3_{T1242A}$ and H3K36me2 occupy the promoter and gene body regions of target genes *Irgm1* and *Prkaa2*. Top panel, schematic of general gene structure and site of primers for *Irgm1* (left panel) and *Prkaa2* (right panel) gene loci. ChIP-qPCR analysis of V5-$NSD3_{T1242A}$ (left panel) and H3K36me2 (right panel) in promoter (p1), gene body (p2) and transcription end site (TES; p3) regions of target genes. TSS: transcription start site. The data were plotted as percent enrichment relative to input. Error bars represent mean ± s.d. from three independent experiments. *P* values indicated determined by two-tailed unpaired t-test. **h**, ChIP-qPCR analysis of $NSD3_{T1242A}$ (left panel) and H3K36me2 (right panel) occupancy at the gene body regions of *Irgm1* and *Prkaa2* gene loci in reconstituted cells as described in Extended Data Fig. 5l. The enrichment was normalized to the sgControl sample and presented as fold change relative to the sgControl sample. Error bars represent mean ± s.d. from three independent experiments. *P* values indicated determined by one-way ANOVA with Tukey's post hoc test.
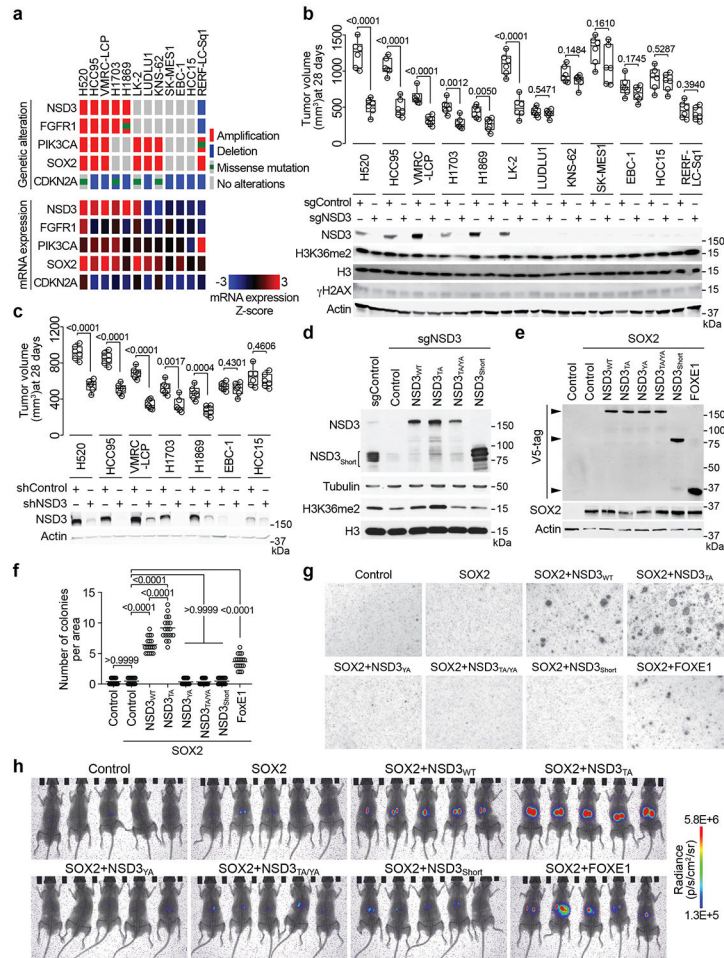
Extended Figure 7. NSD3 depletion attenuates growth of 8p11-amplified and NSD3-overexpressing human LUSC cell lines and ectopic expression of NSD3 cooperates with SOX2 to transform tracheobronchial epithelial cells.

a, Genetic alterations and mRNA expression of LUSC-associated genes in human LUSC cell lines used in this study. **b**, NSD3 depletion by CRISPR/Cas9 attenuates xenograft tumor

growth of 8p11^{AMP} and NSD3-overexpressing LUSC cell lines. Upper panel: Tumor volume quantification of human LUSC cell line xenografts described in (Figure 4a and Extended Data Fig 7a) treated with sgControl or sgNSD3 as indicated and grown in immunocompromised mice at 28 days after implantation (n = 5 mice, for each treatment group). P values indicated determined by one-way ANOVA with Tukey's post hoc test. Data are represented as mean ± s.e.m. Lower panel: Western blot analysis with the indicated antibodies of whole cell lysates from cells in upper panel. γH2AX levels are shown to assess whether sgNSD3 expression induces non-specific DNA damage. H3 and Actin are used as loading controls. (upper panel). **c**, NSD3 depletion by shRNA attenuates xenograft tumor growth of 8p11^{AMP} positive LUSC cell lines. Upper panel: Tumor volume quantification of the indicated 8p11^{AMP}-positive and two control 8p11^{AMP}-negative human LUSC cell line xenografts described as in (**a**) treated with shControl or shNSD3 as indicated and grown in immunocompromised mice at 28 days after implantation (n = 5 mice, for each treatment group). P values indicated determined by one-way ANOVA with Tukey's post hoc test. Data are represented as mean ± s.e.m. Lower panel: Western blot analysis with the indicated antibodies of whole cell lysates from cells in upper panel. Actin is used as a loading control. **d**, NSD3-deficient H520 cells reconstituted with the indicated V5-tagged CRISPR-resistant NSD3 derivatives: NSD3$_{T1232A}$ (NSD3$_{TA}$) NSD3$_{T1232A/Y1174A}$ (NSD3$_{TA/YA}$). Western blots of H520 lysates with indicated antibodies H3 and tubulin were used as loading controls (see Figure 4b). **e**, Western blots of whole cell lysate of AALE cells used in transformation assays for Figure 4c with ectopic expression of SOX2 and the indicated V5-tagged constructs (NSD3$_{WT}$, NSD3$_{TA}$, NSD3$_{YA}$, NSD3$_{TA/YA}$, NSD3$_{Short}$ or FOXE1). **f**, Quantification of soft agar colony formation for AALE tracheobronchial epithelial cells with ectopic expression of SOX2 and NSD3$_{WT}$, NSD3$_{TA}$, NSD3$_{YA}$, NSD3$_{TA/YA}$, NSD3$_{Short}$ or FOXE1 as in (**e**). Data are represented as mean ± s.e.m. of three technical replicates in two independent experiments. **g**, Representative soft-agar images from AALE transformation assays in (**f**). **h**, *In vivo* AALE transformation assay images from Figure 4c. Optical overlay of bioluminescent signal with X-ray image of mice grafted under the renal capsule with AALE cells expressing plasmids as in (**e**) and AkaLuc after substrate (AkaLumine-HCl; see Methods) administration (n = 5 for each condition). The color bar indicates the total bioluminescence radiance (photons/s/cm^2/sr).
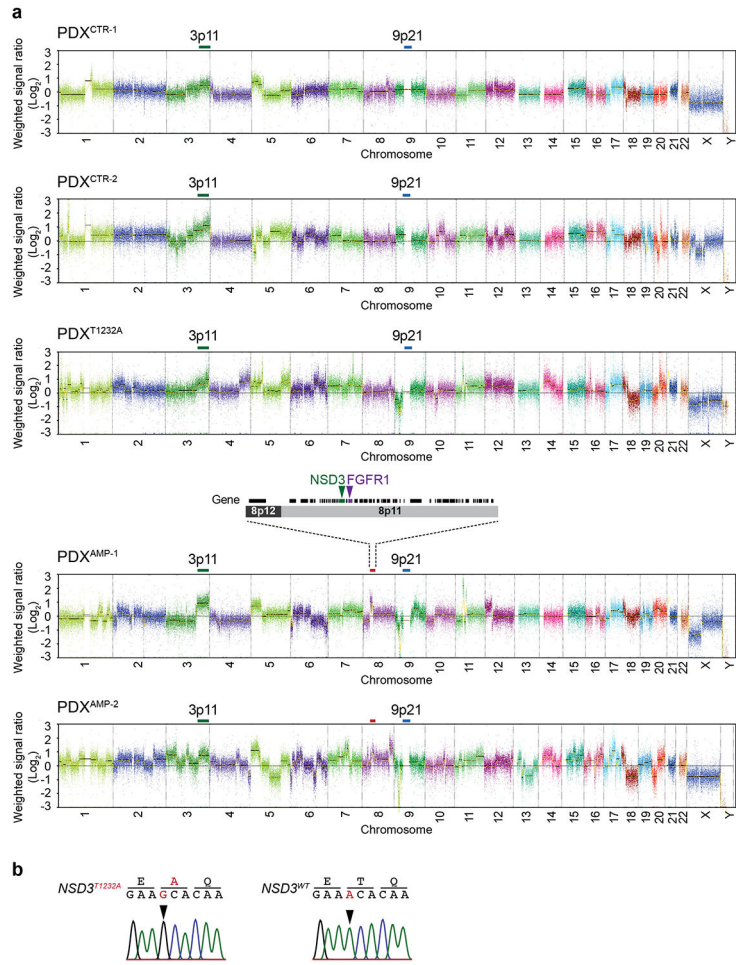
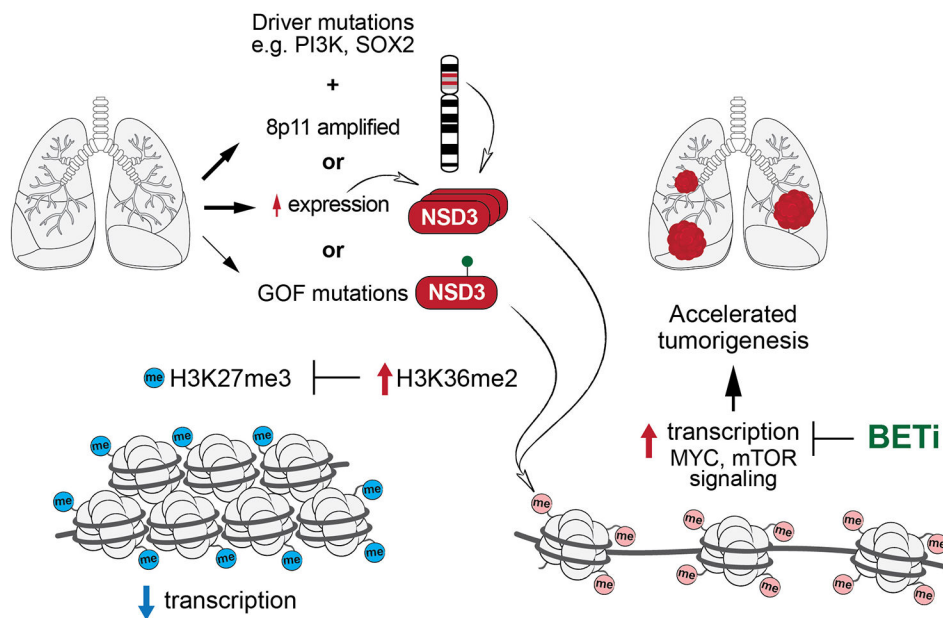**Extended Figure 8. Genomic characterization of patient-derived xenografts (PDX) from primary LUSC.**

a, Distribution of whole-genome copy number alterations in LUSC PDXs utilized in this study by OncoScan (detailed description in Methods). Green lines: 3p11 amplification (including *PIK3CA* and *SOX2*); red lines: marks, 8p11-12 (e.g. *NSD3* and *FGFR1* as shown); blue lines: biallelic deletion in 9p21.3 (including *CDKN2A/B*). **b**, Sanger sequencing analysis of NSD3 SET domain mutation status in collection of 37 LUSC PDX identified *NSD3*[T1232A] mutation in one of the PDX samples (PDX[T1232A]) and *Nsd3* wildtype in other analyzed PDXs.
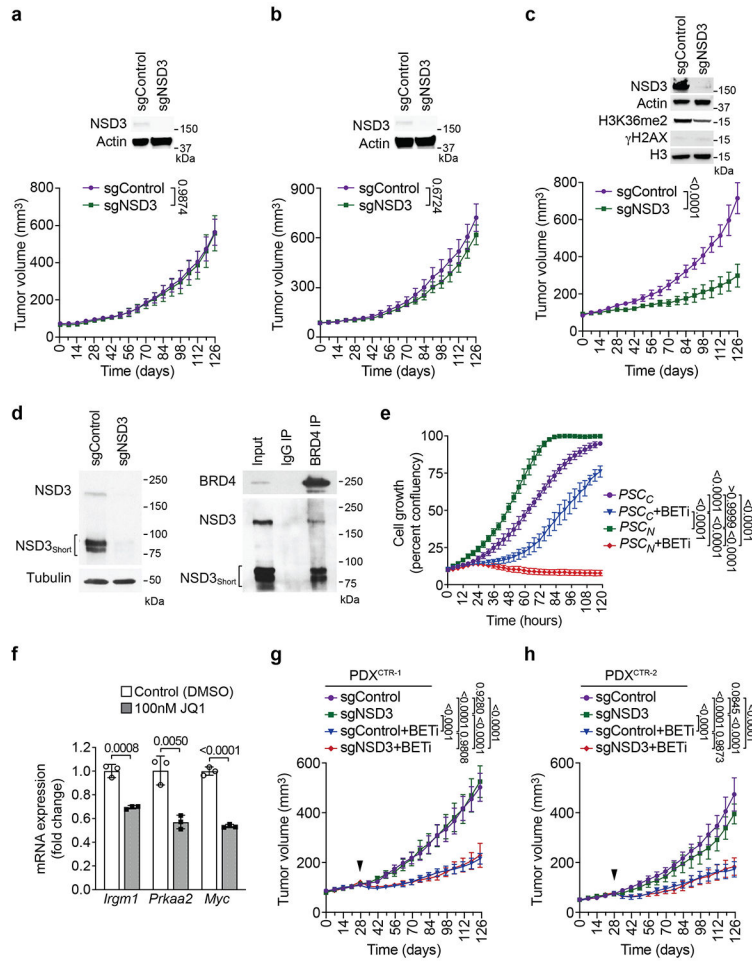
Extended Figure 9. NSD3-dependency renders PDXs from LUSC patients therapeutically vulnerable to BETi.

a-c, Tumor volume quantification of PDX^{CTR-1} (**a**), PDX^{CTR-2} (**b**) and PDX^{AMP-2} (**c**) growing in immunocompromised mice (n = 5 mice, for each treatment group). Western blots with the indicated antibodies of lysates from the PDX ± sgNSD3 is shown. Actin is shown as a loading control. **d**, BRD4 interacts with NSD3 in PSC$_N$ cells. Left panel: Western blot of whole cell lysates from PSC$_N$ cells ± sgNSD3 to show the relative position of NSD3 and NSD3$_{Short}$ in PSC$_N$ cells. Tubulin used as a loading control. Right panel: BRD4 interacts with both NSD3 and NSD3$_{Short}$ isoforms in PSC$_N$ cells. Co-immunoprecipitation (IP) experiments in PSC$_N$ cells with indicated antibodies for IPs and western analyses. Input, PSC$_N$ cell nuclear extract. **e**, Proliferation assay of PSC$_C$ and PSC$_N$ treated with vehicle control or 20nM AZD5153 (BETi). Data represent mean ± s.e.m. of three technical replicates in two independent experiments. **f**, Treatment of PSC$_N$ cells with the BETi JQ1 inhibits expression of the indicated NSD3 target genes. Real-time (RT) qPCR analysis of the indicated mRNAs from PSC$_N$ cells treated with DMSO or 100nM of JQ1. The RT-qPCR data for each target gene were normalized to *Actb* and presented as fold change relative to DMSO treated sample. Error bars represent mean ± s.d. from three independent experiments. *P* values were determined by two-tailed unpaired t-test. **g-h**, Tumor volume quantification of PDX^{CTR-1} (**g**) and PDX^{CTR-2} (**h**) treated with BETi or vehicle (control).

Arrowhead indicates start of the treatment. *P* values were determined by two-tailed unpaired t-test. (**a-c**, **f**) or two-way ANOVA with Tukey's post hoc test (**e**, **g-h**). Data are represented as mean ± s.e.m. (**a-c**, **e**, **f-h**).

Extended Figure 10. Model of the role for elevated NSD3 H3K36 methylation activity in LUSC pathogenesis.

LUSC is characterized by a number of driver mutations such as increased expression of PI3K and SOX2, PI3K activating mutations, deletion of PTEN and other deletions and alterations. One of the more common genetic alterations is amplification of the 8p11-12 genomic region (~20% of LUSC patients). While increased expression of FGFR1 is postulated to be the causative mutation of the 8p11-12 amplicon, our work implicates amplification of the neighboring gene *NSD3* as the main driving alteration. As shown, *NSD3* amplification leads to increased NSD3 expression and hence increased synthesis of H3K36me2, which works in concert with other LUSC driver mutations to promote LUSC pathogenesis. In addition, as shown in Extended Data Fig. 2i, NSD3 overexpression is detected in 60% of LUSC patients, thus NSD3 is frequently highly expressed in LUSC that does not harbor the 8p11-12 amplifications. We also describe $NSD3_{T1232A}$ as a gain-of-function (GOF) variant, that while far less common than the 8p11-12 amplicon and NSD3 overexpression, is an alteration present in human LUSC. The $NSD3_{T1232A}$ variant, due to enhanced catalytic behavior, functionally acts like amplified or overexpressed *NSD3* in increasing H3K36me2 synthesis and cooperating with other LUSC driver mutations to accelerate tumorigenesis. The increase in H3K36me2 by NSD3 overexpression or $NSD3_{T1232A}$ reprograms the chromatin landscape, blocking synthesis of H3K27me3 and increasing H3K36me2, which stimulates transcription of key oncogenic targets including genes involved in mTOR signaling and MYC-associated pathways. We speculate that NSD3-regulated tumors become addicted to H3K36me2-driven transcriptional activation, rendering these tumors particularly vulnerable to BETi as NSD3 and BRD4 interact and cooperate in transcription. This hypersensitivity could potentially be exploited clinically due

to an expanded therapeutic window for BETi and by using these drugs as targeted therapy for the tens of thousands of patients that are 8p11-12 positive. Not shown in the model, NSD3 is overexpressed in many LUSC samples without an underlying known alteration and we speculate that in such cases, increased NSD3 would contribute to tumorigenesis in a similar fashion as the scenarios described above in which NSD3 is hyperactive or overexpressed due to amplification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data Availability

The CUT&RUN and RNA-seq data from tumor-derived cells are deposited in the Gene Expression Omnibus (GEO) database under accession number GSE149482, RNA-seq data from tumor biopsies is deposited in the GEO database under accession number GSE149272. Full blot data for all the western blots were shown in Supplementary Figure 1 and all source data is available in Supplementary Data.

References:

1. Balsara BR et al. Comparative genomic hybridization analysis detects frequent, often high-level, overrepresentation of DNA sequences at 3q, 5p, 7p, and 8q in human non-small cell lung carcinomas. Cancer Res 57, 2116–2120 (1997). [PubMed: 9187106]

2. Tonon G et al. High-resolution genomic profiles of human lung cancer. Proc Natl Acad Sci U S A 102, 9625–9630, doi:10.1073/pnas.0504126102 (2005). [PubMed: 15983384]

3. Rooney C et al. Characterization of FGFR1 Locus in sqNSCLC Reveals a Broad and Heterogeneous Amplicon. PLoS One 11, e0149628, doi:10.1371/journal.pone.0149628 (2016). [PubMed: 26905262]

4. Weiss J et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. Sci Transl Med 2, 62ra93, doi:10.1126/scitranslmed.3001451 (2010).

5. Lim SH et al. Efficacy and safety of dovitinib in pretreated patients with advanced squamous non-small cell lung cancer with FGFR1 amplification: A single-arm, phase 2 study. Cancer 122, 3024–3031, doi:10.1002/cncr.30135 (2016). [PubMed: 27315356]

6. Yang ZQ, Liu G, Bollig-Fischer A, Giroux CN & Ethier SP Transforming properties of 8p11-12 amplified genes in human breast cancer. Cancer Res 70, 8487–8497, doi:10.1158/0008-5472.CAN-10-1013 (2010). [PubMed: 20940404]

7. Turner-Ivey B et al. Development of mammary hyperplasia, dysplasia, and invasive ductal carcinoma in transgenic mice expressing the 8p11 amplicon oncogene NSD3. Breast Cancer Res Treat 164, 349–358, doi:10.1007/s10549-017-4258-9 (2017). [PubMed: 28484924]

8. Travis WD Lung Cancer Pathology: Current Concepts. Clin Chest Med 41, 67–85, doi:10.1016/j.ccm.2019.11.001 (2020). [PubMed: 32008630]

9. Husmann D & Gozani O Histone lysine methyltransferases in biology and disease. Nat Struct Mol Biol 26, 880–889, doi:10.1038/s41594-019-0298-7 (2019). [PubMed: 31582846]

10. Landau DA et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 152, 714–726, doi:10.1016/j.cell.2013.01.019 (2013). [PubMed: 23415222]

11. Qiao Q et al. The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. J Biol Chem 286, 8361–8368, doi:10.1074/jbc.M110.204115 (2011). [PubMed: 21196496]

12. Graham SE, Tweedy SE & Carlson HA Dynamic behavior of the post-SET loop region of NSD1: Implications for histone binding and drug development. Protein Sci 25, 1021–1029, doi:10.1002/pro.2912 (2016). [PubMed: 26940890]

13. Yang S et al. Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. Genes Dev 30, 1611–1616, doi:10.1101/gad.284323.116 (2016). [PubMed: 27474439]

14. Skene PJ & Henikoff S An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife 6, doi:10.7554/eLife.21856 (2017).

15. Munoz DM et al. CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. Cancer Discov 6, 900–913, doi:10.1158/2159-8290.CD-16-0178 (2016). [PubMed: 27260157]

16. Bass AJ et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet 41, 1238–1242, doi:10.1038/ng.465 (2009). [PubMed: 19801978]

17. Zhang Q et al. Structural Mechanism of Transcriptional Regulator NSD3 Recognition by the ET Domain of BRD4. Structure 24, 1201–1208, doi:10.1016/j.str.2016.04.019 (2016). [PubMed: 27291650]

18. Shen C et al. NSD3-Short Is an Adaptor Protein that Couples BRD4 to the CHD8 Chromatin Remodeler. Mol Cell 60, 847–859, doi:10.1016/j.molcel.2015.10.033 (2015). [PubMed: 26626481]

19. Bradbury RH et al. Optimization of a Series of Bivalent Triazolopyridazine Based Bromodomain and Extraterminal Inhibitors: The Discovery of (3R)-4-[2-[4-[1-(3-Methoxy-[1,2,4]triazolo[4,3-b]pyridazin-6-yl)-4-piperidyl]phen oxy]ethyl]-1,3-dimethyl-piperazin-2-one (AZD5153). J Med Chem 59, 7801–7817, doi:10.1021/acs.jmedchem.6b00070 (2016). [PubMed: 27528113]

20. Siegel RL, Miller KD & Jemal A Cancer statistics, 2019. CA Cancer J Clin 69, 7–34, doi:10.3322/caac.21551 (2019). [PubMed: 30620402]

21. Cochran AG, Conery AR & Sims RJ 3rd. Bromodomains: a new target class for drug development. Nat Rev Drug Discov 18, 609–628, doi:10.1038/s41573-019-0030-7 (2019). [PubMed: 31273347]

22. Lin KH et al. Using antagonistic pleiotropy to design a chemotherapy-induced evolutionary trap to target drug resistance in cancer. Nat Genet 52, 408–417, doi:10.1038/s41588-020-0590-9 (2020). [PubMed: 32203462]

23. Bass AJ et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet 41, 1238–1242, doi:10.1038/ng.465 (2009). [PubMed: 19801978]

24. Su Y et al. Novel NanoLuc substrates enable bright two-population bioluminescence imaging in animals. Nat Methods 17, 852–860, doi:10.1038/s41592-020-0889-6 (2020). [PubMed: 32661427]

25. Kuo AJ et al. NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming. Mol Cell 44, 609–620, doi:S1097-2765(11)00813-6[pii]10.1016/j.molcel.2011.08.042 (2011). [PubMed: 22099308]

26. Lowary PT & Widom J New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol 276, 19–42, doi:10.1006/jmbi.1997.1494 (1998). [PubMed: 9514715]

27. Cheema MS & Ausio J Analytical Ultracentrifuge Analysis of Nucleosomes Assembled from Recombinant, Acid-Extracted, HPLC-Purified Histones. Methods Mol Biol 1528, 75–95, doi:10.1007/978-1-4939-6630-1_6 (2017). [PubMed: 27854017]

28. Luger K, Rechsteiner TJ & Richmond TJ Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol Biol 119, 1–16, doi:10.1385/1-59259-681-9:1 (1999). [PubMed: 10804500]

29. Shi X et al. Modulation of p53 function by SET8-mediated methylation at lysine 382. Mol Cell 27, 636–646, doi:10.1016/j.molcel.2007.07.012 (2007). [PubMed: 17707234]

30. Chen S et al. The PZP Domain of AF10 Senses Unmodified H3K27 to Regulate DOT1L-Mediated Methylation of H3K79. Mol Cell 60, 319–327, doi:10.1016/j.molcel.2015.08.019 (2015). [PubMed: 26439302]

31. Mazur PK et al. Combined inhibition of BET family proteins and histone deacetylases as a potential epigenetics-based therapy for pancreatic ductal adenocarcinoma. Nat Med 21, 1163–1171, doi:10.1038/nm.3952 (2015). [PubMed: 26390243]

32. Edelman BL & Redente EF Isolation and Characterization of Mouse Fibroblasts. Methods Mol Biol 1809, 59–67, doi:10.1007/978-1-4939-8570-8_5 (2018). [PubMed: 29987782]

33. Liu S et al. METTL13 Methylation of eEF1A Increases Translational Output to Promote Tumorigenesis. Cell 176, 491–504 e421, doi:10.1016/j.cell.2018.11.038 (2019). [PubMed: 30612740]

34. Adams JR et al. Cooperation between Pik3ca and p53 mutations in mouse mammary tumor formation. Cancer Res 71, 2706–2717, doi:10.1158/0008-5472.CAN-10-0738 (2011). [PubMed: 21324922]

35. Ferone G et al. SOX2 Is the Determining Oncogenic Switch in Promoting Lung Squamous Cell Carcinoma from Different Cells of Origin. Cancer Cell 30, 519–532, doi:10.1016/j.ccell.2016.09.001 (2016). [PubMed: 27728803]

36. Krimpenfort P et al. p15Ink4b is a critical tumour suppressor in the absence of p16Ink4a. Nature 448, 943–946, doi:10.1038/nature06084 (2007). [PubMed: 17713536]

37. Hoch RV & Soriano P Context-specific requirements for Fgfr1 signaling through Frs2 and Frs3 during mouse development. Development 133, 663–673, doi:10.1242/dev.02242 (2006). [PubMed: 16421190]

38. Skarnes WC et al. A conditional knockout resource for the genome-wide study of mouse gene function. Nature 474, 337–342, doi:10.1038/nature10163 (2011). [PubMed: 21677750]

39. Raymond CS & Soriano P High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. PLoS One 2, e162, doi:10.1371/journal.pone.0000162 (2007). [PubMed: 17225864]

40. Chu VT et al. Efficient generation of Rosa26 knock-in mice using CRISPR/Cas9 in C57BL/6 zygotes. BMC Biotechnol 16, 4, doi:10.1186/s12896-016-0234-4 (2016). [PubMed: 26772810]

41. Mazur PK et al. SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. Nature 510, 283–287, doi:10.1038/nature13320 (2014). [PubMed: 24847881]

42. Fraser M et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature 541, 359–364, doi:10.1038/nature20788 (2017). [PubMed: 28068672]

43. Shultz LD et al. Subcapsular transplantation of tissue in the kidney. Cold Spring Harb Protoc 2014, 737–740, doi:10.1101/pdb.prot078089 (2014). [PubMed: 24987138]

44. Iwano S et al. Single-cell bioluminescence imaging of deep tissue in freely moving animals. Science 359, 935–939, doi:10.1126/science.aaq1067 (2018). [PubMed: 29472486]

45. Fushiki H et al. Quantification of mouse pulmonary cancer models by microcomputed tomography imaging. Cancer Sci 100, 1544–1549, doi:10.1111/j.1349-7006.2009.01199.x (2009). [PubMed: 19459854]

46. Wang Z et al. SETD5-Coordinated Chromatin Reprogramming Regulates Adaptive Resistance to Targeted Pancreatic Cancer Therapy. Cancer Cell 37, 834–849 e813, doi:10.1016/j.ccell.2020.04.014 (2020). [PubMed: 32442403]

47. Salzmann M, Pervushin K, Wider G, Senn H & Wuthrich K TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. Proc Natl Acad Sci U S A 95, 13585–13590, doi:10.1073/pnas.95.23.13585 (1998). [PubMed: 9811843]

48. Balwierz W, Armata J, Moryl-Bujakowska A & Pekacki A Is first salvage chemotherapy the last-line chemotherapy in children with Hodgkin's disease? A tentative answer based on long observation of two patients. Folia Haematol Int Mag Klin Morphol Blutforsch 114, 789–796 (1987). [PubMed: 2453408]

49. Li Y et al. Backbone resonance assignments for the SET domain of human methyltransferase NSD3 in complex with its cofactor. Biomol NMR Assign 11, 225–229, doi:10.1007/s12104-017-9753-8 (2017). [PubMed: 28808922]

50. Shen Y & Bax A Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. Methods Mol Biol 1260, 17–32, doi:10.1007/978-1-4939-2239-0_2 (2015). [PubMed: 25502373]

51. Lakomek NA, Ying J & Bax A Measurement of (1)(5)N relaxation rates in perdeuterated proteins by TROSY-based methods. J Biomol NMR 53, 209–221, doi:10.1007/s10858-012-9626-5 (2012). [PubMed: 22689066]

52. Williamson MP Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc 73, 1–16, doi:10.1016/j.pnmrs.2013.02.001 (2013). [PubMed: 23962882]

53. van Zundert GCP et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol 428, 720–725, doi:10.1016/j.jmb.2015.09.014 (2016). [PubMed: 26410586]

54. Morrison MJ et al. Identification of a peptide inhibitor for the histone methyltransferase WHSC1. PLoS One 13, e0197082, doi:10.1371/journal.pone.0197082 (2018). [PubMed: 29742153]

55. Waterhouse A et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46, W296–W303, doi:10.1093/nar/gky427 (2018). [PubMed: 29788355]

56. Yang S et al. Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. Genes Dev 30, 1611–1616, doi:10.1101/gad.284323.116 (2016). [PubMed: 27474439]

57. Zhang Y et al. Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. Sci Rep 7, 43906, doi:10.1038/srep43906 (2017). [PubMed: 28256625]

58. Tisi D et al. Structure of the Epigenetic Oncogene MMSET and Inhibition by N-Alkyl Sinefungin Derivatives. ACS Chem Biol 11, 3093–3105, doi:10.1021/acschembio.6b00308 (2016). [PubMed: 27571355]

59. Kim D, Langmead B & Salzberg SL HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12, 357–360, doi:10.1038/nmeth.3317 (2015). [PubMed: 25751142]

60. Frankish A et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 47, D766–D773, doi:10.1093/nar/gky955 (2019). [PubMed: 30357393]

61. Liao Y, Smyth GK & Shi W featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930, doi:10.1093/bioinformatics/btt656 (2014). [PubMed: 24227677]

62. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550, doi:10.1186/s13059-014-0550-8 (2014). [PubMed: 25516281]

63. Subramanian A et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550, doi:10.1073/pnas.0506580102 (2005). [PubMed: 16199517]

64. Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. Nature 489, 519–525, doi:10.1038/nature11404 (2012). [PubMed: 22960745]

65. Gao J et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6, p11, doi:10.1126/scisignal.2004088 (2013).

66. Cerami E et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2, 401–404, doi:10.1158/2159-8290.CD-12-0095 (2012). [PubMed: 22588877]

67. Hainer SJ & Fazzio TG High-Resolution Chromatin Profiling Using CUT&RUN. Curr Protoc Mol Biol 126, e85, doi:10.1002/cpmb.85 (2019). [PubMed: 30688406]

68. Zhu Q, Liu N, Orkin SH & Yuan GC CUT&RUNTools: a flexible pipeline for CUT&RUN processing and footprint analysis. Genome Biol 20, 192, doi:10.1186/s13059-019-1802-4 (2019). [PubMed: 31500663]

69. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120, doi:10.1093/bioinformatics/btu170 (2014). [PubMed: 24695404]

70. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359, doi:10.1038/nmeth.1923 (2012). [PubMed: 22388286]

71. Shen L, Shao N, Liu X & Nestler E ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics 15, 284, doi:10.1186/1471-2164-15-284 (2014). [PubMed: 24735413]

72. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842, doi:10.1093/bioinformatics/btq033 (2010). [PubMed: 20110278]
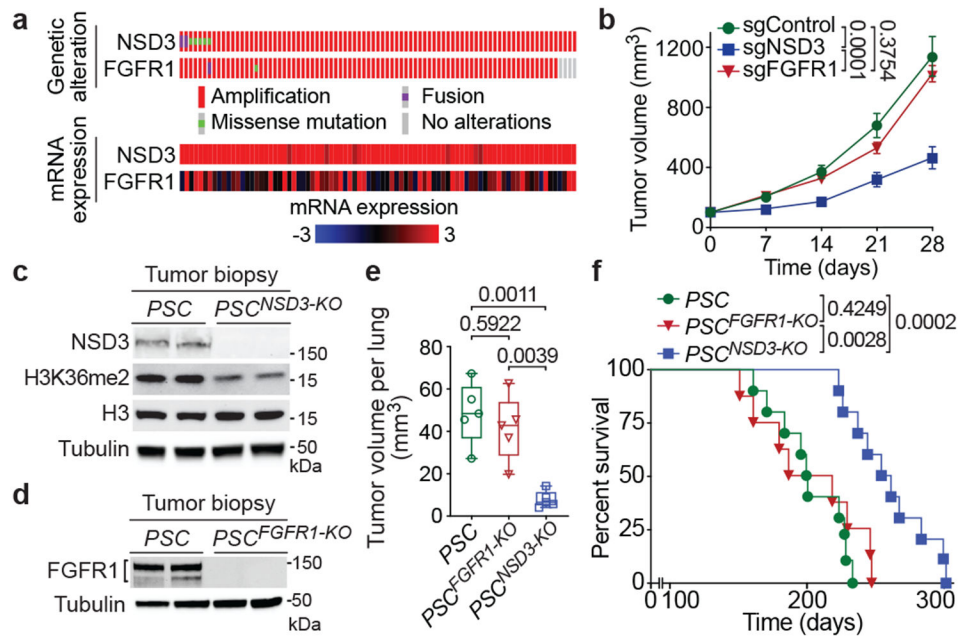
Figure 1. Deletion of NSD3, but not FGFR1, inhibits LUSC tumorigenesis *in vivo*.
a, Analysis of 8p11-12 amplified LUSC datasets from TCGA indicates increased NSD3, but not FGFR1, mRNA expression correlates with gene amplification (n = 85 patients). **b**, Depletion of NSD3, not FGFR1, attenuates xenograft growth of the 8p11^{AMP} LUSC cell line H520. Tumor volume quantification of H520 xenografts in immunocompromised mice (n = 5 mice for each group). **c**, **d**, Western blots with the indicated antibodies of LUSC tumor lysates from *PSC* (control), *PSC$^{NSD3-KO}$* (**c**) and *PSC$^{FGFR1-KO}$* (**d**) mutant mice. Two independent and representative samples are shown for each genotype. Tubulin used as a loading control. **e**, Micro-computed tomography (μCT) analysis of tumor volume of the indicated mouse models (n = 5 mice for each group). In this and subsequent box plots, the line indicates the median, the box marks the 75th and 25th percentiles, and the whiskers minimum and maximum values. All data points are shown. *P* values determined by two-way ANOVA with Tukey's post hoc test (**b**, **e**). f, Kaplan-Meier survival curves of *PSC* (n = 10, median survival = 200.5 days), *PSC$^{FGFR1-KO}$* (n = 8, median survival = 202.5 days) and *PSC$^{NSD3-KO}$* (n = 10, median survival = 257 days) mice, *P* value determined by log-rank test.
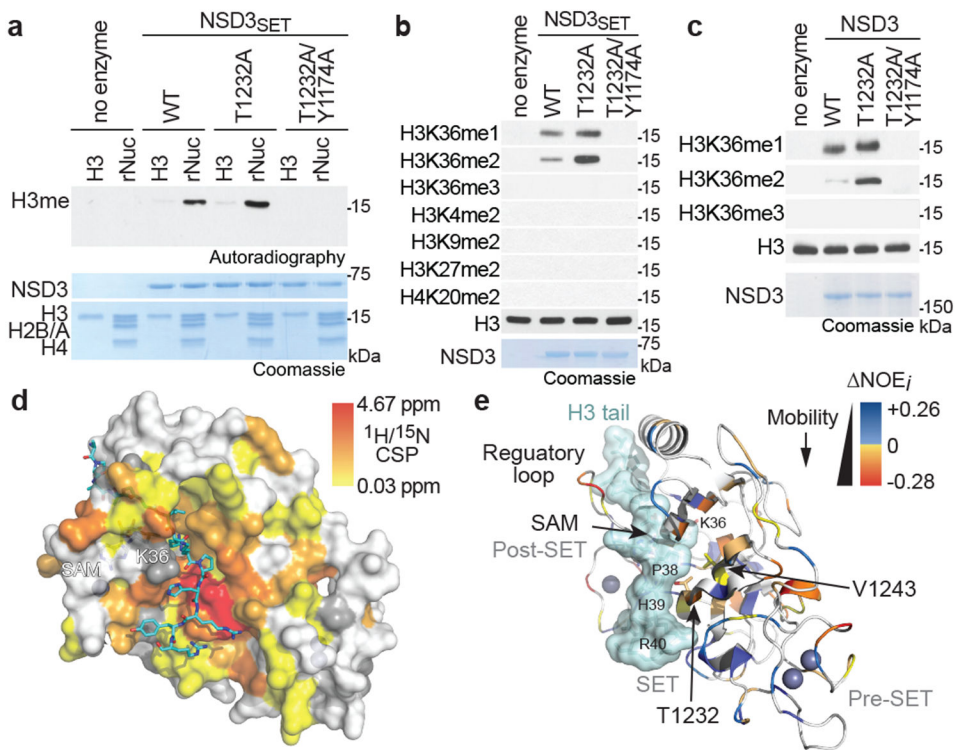
Figure 2. Molecular basis of increased H3K36me2 catalysis by NSD3_{T1232A}.

a, *In vitro* methylation reactions of wild-type (WT) or mutant NSD3_{SET} using recombinant free histone H3 or nucleosome (rNuc) and ^{3}H-SAM as substrates. No enzyme is a negative control. Top panel, autoradiography; bottom panel, Coomassie blue staining. **b**, Methylation assays on rNuc as in (**a**) with non-radiolabeled SAM. Westerns of reaction products shown with the indicated antibodies. H3 is shown as loading control. **c**, Methylation assays as in (**b**) with recombinant full-length wild-type or mutant NSD3. Data in **a-c** are representative of three or more biologically independent experiments with similar results. **d-e**, T1232A substitution induces widespread mobility changes on the NSD3 catalytic domain. **d**, Amide CSPs between NSD3_{SET-T1232A} and NSD3_{SET} is shown mapped onto the surface representation of the NSD3_{SET} structure (PDB: 6CEN) with docked H3.1 residues A29 to R42 (stick representation). Gray: prolines and residues missing amide assignments. **e**, Changes in heteronuclear {^{1}H}-^{15}N nuclear Overhauser effect (NOE) values plotted on the structure of NSD3 as in (d) in ribbon representation. T1232 and V1243 residues are indicated. Uncolored residues: undetectable NOE changes.
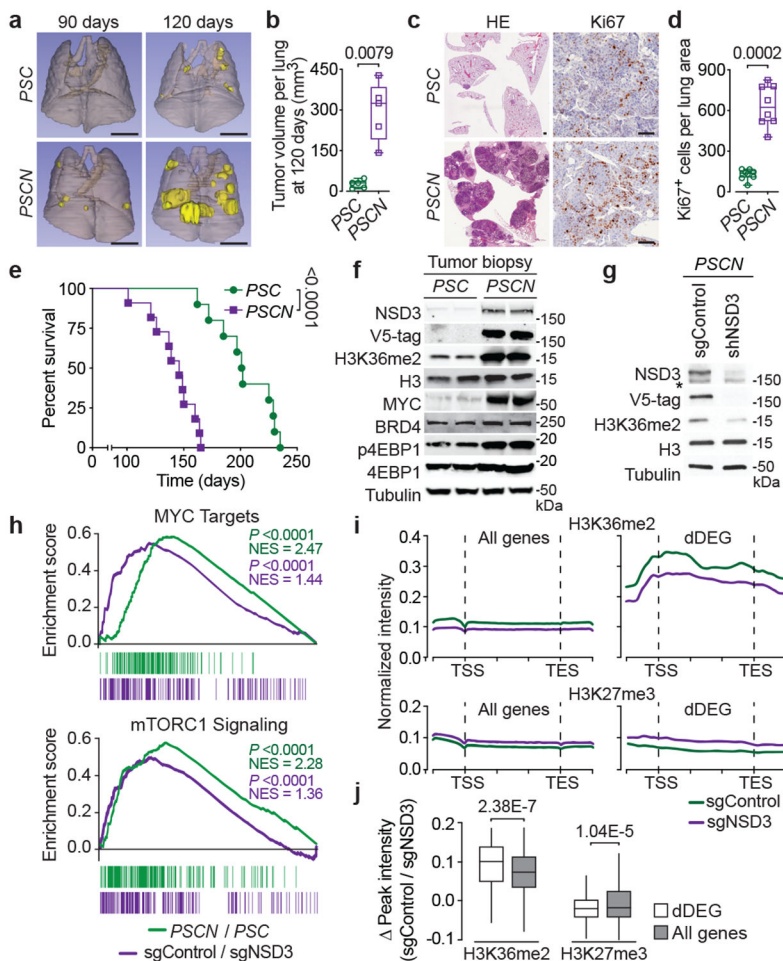
Figure 3. Mutant NSD3-mediated H3K36me2 synthesis promotes oncogenic programming and LUSC tumorigenesis.

a, Representative μCT scans at 90 and 120 days after adenoviral-Cre mediated tumor induction in *PSC* and *PSCN* mice. Scale bars: 10 mm. **b**, Quantitation of tumor volume as in (a) at 120 days (n = 5 mice for each group). **c**, Representative HE and IHC staining of lung tissue from *PSCN* and control *PSC* mutant mice (n = 8 mice for each group). Scale bars: 100 μm. **d**, Quantification of proliferation (Ki67[+]-cells) as in (c). *P* values determined by two-tailed unpaired t-test. (**b**, **d**). **e**, Kaplan-Meier survival curves of *PSCN* (n = 11, median survival = 145 days) and *PSC* (same data as in Fig. 1f, n = 10, median survival = 200.5 days) mutant mice, indicated *P* value determined by log-rank test. **f**, Western blots with the indicated antibodies of LUSC tumor lysates from *PSCN* and *PSC* mice. Two independent and representative samples are shown for each genotype. **g**, Western blots with the indicated antibodies of cell lysates from either control (sgControl) or NSD3-depleted (sgNSD3) PSC_N cells. H3 and tubulin shown as loading controls in (**f**, **g**). **h**, GSEA identifies upregulation of hallmark gene sets: MYC targets and mTOR signaling (MSigDB: M5928, M5924) in RNA-seq data from *PSCN* versus *PSC* tumor biopsies and PSC_N cells ± sgNSD3 (n = 3 biologically independent samples per group). Normalized enrichment scores (NES) and nominal *P* values are provided (detailed statistics description in Methods). **i**, CUT&RUN profiles of H3K36me2 (upper panel) or H3K27me3 (lower panel) over averaged gene body

for all genes or dDEG in PSC$_N$ cells ± sgNSD3 as indicated. **j,** Quantification of H3K36me2 and H3K27me3 peak intensity change in PSC$_N$ cells ±sgNSD3 on the indicated gene sets. dDEG $n = 234$, all genes $n = 16091$. Two-sided t-test P value provided.
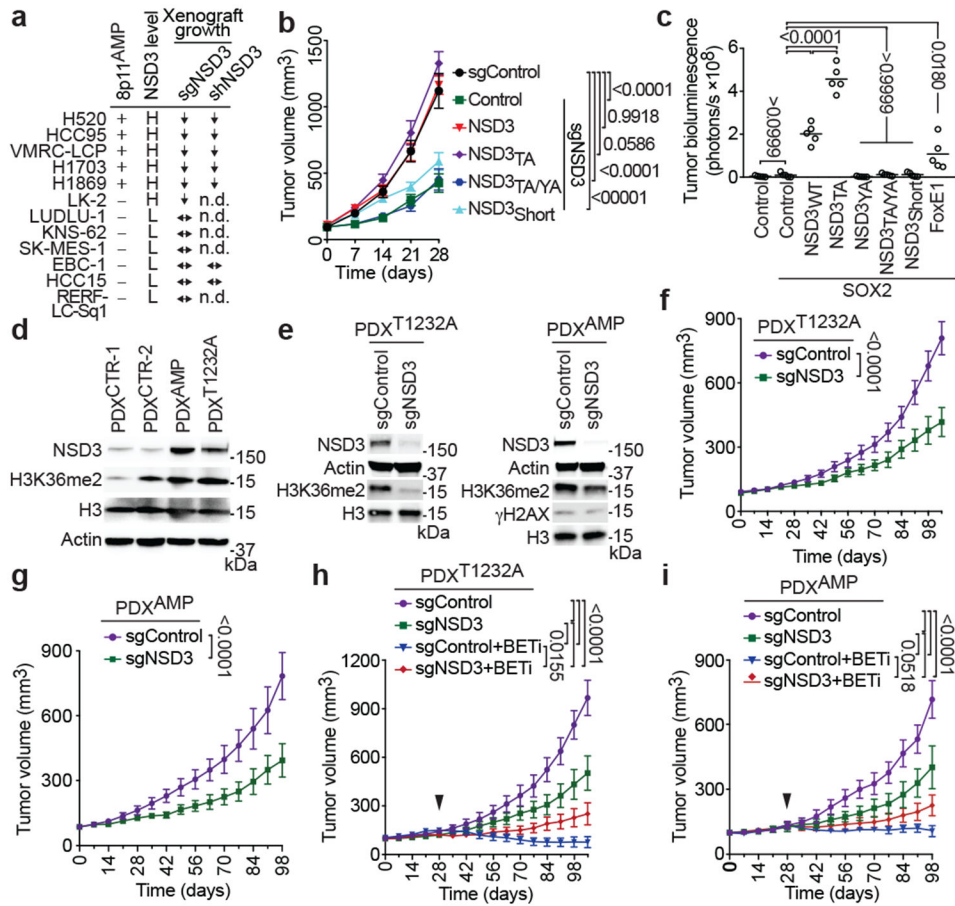
Figure 4. NSD3 promotes human lung cell transformation and xenograft tumor growth of LUSC cells and PDXs and renders PDXs hyper-susceptible to BETi.

a, NSD3 depletion attenuates xenograft tumor growth of 8p11^AMP and NSD3-overexpressing LUSC cell lines. Summary of xenograft tumor growth for the indicated cell lines treated with sgNSD3 or shNSD3 as indicated. NSD3 levels: H, high mRNA (Z-score > 1.0); L, low mRNA (Z-score < 1.0), scoring was consistent with relative protein levels (see Extended Data Figs 7a-b); xenograft tumor growth: ↓: reduced; ↔: no change (see Extended Data Figs. 7b-c); n.d., not determined. **b**, NSD3 catalytic activity required for full H520 xenograft tumor growth. Tumor volume change of H520 xenografts reconstituted with the indicated V5-tagged CRISPR-resistant NSD3 derivatives in immunocompromised mice (n = 5 mice, for each group). **c**, *NSD3$_{WT}$* and *NSD3$_{TA}$* transform *SOX2*-expressing AALE cells *in vivo*. Quantification of tumor size determined by bioluminescence of AALE cells grafted under the renal capsule and expressing the indicated plasmids plus the AkaLuc reporter (n=5 mice for each group; see Methods). FOXE1 is a positive control, NSD3$_{WT}$ and NSD3$_{TA}$ *vs.* FOXE1 *P* < 0.0001. Lines denote median. **d**, Western blots of lysates from the indicated LUSC PDX samples and using the indicated antibodies. H3 and Actin are loading controls. **e**, Western blots as in (**d**) of lysates from PDX^T1232A (left) and PDX^AMP (right) ±sgNSD3 treatment. **f, g**, PDX tumor volume quantification of **f**, PDX^T1232A and **g**, PDX^AMP ±sgNSD3 in immunocompromised mice (n = 5 for each group). **h, i**, Tumor volume quantification of **h**, PDX^T1232A and **i**, PDX^AMP ±sgNSD3 and ±BETi (AZD5153 2.5mg/kg,

i.p.) treatments as indicated (n = 5 mice for each group). Control animals received vehicle (placebo) treatment. Arrowhead indicates start of the treatment. *P* values determined by two-way ANOVA with Tukey's post hoc test (**b, c, h, i**) or two-tailed unpaired t-test. (**f**, **g**). Data are represented as mean ± s.e.m. (**b**, **f-i**).