

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/349007559>

Elevated NSD3 histone methylation activity drives squamous cell lung cancer

Article in *Nature* · February 2021

DOI: 10.1038/s41586-020-03170-y

CITATIONS

0

READS

214

22 authors, including:



Gang Yuan

Stanford University

6 PUBLICATIONS 85 CITATIONS

[SEE PROFILE](#)



Simone Hausmann

University of Texas MD Anderson Cancer Center

26 PUBLICATIONS 765 CITATIONS

[SEE PROFILE](#)



Shane M Lofgren

Stanford University

17 PUBLICATIONS 597 CITATIONS

[SEE PROFILE](#)



Maria Angulo-Ibáñez

Stanford Medicine

11 PUBLICATIONS 67 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Mechanisms of cancer progression and drug resistance [View project](#)



SETD7 Drives Cardiac Lineage Commitment through Stage-Specific Transcriptional Activation [View project](#)

Elevated NSD3 Histone Methylation Activity Drives Squamous Cell Lung Cancer

Gang Yuan^{1,11}, Natasha M. Flores^{2,11}, Simone Hausmann², Shane M. Lofgren², Vladlena Kharchenko³, Maria Angulo-Ibanez^{4,5}, Deepawita Sengupta¹, Xiaoyin Lu², Iwona Czaban³, Dulat Azhibek³, Silvestre Vicent⁶, Wolfgang Fischle³, Mariusz Jaremko³, Bingliang Fang⁷, Ignacio I. Wistuba⁸, Katrin Chua^{4,5}, Jack A. Roth⁷, John D. Minna⁹, Ning-Yi Shao^{10,#}, Łukasz Jaremko^{3,#}, Pawel K. Mazur^{2,#} and Or Gozani^{1,#}

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

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

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

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

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

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

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

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

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

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

¹¹These 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)^{1,2}. FGFR1 (fibroblast growth factor receptor 1) is the main candidate driver within this region^{3,4}. However, clinical trials evaluating FGFR1 inhibition as a targeted therapy have been unsuccessful^{5,6}. 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

There are many genes present within the 8p11-12 amplicon that 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*^{3,7-13}. 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)^{4,14}. 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)^{2,7,8}. 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.

To directly compare the role of *NSD3* and *FGFR1* in LUSC pathogenesis, guided by previous studies^{15,16}, we established a robust LUSC mouse model. Canonical LUSC alterations that co-occur with 8p11-12 amplification¹⁴ were engineered to be under Cre-recombinase regulation: (1) constitutively active PI3K (*Pik3ca*^{H1047R}), (2) SOX2 overexpression (*Sox2*^{OE}), and (3) *CDKN2A/B* deletion (*Cdkn2a/b*^{KO}), hereto referred to as *PSC* mice (Fig. 1c; Extended Data Fig. 2f). The *PSC* mice develop with high penetrance lung tumors characterized by the diagnostic molecular hallmarks that define lung squamous cell carcinoma¹⁵⁻¹⁷, including staining positive for keratin 5 and p63 and negative for TTF1 (Extended Data Fig. 2g). We note that in this model, increased

NSD3 expression tracks with tumor progression (Extended Data Figs. 2h-i). These data are consistent with NSD3 overexpression being observed in ~60% of LUSC samples (beyond the 20% of samples harboring 8p11-12 amplification) and frequently co-occurring with the molecular alterations present in the *PSC* signature (Extended Data Fig. 2j).

Conditional *FGFR1*^{loxP/loxP} and *NSD3*^{loxP/loxP} mutant mice were generated (Fig. 1c), which develop normally and are viable and fertile (data not shown). Targeted gene deletion in the lung was achieved by delivering Cre-recombinase via adenoviral tracheal lavage (Extended Data Figs. 2k-n). In the *PSC* background, deletion of NSD3 (*PSC*^{NSD3-KO}) (Fig. 1d), but not FGFR1 (*PSC*^{FGFR1-KO}) (Fig. 1e), significantly attenuated tumor growth (Figs. 1f-g) and cancer cell proliferation (Figs. 1f-h), while increasing apoptosis (Extended Data Figs. 2o-p). In survival studies, deletion of *NSD3* extended lifespan of *PSC* mice by ~30%, whereas *FGFR1* knockout had no impact (Fig. 1i). 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, are all linked to diverse cancer pathways (reviewed in ^{10,11}) and comprise the four enzymes in mammals that specifically synthesize H3K36me2, a chromatin modification linked to transcription activation^{10,11,18,19}. Depending on cell type, NSD1 and NSD2 are the major H3K36me2-generating enzymes^{20,21}, whereas the physiologic context in which NSD3 regulates H3K36me2 is less clear¹⁸. We observed lower global H3K36me2 levels in NSD3-deleted lung tumor tissue compared to control tumors (Fig. 1d), suggesting an etiological role for NSD3-catalyzed H3K36me2 synthesis in *NSD3* amplified-LUSC. We reasoned that this idea could be tested in mice by transgenic expression of an NSD3 hyperactive catalytic

mutant, which would model the consequence of increased NSD3 enzymatic activity due to elevated *NSD3* dosage on tumorigenesis. Indeed, a recurrent hyperactivating mutation within the NSD2 catalytic SET domain is present in hematologic malignancies and solid tumors^{11,22,23}. We hypothesized 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-b). The variant showing the greatest increase in histone methylation activity compared to wild-type enzyme was the T1232A substitution (NSD3_{T1232A}), a recurrent cancer-associated mutation²⁴.

The *in vitro* methylation activity on nucleosomes of recombinant NSD3 catalytic domain harboring the T1232A substitution (NSD3_{SET-T1232A}) was stronger than wild-type (NSD3_{SET}) enzyme (Figs. 2a-b; Extended Data Fig. 3c). This activity was abrogated when paired with a catalytic mutant (NSD3_{SET-T1232A/Y1174A}) (Fig. 2b)²¹. Further, NSD3_{SET-T1232A} was more efficient in specifically generating H3K36me2 relative to wild-type enzyme (Fig. 2c). Despite the higher activity, NSD3_{SET-T1232A} behaved similar to NSD3_{SET} with respect to (i) nucleosomal versus free H3 substrate preference (Fig. 2b), (ii) methylation state specificity, generating H3K36me1/2 but not H3K36me3 (Fig. 2c), and (iii) site-specific selectivity, methylating K36, but not several other histone lysine residues (Fig. 2c). Recombinant full-length NSD3_{T1232A} behaves like the isolated catalytic domain in being hyperactive while retaining the same catalytic selectivity profile (Fig. 2d). 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. 3d)²¹. Taken together, our findings identify NSD3_{T1232A} as a hyperactive cancer-associated mutant.

The enhanced *in vitro* methylation activity of NSD3_{SET-T1232A} versus NSD3_{SET} is also evident when using nucleosomes with linker DNA (wrapped with 187 base-pairs (bp) rather than 147bp

of DNA) (Extended Data Fig. 3e). The mechanism behind the increased methylation activity, irrespective of linker DNA, is not mediated by the T1232A substitution significantly altering NSD3's binding affinity for nucleosomes or DNA (Extended Data Figs. 4a-c; data not shown). Further, NSD3_{SET-T1232A} and NSD3_{SET} bind the cofactor methyl donor *S*-adenosyl-L-methionine (SAM) with nearly comparable affinities (Extended Data Fig. 4d-e). 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. 4f-i). In contrast, chemical shift perturbations (CSPs) of the backbone amide ¹H/¹⁵N resonances, which serve as sensitive probes for local structural and allosteric effects, were observed throughout the SET domain (Fig. 2e; Extended Data Figs. 4i-j). The highest CSPs are centered around T1232A, with effects extending throughout the substrate-binding cleft and include the conserved NSD3 auto-regulatory loop (Fig. 2e; Extended Data Figs. 4j-k)^{25,26}. Consistent with these findings, backbone ¹⁵N 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. 2f; Extended Data Figs. 4l-n)²⁷. These data suggest that the molecular basis by which the T1232A substitution confers increased activity on NSD3 is via (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. In this regard, the loss of the polar hydroxyl group with the T1232A substitution increases local hydrophobicity, which can alter H3 anchoring dynamics (Figs. 2e-f)²⁸. Introduction of a second alanine substitution at V1243, a residue adjacent to A1232 in the structure, (T1232A/V1243A) is predicted to restore local surface hydrophobicity²⁸. Accordingly, the methylation activity of a double alanine mutant (NSD3_{SET-T1232A/V1243A}) is similar

to wild-type enzyme and lower than the activity of the single mutant (NSD3_{SET-T1232A}), while the NSD3_{SET-V1243A} shows no activity and the V1243A substitution decreased thermal stability (Fig. 2g; Extended Data Figs 4o-p). Thus, the 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

Expression of the NSD3_{T1232A} variant was next used to model consequences of elevated NSD3 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 expression of the transgene 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} (*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 the lifespan of *PSC* mice 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 referred to as *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 *PSC_N* to *PSC* tumor biopsies than for *PSC_N* cells \pm NSD3 depletion, in which NSD3 was depleted for a relatively short time prior to the analysis (Extended Data Figs. 6a-b and Tables 1-2). 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. 6c). 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. 6d-e).

Next, Cleavage Under Targets and Release Using Nuclease (CUT&RUN)²⁹ was used to profile the genomic distribution of H3K36me2 in *PSC_N* cells \pm NSD3 depletion. We also mapped H3K27me3 as the deposition of this important chromatin modification is directly antagonized by H3K36 methylation^{22,30-35}. 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. 6f). 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. 6f). 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. 6f)²¹. Furthermore, quantitation of the difference between H3K36me2 peak signal intensities \pm NSD3 depletion reveals that the magnitude of the peak change at the gene bodies of dDEGs is significantly greater than the change observed genome-wide (Fig. 3j). Notably, relative to other genes, the dDEGs show both high H3K36me2 chromatin signal and high mRNA transcript levels (Extended Data Fig. 6g). 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 both genome-wide and at dDEGs, with no quantitative difference in the magnitude of the change between the two groups of genes (Figs. 3i-j). The inverse relationship between H3K36me2 and H3K27me3 in relation to NSD3 status also extends to intergenic regions (Extended Data Fig. 6h). The pattern of diminished H3K36me2 and increased H3K27me3 in response to NSD3 depletion is also observed at the level of individual dDEG target gene tracks (Extended Data Fig. 6i). 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 NSD3, but not with catalytic dead NSD3 (Extended Data Figs. 6j-l). 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 NSD3-overexpressing cell line with no 8p11-12 amplification and a similar genetic background as the PSC mice, and six control cell lines lacking NSD3 overexpression and 8p11-12 amplification (Fig. 4a; Extended Data Fig. 7a). NSD3 depletion impaired xenograft tumor formation in the 8p11^{AMP} and 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-c). It is unlikely that the phenotype in the five 8p11^{AMP} LUSC lines is a result of non-specific DNA damage due to sgNSD3-targeted cutting within amplified regions³⁶ as: (i) cutting in amplified regions by sgFGFR1 and sgPLPP5 had no effect on xenograft growth in 8p11^{AMP}-positive H520 cells (Extended Data Fig. 2e), (ii) γ H2AX levels were not altered by sgNSD3 treatment in the 8p11^{AMP} cell lines (Extended Data Figs. 2d and 7b), and (iii) depletion of NSD3 in the 8p11^{AMP} LUSC cell lines by RNAi as an independent method attenuated xenograft growth (Fig. 4a; Extended Data Figs. 7d-e). 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 pathogenesis¹² both failed to reconstitute full xenograft growth (Figs. 4b-c).

We next investigated the ability of NSD3 to cooperate with SOX2 to transform human tracheobronchial epithelial (AALE) cells³⁷. As previously reported, ectopic expression of SOX2 with a second transcription factor (FOXE1) promotes anchorage-independent growth of AALE cells³⁷. 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* (Fig. 4d; Extended Data Figs. 7f-g)

and tumor growth *in vivo* (Fig. 4e; Extended Data Fig. 7h). In contrast, catalytically dead NSD3 and NSD3_{short} failed to promote AALE transformation (Figs. 4d-e; Extended Data Figs. 7f-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.

Patient-derived xenografts (PDX) from primary LUSC were used to further investigate roles of the NSD3-H3K36me2 axis in human cancer. A collection of 37 LUSC PDX samples was assembled from the NCI and MD Anderson PDX repositories and screened for NSD3 alterations. Of the 37 samples, 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 Figs. 8a-b; 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 (i.e. PI3K/SOX2 activation and *CDKNA/B* deletion), were selected for further analysis (Extended Data Fig. 8a). In comparison to the PDX^{CTR} samples, NSD3 levels are higher in PDX^{T1232A} cells and highest in PDX^{AMP-1} cells (Fig. 4e). H3K36me2 levels in the PDX^{T1232A} and PDX^{AMP-1} cells are comparable to one another and elevated in comparison to control samples (Fig 4e); the similar abundance of H3K36me2 in PDX^{T1232A} and PDX^{AMP-1} cells, while the former has lower NSD3 levels, is consistent with the NSD3^{T1232A} variant being more active than wild-type enzyme. Depletion of NSD3, which resulted in H3K36me2 loss, attenuated growth of the PDX^{T1232A} and PDX^{AMP-1} xenograft tumors (Figs. 4f-h) but had no effect on growth of the PDX^{CTR} samples (Extended Data Figs. 9a-b). Inhibition of PDX growth 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 tumors to drugs targeting known cancer pathways. To test this idea, a comparative cell-based screen using a library assembled of 285 characterized inhibitors covering ~170 cancer targets was performed to identify drugs that have increased efficacy in PSC_N versus PSC_C cells (Extended Data Table 3)³⁸. The four BETi drugs present in the library exhibited the highest differential lethality of 65-70% in the screen, consistent with BRD4 interacting directly with NSD3 in LUSC cells (Extended Data Figs. 9d-e)^{39,40}. Under normal conditions, PSC_N cells proliferate more rapidly than PSC_C cells (Fig. 4i). However, treatment with the BETi AZD5153⁴¹, which has a modest effect on proliferation of PSC_C cells, inhibits NSD3-target gene expression in and proliferation of PSC_N cells (Fig. 4i; Extended Data Fig. 9f). The NSD3-dependent sensitivity to BETi was also observed in human PDX samples. Actively growing PDX^{T1232A} and PDX^{AMP} tumors stall in response to BETi treatment (Figs. 4f-h), a phenotype that is weakened by NSD3 depletion (Figs. 4j-l). In contrast, PDX^{CTR} samples are moderately responsive to BETi therapy in an NSD3-independent fashion (Extended Data Figs. 9g-h). These data suggest that the oncogenic advantage provided by NSD3 to LUSC tumors 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^{4,14}. 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 8p11-12 positive LUSC^{14,42}. 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 transcription⁴³. Indeed, NSD3 depletion was recently identified in a screen of 2240 genes as a strong sensitizer of acute myeloid leukemia cells to BETi⁴⁴. In this context, NSD3 directly interacts with the BET protein BRD4^{39,40} (Extended Data Fig. 9e), suggesting a direct role for BRD4 in NSD3-mediated epigenetic deregulation. While clinical compounds targeting NSD3 are presently unavailable⁴⁵, the increased sensitivity of NSD3-regulated LUSC to BETi may expand the narrow therapeutic window for these potentially efficacious but toxic medicines⁴³. Finally, NSD3's oncogenic activity likely extends beyond LUSC, as the 8p11-12 amplicon is a common molecular signature of breast cancer and many other malignancies^{8,13}.

Acknowledgements

We thank members of the Gozani and Mazur labs for critical reading of the manuscript. We thank Michael Lin for the AkaLuc vector. This work was supported in part by grants from the NIH to O.G. (R01 GM079641), K.F.C (R01AG050997) and P.K.M. (R00 CA197816), 1U54CA224065 to B.F. and J.A.R., VA Merit Award to K.F.C., intramural funds from KAUST to W.F., L.J. and M.J, the P. Neuroendocrine Tumor Research Foundation, AACR, DOD PRCRP Career Development Award (CA181486), Career Enhancement Grant - The University of Texas NIH SPORE in Lung Cancer (P50CA070907), the Andrew Sabin Family Foundation Scientist and CPRIT Scholar in Cancer Research (RR160078) to P.K.M. N.Y.S. was supported by SRG2019-00177-FHS, X.L by CPRIT Research Training Grant (RP170067) and S.H. by a Deutsche Forschungsgemeinschaft Fellowship.

Author Contributions

G.Y. and N.M.F. contributed equally to this work. They were responsible for the experimental design, execution, data analysis, and manuscript preparation. S.H. performed PDX, and AALE studies and IHC analysis. D.S., X.L. G.Y. and S.V. helped with PDX and AALE studies. M.A.I. helped G.Y. with CUT&RUN experiments and K.C. supervised. S.M.L. performed bioinformatic analysis of tumor RNA-seq. V.K. I.C. D.A. W.F. M.J. and L.J. performed biophysical and NMR experiments. I.I.W. evaluated histopathology, I.I.W., B.F. J.A.R., J.D.M. provided PDX samples. N.Y.S. analyzed RNA-seq and CUT&RUN data. O.G., P.K.M., L.J. and N.Y.S. were equally responsible for supervision of research, data interpretation and manuscript preparation.

Competing Interests

Or Gozani is a co-founder of EpiCypher, Inc. and Athelas Therapeutics, Inc. Jack A. Roth is a consultant, stock owner (including pending patents) in Genprex, Inc. Pawel K. Mazur is a co-founder of Amplified Medicines, Inc.

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. Review access tokens: GSE149482: gdgbcaimvhgfrcl; GSE149272: mpapgggsnnefhkp

References:

- 1 Balsara, B. R. *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).
- 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).
- 3 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).
- 4 Hirsch, F. R., Suda, K., Wiens, J. & Bunn, P. A., Jr. New and emerging targeted treatments in advanced non-small-cell lung cancer. *Lancet* **388**, 1012-1024, doi:10.1016/S0140-6736(16)31473-8 (2016).
- 5 Lim, S. H. *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).
- 6 Paik, P. K. *et al.* A Phase Ib Open-Label Multicenter Study of AZD4547 in Patients with Advanced Squamous Cell Lung Cancers. *Clin Cancer Res* **23**, 5366-5373, doi:10.1158/1078-0432.CCR-17-0645 (2017).
- 7 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).
- 8 Parris, T. Z. *et al.* Genome-wide multi-omics profiling of the 8p11-p12 amplicon in breast carcinoma. *Oncotarget* **9**, 24140-24154, doi:10.18632/oncotarget.25329 (2018).
- 9 Dutt, A. *et al.* Inhibitor-sensitive FGFR1 amplification in human non-small cell lung cancer. *PLoS One* **6**, e20351, doi:10.1371/journal.pone.0020351 (2011).
- 10 Li, J., Ahn, J. H. & Wang, G. G. Understanding histone H3 lysine 36 methylation and its deregulation in disease. *Cell Mol Life Sci* **76**, 2899-2916, doi:10.1007/s00018-019-03144-y (2019).
- 11 Bennett, R. L., Swaroop, A., Troche, C. & Licht, J. D. The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in Cancer. *Cold Spring Harb Perspect Med* **7**, doi:10.1101/cshperspect.a026708 (2017).
- 12 Yang, Z. Q., Liu, G., Bollig-Fischer, A., Giroux, C. N. & Ethier, S. P. 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).
- 13 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).
- 14 Ruiz-Cordero, R. & Devine, W. P. Targeted Therapy and Checkpoint Immunotherapy in Lung Cancer. *Surg Pathol Clin* **13**, 17-33, doi:10.1016/j.path.2019.11.002 (2020).
- 15 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).
- 16 Mukhopadhyay, A. *et al.* Sox2 cooperates with Lkb1 loss in a mouse model of squamous cell lung cancer. *Cell Rep* **8**, 40-49, doi:10.1016/j.celrep.2014.05.036 (2014).

- 17 Travis, W. D. Lung Cancer Pathology: Current Concepts. *Clin Chest Med* **41**, 67-85, doi:10.1016/j.ccm.2019.11.001 (2020).
- 18 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).
- 19 Li, Y. *et al.* The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J Biol Chem* **284**, 34283-34295, doi:M109.034462 [pii]10.1074/jbc.M109.034462 (2009).
- 20 Papillon-Cavanagh, S. *et al.* Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. *Nat Genet* **49**, 180-185, doi:10.1038/ng.3757 (2017).
- 21 Kuo, A. J. *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).
- 22 Jaffe, J. D. *et al.* Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat Genet* **45**, 1386-1391, doi:10.1038/ng.2777 (2013).
- 23 Oyer, J. A. *et al.* Point mutation E1099K in MMSET/NSD2 enhances its methyltransferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia* **28**, 198-201, doi:10.1038/leu.2013.204 (2014).
- 24 Landau, D. A. *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).
- 25 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).
- 26 Graham, S. E., Tweedy, S. E. & Carlson, H. A. 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).
- 27 Lakomek, N. A., 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).
- 28 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).
- 29 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, doi:10.7554/eLife.21856 (2017).
- 30 Lu, C. *et al.* Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science* **352**, 844-849, doi:10.1126/science.aac7272 (2016).
- 31 Fang, D. *et al.* The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. *Science* **352**, 1344-1348, doi:10.1126/science.aae0065 (2016).
- 32 Huang, C. & Zhu, B. Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. *Biophys Rep* **4**, 170-177, doi:10.1007/s41048-018-0063-1 (2018).
- 33 Streubel, G. *et al.* The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol Cell* **70**, 371-379 e375, doi:10.1016/j.molcel.2018.02.027 (2018).
- 34 Popovic, R. *et al.* Histone methyltransferase MMSET/NSD2 alters EZH2 binding and reprograms the myeloma epigenome through global and focal changes in H3K36 and H3K27 methylation. *PLoS Genet* **10**, e1004566, doi:10.1371/journal.pgen.1004566 (2014).

- 35 Swaroop, A. *et al.* An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. *Oncogene* **38**, 671-686, doi:10.1038/s41388-018-0474-y (2019).
- 36 Munoz, D. M. *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).
- 37 Bass, A. J. *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).
- 38 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).
- 39 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).
- 40 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).
- 41 Bradbury, R. H. *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-dimethylpiperazin-2-one (AZD5153). *J Med Chem* **59**, 7801-7817, doi:10.1021/acs.jmedchem.6b00070 (2016).
- 42 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J Clin* **69**, 7-34, doi:10.3322/caac.21551 (2019).
- 43 Cochran, A. G., Conery, A. R. & Sims, R. J., 3rd. Bromodomains: a new target class for drug development. *Nat Rev Drug Discov* **18**, 609-628, doi:10.1038/s41573-019-0030-7 (2019).
- 44 Lin, K. H. *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).
- 45 Bottcher, J. *et al.* Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3. *Nat Chem Biol* **15**, 822-829, doi:10.1038/s41589-019-0310-x (2019).

FIGURE LEGENDS

Figure 1. Deletion of NSD3, but not FGFR1, inhibits tumorigenesis *in vivo* in a LUSC mouse

model. a, Analysis of 8p11-12 amplified LUSC datasets from the TCGA indicates that increased NSD3, but not FGFR1, mRNA expression strongly correlates with gene amplification. **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 = 6 mice, for each treatment group). *** $P < 0.0001$, n.s. not significant by two-way ANOVA with Tukey's testing for multiple comparisons. Data are represented as mean \pm s.e.m. **c**, Schematic of a LUSC mouse model with constitutively active *Pik3ca*^{H1047R}, *Sox2* overexpression, and *Cdkn2a/b* deletion (PSC) and generation of conditional *Fgfr1* and *Nsd3* mutants in the PSC background as indicated. **d**, Western blots with the indicated antibodies of LUSC tumor lysates from *PSC*^{NSD3-KO} and *PSC* (control) mutant mice. Two independent and representative samples are shown for each genotype. Tubulin is shown as a loading control. **e**, Western blots with the indicated antibodies of LUSC tumor lysates from *PSC*^{FGFR1-KO} and *PSC* (control) mutant mice as in (c). **f**, Representative HE and immunohistochemical (IHC) staining for cell proliferation (Ki67) of lung tissue from *PSC*^{FGFR1-KO}, *PSC*^{NSD3-KO}, and control *PSC* mutant mice. Representative of n = 8 mice for each experimental group). Scale bars: 100 μ m. **g**, Micro-computed tomography (μ CT) analysis of tumor volume of the indicated mouse models (n = 6 mice for each experimental group). Boxes: 25th to 75th percentile, whiskers: min. to max., center: median; *** $P < 0.0001$, n.s., not significant by two-way ANOVA with Tukey's testing for multiple comparisons. **h**, Quantification of proliferation (Ki67⁺ positive cells) in samples as in (e). Boxes: 25th to 75th percentile, whiskers: min. to max., center: median; *** $P < 0.0001$, n.s., not significant by two-way ANOVA with Tukey's testing for multiple comparisons. **i**, 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}* (control, n = 10, median survival = 257 days) mice, *** $P = 0.001$, n.s., not significant by log-rank test for significance.

Figure 1

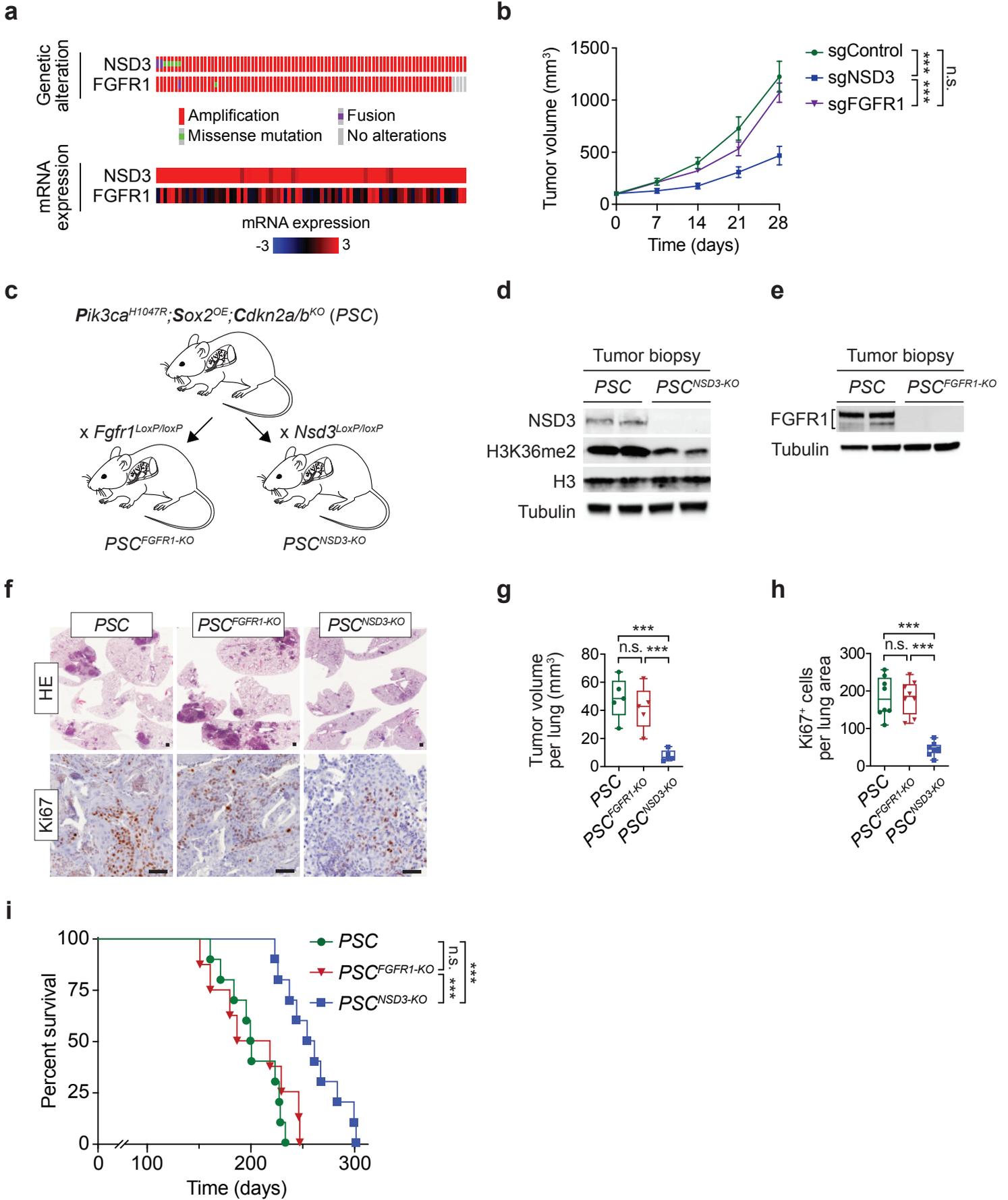


Figure 2. Molecular basis of increased H3K36me2 catalysis by NSD3_{T1232A}. **a**, Schematic diagram of NSD3 fragments used in this figure. **b**, *In vitro* methylation reactions of recombinant wild-type (WT) or mutant NSD3_{SET} using recombinant free histone H3 or nucleosome (rNuc) and ³H-SAM as substrates. No enzyme is a negative control. Top panel, autoradiography; bottom panel, Coomassie blue staining. **c**, Methylation assays on rNuc as in (b) with non-radiolabeled SAM. Western blots of the reaction products with the indicated antibodies. H3 is shown as a loading control. **d**, Methylation assays as in (c) with recombinant full-length wild-type or mutant NSD3. Top panel, Western blot of the reaction products with indicated antibodies. Bottom panel, Coomassie blue staining of NSD3 proteins. **e-f**, T1232A substitution induces widespread mobility changes on the NSD3 catalytic domain. **e**, 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. **f**, Changes in heteronuclear {¹H}-¹⁵N nuclear Overhauser effect (NOE) values plotted on the structure of NSD3 as in (e) in ribbon representation. T1232 and V1243 residues are indicated. Uncolored residues: undetectable NOE changes. **g**, Methylation assays as in (b) with rNuc(187bp) and the indicated NSD3 variant.

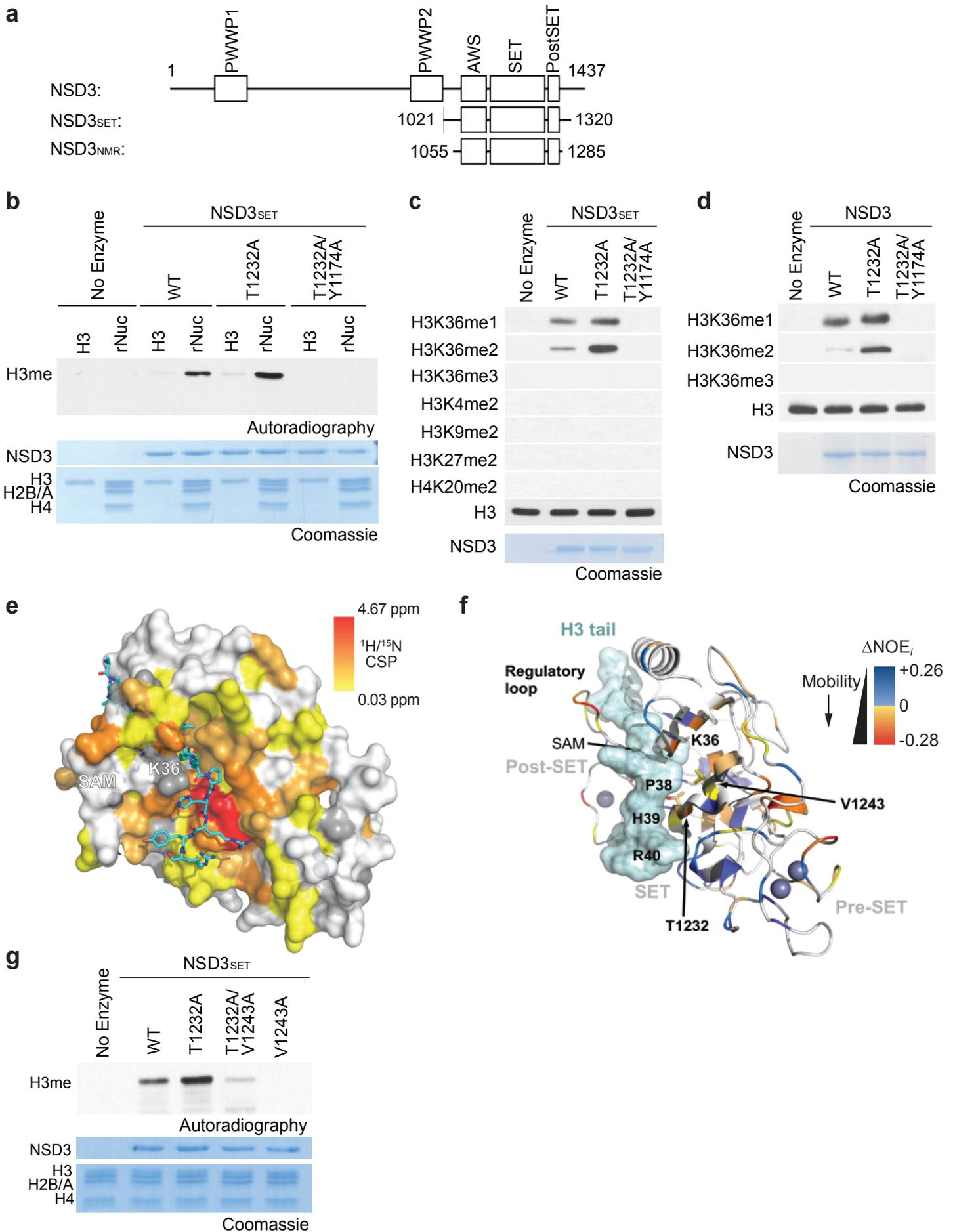
Figure 2

Figure 3. Mutant NSD3-mediated H3K36me2 synthesis promotes oncogenic programming and LUSC tumorigenesis *in vivo*. **a**, Representative μ CT 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 post Cre-induction (n = 6 mice for each experimental group). Boxes: 25th to 75th percentile, whiskers: min. to max., center: median; *** $P < 0.0001$ by two-tailed unpaired t-test. **c**, Representative HE and IHC staining of lung tissue from *PSCN* and control *PSC* mutant mice (representative of n = 8 mice for each experimental group). Scale bars: 100 μ m. **d**, Quantification of proliferation (Ki67⁺ positive cells) as in (c). Boxes: 25th to 75th percentile, whiskers: min. to max., center: median; *** $P < 0.0001$ by two-tailed unpaired t test, n = 8 mice for each experimental group. **e**, Kaplan-Meier survival curves of *PSCN* (n = 11, median survival = 145 days) and *PSC* (control, same data as shown in Fig. 1h, n = 10, median survival = 200.5 days) mutant mice. *** $P = 0.001$ by log-rank test for significance. **f**, Western blots with the indicated antibodies of LUSC tumor lysates from *PSCN* and *PSC* mutant mice. Two independent and representative samples are shown for each genotype. Tubulin is shown as a loading control. **g**, Western blot analysis with the indicated antibodies of whole cell lysates from either control (sgControl) or NSD3-depleted (sgNSD3) *PSC_N* cells. H3 and tubulin were used as loading controls. **h**, GSEA analyses of *PSCN* versus *PSC* tumor biopsies and *PSC_N* cells \pm sgNSD3. 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 as indicated in *PSC_N* cells \pm sgNSD3 as indicated. **j**, Quantification of H3K36me2 and H3K27me3 peak intensity change in *PSC_N* cells \pm sgNSD3 on the indicated gene sets. P value provided (detailed statistics description in **Methods**).

Figure 3

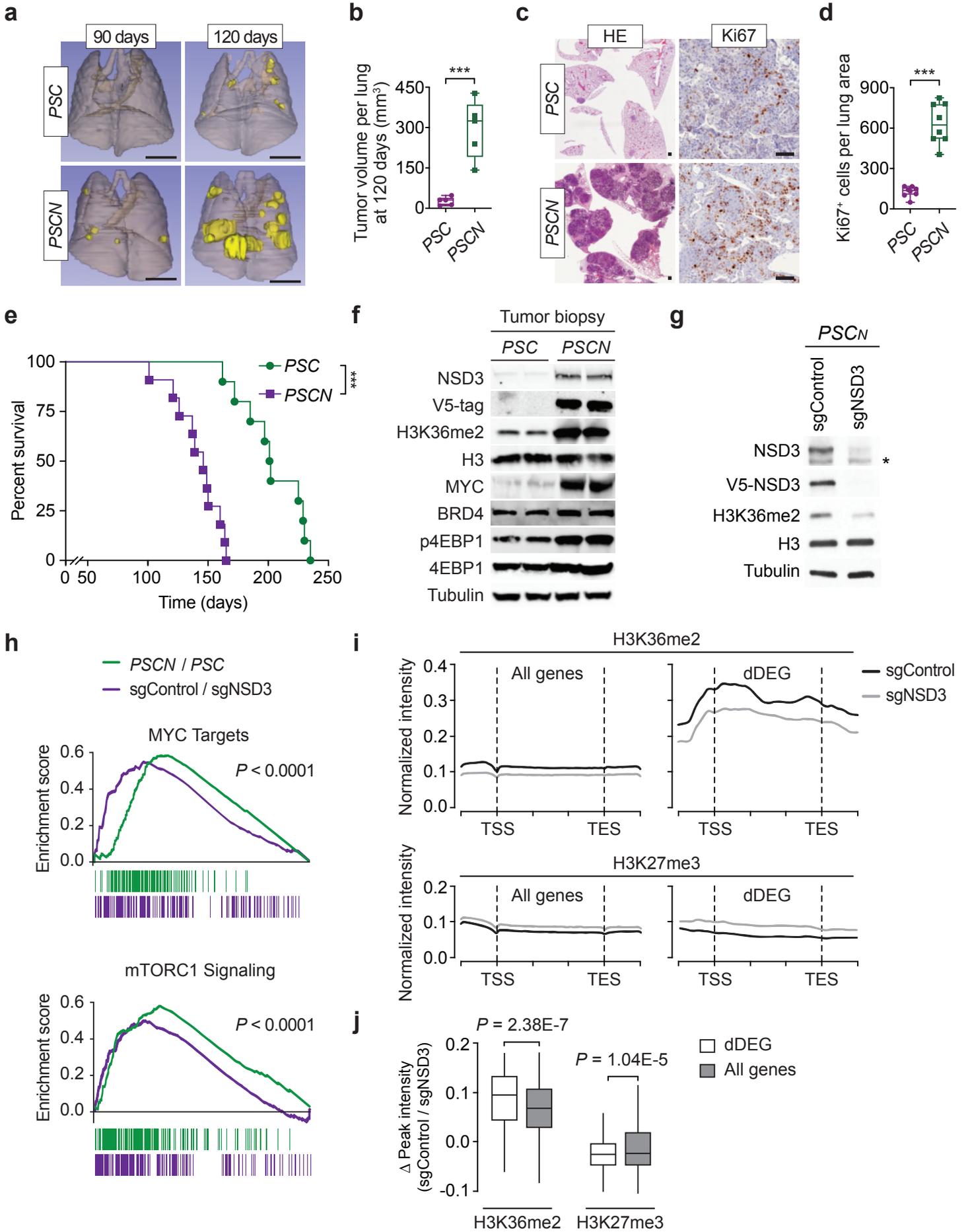
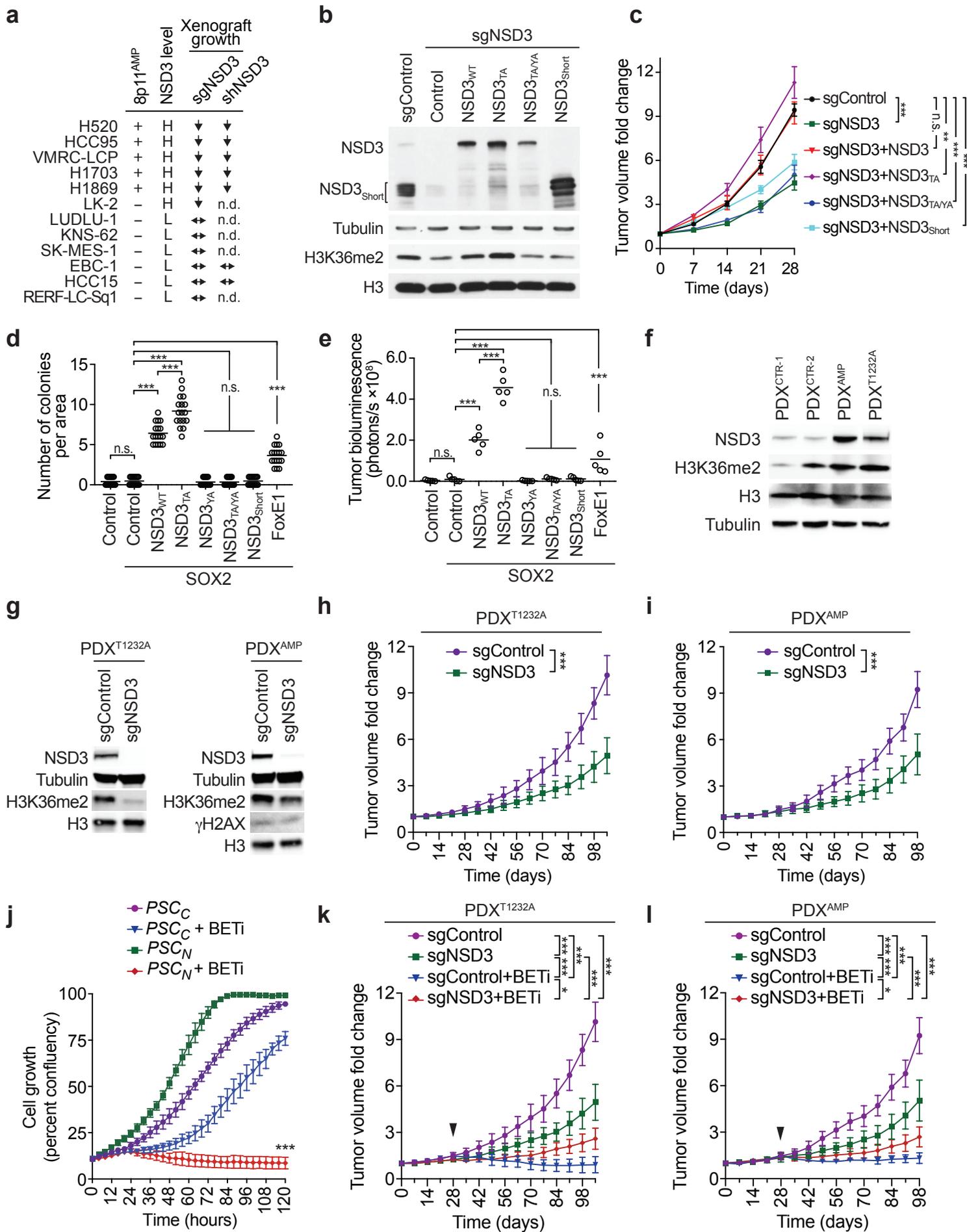


Figure 4. NSD3 promotes human lung cell transformation and xenograft tumor growth of LUSC cells and PDXs and renders PDXs hyper-susceptible to BETi. **a**, NSD3 depletion attenuates xenograft tumor growth of 8p11^{AMP} and NSD3 overexpressing LUSC cell lines. Summary of xenograft tumor growth of the indicated cell lines with the 8p11-12 amplicon status shown and 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 upon NSD3 depletion: ↓: reduced growth; ↔: no change in growth (see Extended Data Figs 7c, e); n.d., not determined. **b-c**, NSD3 catalytic activity required for full H520 xenograft tumor growth. NSD3-deficient H520 cells reconstituted with the indicated V5-tagged CRISPR-resistant NSD3 derivatives. **b**, Western blots of H520 lysates with indicated antibodies (NSD3_{T1232A}: NSD3_{TA}; NSD3_{T1232A/Y1174A}: NSD3_{TA/YA}). H3 and tubulin were used as loading controls. **c**, Tumor volume change quantification of H520 xenografts generated in (b) in immunocompromised mice (n = 6 mice, for each treatment group). ***P* < 0.002; ****P* < 0.0001, n.s., not significant by two-way ANOVA with Tukey's testing for multiple comparisons. Data are represented as mean ± s.e.m. **d-e**, NSD3_{WT} and NSD3_{TA} transform SOX2 expressing immortalized tracheobronchial epithelial (AALE) cells *in vitro* and *in vivo*. **d**, Quantification of soft agar colony formation for AALE tracheobronchial epithelial cells with ectopic expression of SOX2 and NSD3_{WT}, NSD3_{TA}, NSD3_{YA}, NSD3_{Short} or FOXE1. Data are represented as mean ± s.e.m. of three technical replicates in two independent experiments. **e**, Quantification of tumor size determined by bioluminescence for mice (n = 5) grafted under the renal capsule with AALE cells expressing plasmids as in d and expressing the AkaLuc reporter (see **Methods**). ****P* < 0.0001, n.s., not significant by two-way ANOVA with Tukey's testing for multiple comparisons. FOXE1 was used as a positive control, NSD3_{WT} and NSD3_{TA} vs. FOXE1

$P < 0.0001$. **f**, Western blots of lysates from the indicated LUSC PDX samples and using the indicated antibodies. H3 and Tubulin are shown as loading controls. **g**, Western blots as in (e) of lysates from PDX^{T1232A} (left) and PDX^{AMP} (right) \pm sgNSD3 treatment. **h-i**, Tumor volume quantification of **h**, PDX^{T1232A} and **i**, PDX^{AMP} \pm sgNSD3 treatment in immunocompromised mice ($n = 6$ mice, for each treatment group). $***P < 0.0001$ by two-tailed unpaired t-test. Data are represented as mean \pm s.e.m. **j**, Proliferation assay of PSC_C and PSC_N treated with vehicle control or 20nM AZD5153 (BETi). Data represent mean \pm s.e.m. of three technical replicates in two independent experiments. $***P < 0.0001$ by two-way ANOVA with Tukey's testing for multiple comparisons. **k-l**, Tumor volume quantification of **k**, PDX^{T1232A} and **l**, PDX^{AMP} \pm sgNSD3 and \pm BETi (2.5mg/kg, i.p.) treatments as indicated ($n = 6$ mice, for each treatment group). Control animals received vehicle (placebo) treatment. $*P < 0.033$, $***P < 0.0001$ by two-way ANOVA with Tukey's testing for multiple comparisons. Data are represented as mean \pm s.e.m. Arrowhead indicates start of the treatment.

Figure 4



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-FOXE1⁴⁶ and AkaLuc (gift from Dr. Michael Z. Lin)⁴⁷ 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. 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)⁴⁸ 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.

Antibodies. H3K36me2 (Thermo Fisher Scientific 701767); H3K36me3 (Thermo Fisher Scientific MA5-24687); H3K27me3 (Thermo Fisher Scientific MA5-33081); V5 (Thermo Fisher Scientific R960-25); H3K36me1 (Abclonal #A2364); H3K4me2 (CST #9725S); H3K9me2 (CST #4658S); H3K27me2 (CST #9728S); γ H2AX (CST #9718), SOX2 (CST #3579), c-Myc (CST #5605S); Phospho-4E-BP1 (CST #2855S); 4E-BP1 (CST #9644S); FGFR1 (CST #9740S); H4K20me2 (Abcam #ab9052); H3 (Abcam #ab1791); BRD4 (Abcam #ab128874); β -Tubulin (Millipore 05-661); NSD2⁴⁸; NSD3 (CST#92056S), for detecting both NSD3 and NSD3_{Short};

NSD3: protein fragment corresponding to amino acid residues 1021-1320 of NSD3 was used as the epitope to generate antiserum in Genemed Synthesis Inc., antiserum purification was performed with immortalized epitope for NSD3 specific antibody.

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 glutathione 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-[¹⁵N], U-[¹³C, ¹⁵N] NSD3_{SET} and NSD3_{SET-T1232A} versions were prepared from pQE30 vector by expression in M9 media. ¹⁵NH₄Cl at a concentration of 1 g/L served as sole source of nitrogen, for U-[¹³C, ¹⁵N] labeling a ¹³C-glucose (CIL) at 3 g/L was added. Recombinant nucleosomes were assembled using 601 Widom DNA⁴⁹. 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 protocol⁵⁰. Histone octamers and nucleosomes were assembled according to methods described previously⁵¹.

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 MgCl₂, 10 mM KCl to deplete cytoplasmic fractions, the pellet was treated with buffer containing 20 mM HEPES (pH = 7.9), 1.5mM MgCl₂, 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 described⁵² using 2 µg of enzyme (unless otherwise noted) and 2 µg of recombinant nucleosomes (EpiCypher, catalog#: 16-0009 and 16-2004) or recombinant histone H3 (generated as describe in ⁵³) 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 described⁵⁴. 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 methods⁵⁵. Briefly, mice were euthanized by CO₂ 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% CO₂. 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 described⁵⁶. 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 \pm 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 μ g /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 shRNA oligo sequences were utilized: sgControl CTTCGAAATGTCCGTTCCGGT, sgNSD2 ACTCGTTAACAAATTCTCCC (made by Saumya Sankaran), sgNSD3 GGATACTGATTATATGAC, sgFGFR1 TCCCCGACCTTGCTGAACA, sgPLPP5 CAATAAAACTGATCGTAGGG, shControl CCTAAGGTTAAGTCGCCCTCG, shNSD3 GCAGATTGTTTGGGTCAAATT. After 48 hours of transfection, target cells were transduced with 0.45 μ m filtered viral supernatant and 8 μ g/mL polybrene. Cells were selected 24 hours after media replacement with 2 μ g /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 μ m filtered viral

supernatant and 8 µg/mL polybrene. The subsequent selection was carried out with 250 µg/mL hygromycin.

Mice. *ROSA26^{LSL-Pik3ca(H1047R)}*, *Colla1^{LSL-Sox2}*, *Cdkn2a/b^{LoxP/LoxP}*, *Fgfr1^{LoxP/LoxP}* mice have been described before⁵⁷⁻⁶⁰. Reporter-tagged insertion with conditional potential *Nsd3^{tm1a(EUCOMM)}* mouse strain originates from ES clone #HEPD0563-4-G02 obtained from European Mouse Mutant Archive repository⁶¹. 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 strain⁶² 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 described⁶³. 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 fed a standard chow diet *ad*

libitum and housed in pathogen-free facility with standard controlled temperature, humidity, and light-dark cycle (12 h) conditions with no more than 5 mice per cage under the supervision of veterinarians, in an AALAC-accredited animal facility at the University of Texas M.D. Anderson Cancer Center. All animal procedures were reviewed and approved by the MDACC Institutional Animal Care and Use Committee (IACUC 00001636, PI: Mazur).

Lung Squamous Cell Carcinoma Mouse Models. To generate tumors in the lungs of *ROSA26^{LSL-Pik3ca(H1047R)}*, *Colla1^{LSL-Sox2}*; *Cdkn2a/b^{LoxP/LoxP}* (PSC), *ROSA26^{LSL-Pik3ca(H1047R)}*; *Colla1^{LSL-Sox2}*; *Cdkn2a/b^{LoxP/LoxP}*; *Fgfr1^{LoxP/LoxP}* (PSC^{FGFR1-KO}), *ROSA26^{LSL-Pik3ca(H1047R)}*; *Colla1^{LSL-Sox2}*; *Cdkn2a/b^{LoxP/LoxP}*; *Nsd3^{LoxP/LoxP}* (PSC^{NSD3-KO}), *ROSA26^{LSL-Pik3ca(H1047R)}*; *Colla1^{LSL-Sox2}*; *Cdkn2a/b^{LoxP/LoxP}*; *LSL-Nsd3^{T1242A}* (PSCN) mutant mice, we used replication-deficient adenoviruses expressing Cre-recombinase (Ad-Cre) as previously described⁶⁴. 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 previously⁵⁶. Genome-wide copy number analysis was performed for each PDX using an OncoScan CNV Assay Kit (Thermo Fisher Scientific) as previously described⁶⁵. 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 puromycin 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 μ l 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. Tumor volume was calculated using the formula: $\text{Volume} = (\text{width})^2 \times \text{length} / 2$ where *length* represents the largest tumor diameter and *width* represents the perpendicular tumor diameter.

Tumorigenicity Assay *in vitro* and *in vivo*. Immortalized human tracheobronchial epithelial (AALE) were described before⁴⁶. 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 – AALE cells were transplanted bilaterally under sub-renal capsule into immunocompromised NSG mice using protocol previously described⁶⁶. The cells resuspended in Matrigel (10^6 in 40 μ l) 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^{47,67}. 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 ($\text{photons/s/cm}^2/\text{sr}$) and plotted as mean \pm 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 described⁶⁸. 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 x 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 μCT 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 $93 \mu\text{m}^3$ 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 μ m 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 before⁶⁴. 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 describe⁶⁹. 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 μ M concentration, NSD3_{SET} or NSD3_{SET-T1232A} domains at 100-125 μ M, and ITC measurements were performed in a MicroCal PEAQ-ITC instrument (Malvern Panalytical) at 20°C with 55 injections of 0.7 μ L 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) D₂O/H₂O, 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-[¹³C,¹⁵N] NSD3_{SET} or NSD3_{SET-T1232A} samples using transverse relaxation optimized spectroscopy, TROSY-based triple resonance three-dimensional (3D) experiments of HNC0, HNcaCO, HNCA, HNcoCA, HNCACB, CBCAcoNH with further support from TROSY-version of 3D ¹⁵N-edited NOESY (100 ms mixing time) recorded on 370 μM U-[¹⁵N] samples⁷⁰⁻⁷². 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 software⁷³. The mobility of the protein backbone was studied by NMR ¹⁵N spin relaxation measurements using TROSY-based ¹⁵N[¹H] nuclear Overhauser effect (NOE) experiments optimized for high-molecular weight proteins on 370 μM uniformly labeled U-[¹⁵N] NSD3_{SET} or NSD3_{SET-T1232A} preparations containing variants⁷⁴. The recycle delay for ¹⁵N[¹H] NOE was 10 s and 4 s of ¹H saturation. The errors of ¹⁵N[¹H] 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 ¹H/¹⁵N resonances was determined with the

equation: $CSP_i = [(\Delta\delta_{H,i})^2 + 0.25 \cdot (\Delta\delta_{N,i})^2]^{0.575}$, where $\Delta\delta_{H,i}$ and $\Delta\delta_{N,i}$ 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 program⁷⁶ 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 structure⁷⁷ with the G¹²⁶⁷-N¹²⁶⁸ residues modelled which are lacking from the experimental electron density in the SwissModel⁷⁸. 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^{79,80}. 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 (S³² to A³²) and used to define ambiguous interaction restraints (AIRs) driving the docking process. Remaining parameters were treated according to standard docking protocol⁷⁶. The NSD3_{SET} with the regulatory loop in the closed conformation (auto-inhibited state) was modelled using SwissModel⁷⁸ and the crystal structure of NSD2_{SET} [PDB id: 5LSU:A; fragment 983-1203]⁸¹ 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)⁸². Then the transcriptomes were annotated to GENCODE M23 (gencodegenes.org)⁸³ by FeatureCounts⁸⁴. Differentially expressed genes (DEGs) were detected by DESeq2 package from Bioconductor (bioconductor.org/packages/release/bioc/html/DESeq2)⁸⁵ using likelihood ratio test (LRT, adjusted *P*-value < 0.001). The expression levels of genes were normalized and log₂-transformed, and the variances-stabilizing transformation were implemented. Bioconductor package EnhancedVolcano (bioconductor.org/packages/release/bioc/html/EnhancedVolcano) 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)⁸⁶ and the Bioconductor package (bioconductor.org/packages/release/bioc/html/GeneAnswers) GeneAnswers. The comparisons of expression levels of DEGs and all genes were implemented by robust t-test, and the trim level was 0.1.

CUT&RUN. This assay was performed as described previously with modifications⁸⁷ 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 μ m strainer before proceeding to next step. Concanavalin A beads were equilibrated in ice-cold binding buffer and mixed with the isolated nuclei (10 μ l 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 μ L were aliquoted into each sample tube. While gently vortexing, 100 μ L of 2x 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 μ L of ice-cold WB. While gently vortexing, 25 μ L of 2x 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 μ L of ice cold WB. While gently vortexing, 3 μ L of pre-chilled 100 μ M 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 μ L of 2x STOP buffer. To isolate the digested DNA through 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 μ L of 10% SDS and 2.5 μ L 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/)⁸⁸. The reads were trimmed by Trimmomatic⁸⁹, then aligned to the mouse genome mm10 by Bowtie2 (bio.sourceforge.net/bowtie2/)⁹⁰. The duplicate reads at same locations were removed by MarkDuplicates of Picard package (broadinstitute.github.io/picard). 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)⁹¹. 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 (Δ Peaks), the differences between them calculated, and compared to the distribution of Δ Peaks of DEGs and the genome-wide genes with robust t-test at the trim level 0.1. The intergenic region annotation was generated by Complement of BedTools⁹², and then the average profile of the intergenic regions was visualized by ngs.plot.

ChIP-qPCR. ChIP assays were performed as described⁴⁸. 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.

References:

- 46 Bass, A. J. *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).
- 47 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).
- 48 Kuo, A. J. *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).
- 49 Lowary, P. T. & 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).
- 50 Cheema, M. S. & 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).
- 51 Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol Biol* **119**, 1-16, doi:10.1385/1-59259-681-9:1 (1999).
- 52 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).
- 53 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).

- 54 Mazur, P. K. *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).
- 55 Edelman, B. L. & Redente, E. F. Isolation and Characterization of Mouse Fibroblasts. *Methods Mol Biol* **1809**, 59-67, doi:10.1007/978-1-4939-8570-8_5 (2018).
- 56 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).
- 57 Adams, J. R. *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).
- 58 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).
- 59 Krimpenfort, P. *et al.* p15Ink4b is a critical tumour suppressor in the absence of p16Ink4a. *Nature* **448**, 943-946, doi:10.1038/nature06084 (2007).
- 60 Hoch, R. V. & 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).
- 61 Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337-342, doi:10.1038/nature10163 (2011).
- 62 Raymond, C. S. & Soriano, P. High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS One* **2**, e162, doi:10.1371/journal.pone.0000162 (2007).
- 63 Chu, V. T. *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).
- 64 Mazur, P. K. *et al.* SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. *Nature* **510**, 283-287, doi:10.1038/nature13320 (2014).
- 65 Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359-364, doi:10.1038/nature20788 (2017).
- 66 Shultz, L. D. *et al.* Subcapsular transplantation of tissue in the kidney. *Cold Spring Harb Protoc* **2014**, 737-740, doi:10.1101/pdb.prot078089 (2014).
- 67 Iwano, S. *et al.* Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science* **359**, 935-939, doi:10.1126/science.aag1067 (2018).
- 68 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).
- 69 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).
- 70 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).
- 71 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).

- 72 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).
- 73 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).
- 74 Lakomek, N. A., 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).
- 75 Williamson, M. P. 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).
- 76 van Zundert, G. C. P. *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).
- 77 Morrison, M. J. *et al.* Identification of a peptide inhibitor for the histone methyltransferase WHSC1. *PLoS One* **13**, e0197082, doi:10.1371/journal.pone.0197082 (2018).
- 78 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).
- 79 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).
- 80 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).
- 81 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/acscchembio.6b00308 (2016).
- 82 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).
- 83 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).
- 84 Liao, Y., Smyth, G. K. & 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).
- 85 Love, M. I., 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).
- 86 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).
- 87 Hainer, S. J. & Fazio, T. G. High-Resolution Chromatin Profiling Using CUT&RUN. *Curr Protoc Mol Biol* **126**, e85, doi:10.1002/cpmb.85 (2019).
- 88 Zhu, Q., Liu, N., Orkin, S. H. & Yuan, G. C. CUT&RUNTools: a flexible pipeline for CUT&RUN processing and footprint analysis. *Genome Biol* **20**, 192, doi:10.1186/s13059-019-1802-4 (2019).
- 89 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

- 90 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 91 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).
- 92 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).