Article

Elevated NSD3 histone methylation activity drives squamous cell lung cancer

https://doi.org/10.1038/s41586-020-03170-y

Received: 30 April 2020

Accepted: 23 December 2020


Check for updates

Gang Yuan[1,13], Natasha M. Flores[2,13], 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[9,10,11], Ning-Yi Shao[12]✉, Łukasz Jaremko[3]✉, Pawel K. Mazur[2]✉ & Or Gozani[1]✉

Amplification of chromosomal region 8p11–12 is a common genetic alteration that has been implicated in the aetiology of lung squamous cell carcinoma (LUSC)[1–3]. The *FGFR1* gene is the main candidate driver of tumorigenesis within this region[4]. However, clinical trials evaluating FGFR1 inhibition as a targeted therapy have been unsuccessful[5]. Here we identify the histone H3 lysine 36 (H3K36) methyltransferase NSD3, whose gene is located in the 8p11–12 amplicon, as a key regulator of LUSC tumorigenesis. In contrast to other 8p11–12 candidate LUSC drivers, increased expression of *NSD3* correlated strongly with its gene amplification. Ablation of *NSD3*, but not of *FGFR1*, attenuated tumour growth and extended survival in a mouse model of LUSC. We identify an LUSC-associated variant NSD3(T1232A) that shows increased catalytic activity for dimethylation of H3K36 (H3K36me2) in vitro and in vivo. Structural dynamic analyses revealed that the T1232A substitution elicited localized mobility changes throughout the catalytic domain of NSD3 to relieve auto-inhibition and to increase accessibility of the H3 substrate. Expression of NSD3(T1232A) in vivo accelerated tumorigenesis and decreased overall survival in mouse models of LUSC. Pathological generation of H3K36me2 by NSD3(T1232A) reprograms the chromatin landscape to promote oncogenic gene expression signatures. Furthermore, NSD3, in a manner dependent on its catalytic activity, promoted transformation in human tracheobronchial cells and growth of xenografted human LUSC cell lines with amplification of 8p11–12. Depletion of NSD3 in patient-derived xenografts from primary LUSCs containing *NSD3* amplification or the NSD3(T1232A)-encoding variant attenuated neoplastic growth in mice. Finally, NSD3-regulated LUSC-derived xenografts were hypersensitive to bromodomain inhibition. Thus, 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 bromodomain inhibition.

Q1

Many genes within the 8p11–12 amplicon could potentially contribute to oncogenesis (Extended Data Fig. 1a). However, the minimal region of amplification across many neoplasms with the 8p11–12 amplicon centres around *FGFR1* and the neighbouring gene and candidate oncogene *NSD3*[3,4,6,7]. Analysis of LUSC datasets from The Cancer Genome Atlas (TCGA) shows that *NSD3* and its immediate neighbours (for example, *FGFR1*) are the most commonly amplified 8p11–12 genes (Extended Data Fig. 1a). Overall, amplification of the genomic area spanning *NSD3* and *FGFR1* is one of the more common molecular alterations in LUSC (Extended Data Fig. 1b). Notably, for *NSD3*, gene amplification correlates strongly with increased mRNA expression; by contrast, there is little correlation between *FGFR1* gene copy number and mRNA expression[2,3] (Fig. 1a, Extended Data Fig. 1a, b). Accordingly, depletion of NSD3, but not depletion of FGFR1 or PLPP5 (encoded by the gene immediately adjacent to *NSD3*), in the 8p11–12-amplified (8p11^{AMP}) H520 LUSC cell line inhibited xenograft tumour growth (Fig. 1b, Extended

[1]Department of Biology, Stanford University, Stanford, CA, USA. [2]Department of Experimental Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [3]Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. [4]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [5]Geriatric Research, Education, and Clinical Center, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA. [6]University of Navarra, Center for Applied Medical Research, Pamplona, Spain. [7]Department of Thoracic and Cardiovascular Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [8]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [9]Hamon Center for Therapeutic Oncology Research, University of Texas Southwestern Medical Center, Dallas, TX, USA. [10]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA. [11]Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX, USA. [12]Faculty of Health Sciences, University of Macau, Macau SAR, China. [13]These authors contributed equally: Gang Yuan, Natasha M. Flores. ✉e-mail: nshao@um.edu.mo; lukasz.jaremko@kaust.edu.sa; pkmazur@mdanderson.org; ogozani@stanford.edu

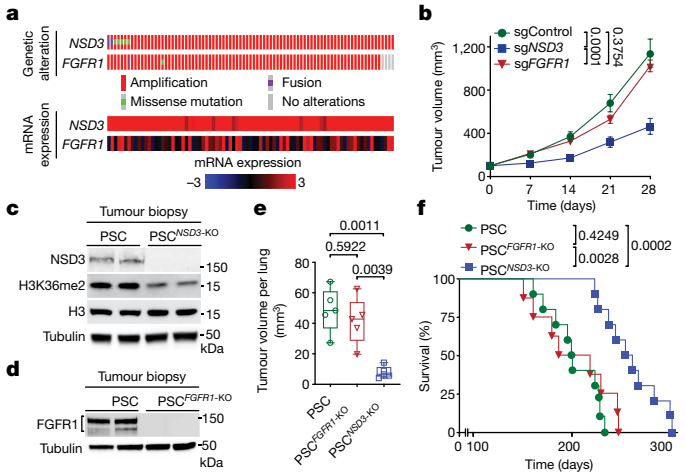


Fig. 1 | Deletion of *NSD3*, but not *FGFR1*, inhibits LUSC tumorigenesis in vivo. a, Analysis of 8p11–12-amplified LUSC datasets from TCGA indicates that increased mRNA expression (*Z*-score) of *NSD3*, but not *FGFR1*, correlates with gene amplification ($n = 85$ patients). **b**, Depletion of *NSD3*, but not *FGFR1*, attenuates xenograft growth of the 8p11^{AMP} LUSC cell line H520. Tumour volume quantification of H520 xenografts in immunocompromised mice ($n = 5$ mice for each group). **c**, **d**, Western blots of LUSC tumour lysates from PSC (control), PSC$^{NSD3\text{-}KO}$ (**c**) and PSC$^{FGFR1\text{-}KO}$ (**d**) mice with the indicated antibodies. Two independent and representative samples are shown for each genotype. Tubulin used as a loading control. **e**, Micro-computed tomography (μCT) analysis of tumour volume of the indicated mouse models ($n = 5$ mice for each group). Centre line, median; box limits, 75th and 25th percentiles; 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$^{FGFR1\text{-}KO}$ ($n = 8$, median survival = 202.5 days) and PSC$^{NSD3\text{-}KO}$ ($n = 10$, median survival = 257 days) mice; *P* values determined by log-rank test.

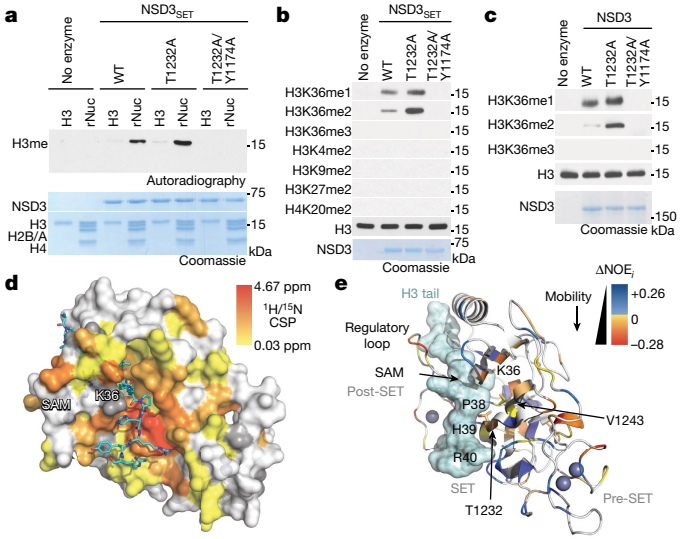


Fig. 2 | Molecular basis of increased H3K36me2 catalysis by NSD3(T1232A). a, In vitro methylation reactions of wild-type (WT) or mutant NSD3$_{SET}$ using recombinant free histone H3 or nucleosome (rNuc) and ^{3}H-SAM as substrates. No enzyme is a negative control. Top, autoradiography; bottom, Coomassie blue staining. **b**, Methylation assays on rNuc as in **a** with non-radiolabelled SAM. Western blots 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 in the NSD3 catalytic domain. **d**, Amide chemical shift perturbations between NSD3$_{SET\text{-}T1232A}$ and NSD3$_{SET}$ are mapped onto the surface representation of the NSD3$_{SET}$ structure (PDB: 6CEN) with docked H3.1 residues A29 to R42 (stick representation). Grey, 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. Uncoloured residues, undetectable NOE changes. *i*, individual peaks.

Data Fig. 2a–e). These data point to NSD3, rather than FGFR1, as the mutational driver within the 8p11–12 amplicon in LUSC.

We established a robust mouse model of LUSC consisting of canonical LUSC alterations co-occurring with 8p11–12 amplification (constitutively active PI3K, overexpression of SOX2, and deletion of *CDKN2A* and *CDKN2B*, here named PSC mice; see Extended Data Fig. 1b and Methods). PSC mice develop with high penetrance lung tumours characterized by LUSC-defining molecular hallmarks[8] (Extended Data Fig. 2f). In this model, increased NSD3 expression tracks with tumour progression, consistent with NSD3 overexpression being observed in about 60% of LUSC samples (beyond the 20% that harbour 8p11^{AMP}) and frequently co-occurring with PSC molecular alterations (Extended Data Fig. 2g–i).

Conditional *Fgfr1*$^{loxP/loxP}$ and *Nsd3*$^{loxP/loxP}$ mutant mice were generated. The mice were viable and fertile, and developed normally (see Methods). In the PSC background, lung-specific deletion of *Nsd3* (PSC$^{NSD3\text{-}KO}$; Fig. 1c), but not *Fgfr1* (PSC$^{FGFR1\text{-}KO}$; Fig. 1d), significantly attenuated tumour growth and cancer cell proliferation, while increasing apoptosis (Fig. 1e, Extended Data Fig. 2j–l). In survival studies, deletion of *Nsd3* extended the lifespan of PSC mice by about 30%, whereas *Fgfr1* knockout had no effect (Fig. 1f). Together, these data support a key in vivo role for NSD3 in LUSC tumorigenesis.

NSD3(T1232A) is a hyperactive variant

NSD3, NSD1, NSD2, and ASH1L are the four enzymes in mammals that specifically synthesize the euchromatin-associated H3K36me2 modification[9]. Depending on cell type, NSD1 and NSD2 are the main H3K36me2-generating enzymes, whereas the physiological context in which NSD3 regulates H3K36me2 is less clear[9]. Tumours lacking NSD3 showed lower global H3K36me2 than control tumours (Fig. 1c), suggesting an aetiological role for NSD3-catalysed H3K36me2 synthesis in *NSD3*-amplified LUSCs. We reasoned that we could test this idea in mice by transgenic expression of a hyperactive NSD3 variant, which would model the consequences of increased NSD3 catalysis resulting from elevated *NSD3* dosage on tumorigenesis. A recurrent hyperactivating NSD2 mutation is present in various cancers[9], suggesting that a functionally analogous cancer-associated mutation may be present in NSD3. We tested in vitro histone methylation activity for 35 TCGA-documented mutations within the NSD3 catalytic domain (Extended Data Fig. 3a–c). The variant with the highest activity was the T1232A substitution (NSD3(T1232A)), a recurrent cancer-associated mutation[10].

A recombinant NSD3 catalytic domain harbouring the T1232A substitution (NSD3$_{SET}$(T1232A)) shows stronger in vitro methylation activity on nucleosomes than does the wild-type domain (NSD3$_{SET}$), but this higher activity is abrogated when the substitution is paired with a catalytic mutant (NSD3$_{SET}$(T1232A/Y1174A); Fig. 2a, b, Extended Data Fig. 3d). Furthermore, NSD3$_{SET}$(T1232A) was more efficient than the wild-type enzyme in selectively generating H3K36me2 (Fig. 2b). Despite its higher activity, NSD3$_{SET}$(T1232A) behaved similarly to NSD3$_{SET}$ with respect to (i) nucleosomal versus free H3 substrate preference (Fig. 2a); (ii) methylation-state specificity, generating H3K36me1 or H3K36me2 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) behaved like the isolated catalytic domain in being hyperactive while retaining the

same catalytic selectivity profile (Fig. 2c). Finally, overexpression of NSD3(T1232A), but not wild-type NSD3 or NSD3(T1232A/Y1174A), increased 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 $NSD3_{SET}$ was more evident in nucleosomes with longer linker DNA (187 versus 147 base pairs; Extended Data Fig. 3f). The increase in catalysis, irrespective of linker DNA, was not mediated by the T1232A substitution significantly altering the binding affinity of NSD3 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 backbones of $NSD3_{SET}$(T1232A) and $NSD3_{SET}$ in solution showed that the static structure is largely unaffected by the T1232A substitution (Extended Data Fig. 4b, c, Supplementary Table 1). By contrast, we found chemical shift perturbations of the backbone amide $^{1}H/^{15}N$ resonances, which serve as sensitive probes for local structural and allosteric effects, throughout the SET domain (Fig. 2d, Extended Data Fig. 4c, d). The highest chemical shift perturbations were centred around T1232A, with effects extending throughout the substrate-binding cleft and including the conserved NSD3 auto-regulatory loop[11,12] (Fig. 2d, Extended Data Fig. 4d, e). 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 Fig. 4g, h, Supplementary Table 1). These data suggest that T1232A increases activity by enhancing the mobility of the auto-regulatory loop, which destabilizes the inhibitory state, acting together with the active site being rendered more accessible to the H3 substrate and direct, localized modification of the H3 binding surface. Indeed, the loss of the polar hydroxyl group in the T1232A mutant increases local hydrophobicity, which can alter H3 anchoring dynamics[13] (Fig. 2d, e). 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, the methylation activity of a double alanine mutant ($NSD3_{SET}$(T1232A/V1243A)) was similar to that of the wild-type enzyme and lower than that of the single mutant $NSD3_{SET}$(T1232A), while $NSD3_{SET}$(V1243A) showed no activity and decreased thermal stability (Extended Data Fig. 4i, Supplementary Table 1). Thus, T1232A substitution induces changes in local dynamics at many residues within three key functional regions of the NSD3 catalytic domain that are likely to work together to increase catalysis of H3K36 methylation.

H3K36me2 deregulation accelerates LUSC

We used NSD3(T1232A) to model the consequences of elevated catalytic activity that would result from *NSD3* amplification for LUSC tumorigenesis in vivo. Mouse and human NSD3 are highly similar (94% overall; 98% in the catalytic domain), with mouse NSD3 (mNSD3) T1242 corresponding to human T1232 (Extended Data Fig. 5a). We generated Cre-inducible V5-tagged mNSD3(T1242A) mice and verified transgene expression in lung fibroblast cells transduced with Ad-Cre ex vivo (Extended Data Fig. 5b–e). In the PSC background, expression of mNSD3(T1242A) (here referred to as PSCN mice) accelerated tumour growth (Fig. 3a, b) and proliferation (Fig. 3c, d) while decreasing apoptosis (Extended Data Fig. 5f, g) relative to the control PSC group. For example, 90 days after Cre induction, when PSC mice have no detectable tumours, multiple growths were observed in the PSCN cohort (Fig. 3a). At 120 days, the tumour burden in PSCN mice was more than an order of magnitude greater than in control PSC mice (Fig. 3a, b). Consistent with these data, expression of mNSD3(T1242A) reduced lifespan in PSC mice by about 30% (Fig. 3e). In independent LUSC tumour biopsies, high NSD3 levels (due to mNSD3(T1242A) expression) resulted in markedly elevated H3K36me2 (Fig. 3f, Extended Data Fig. 5f). Moreover, mNSD3(T1242A)-driven tumours had higher levels of the oncoproteins

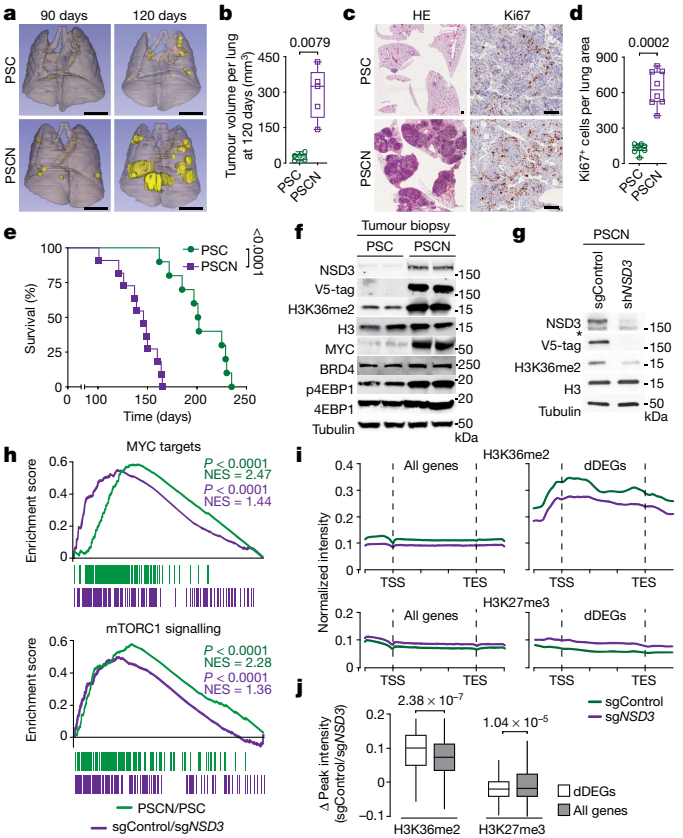


Fig. 3 | Mutant NSD3-mediated H3K36me2 synthesis promotes oncogenic programming and LUSC tumorigenesis. a, Representative μCT scans at 90 and 120 days after adenoviral Cre-mediated tumour induction in PSC and PSCN mice. Scale bars, 10 mm. **b**, Quantification of tumour volume as in **a** at 120 days ($n = 5$ mice for each group). **c**, Representative haematoxylin and eosin (HE) and immunohistochemical staining (Ki67) 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) mice; *P* value determined by log-rank test. **f**, Western blots of LUSC tumour lysates from PSCN and PSC mice with the indicated antibodies. Two independent and representative samples are shown for each genotype. **g**, Western blots of cell lysates from control (sgControl) or NSD3-depleted (sg*NSD3*) PSC_N cells with the indicated antibodies. H3 and tubulin shown as loading controls in **f**, **g**. **h**, GSEA identifies upregulation of hallmark gene sets (MYC targets and mTOR signalling (MSigDB: M5928, M5924)) in RNA sequencing (RNA-seq) data from PSCN versus PSC tumour biopsies and PSC_N cells with or without sg*NSD3* ($n = 3$ biologically independent samples per group). Normalized enrichment scores (NES) and nominal *P* values are provided (see Methods). **i**, CUT&RUN profiles of H3K36me2 (top) and H3K27me3 (bottom) over averaged gene body for all genes or dDEGs in PSC_N cells with or without sg*NSD3*. **j**, Quantification of H3K36me2 and H3K27me3 peak intensity change in PSC_N cells with or without sg*NSD3* on the indicated gene sets. dDEGs $n = 234$, all genes $n = 16{,}091$. *P* values calculated by two-sided *t*-test.

MYC and BRD4 and phosphorylated 4EBP1 (a proxy for mTOR pathway activation; Fig. 3f). Together, these data indicate that the activity of NSD3 substantially affects the clinical course and outcome of LUSC in vivo.

Comparison of cell lines derived from mNSD3(T1242A) and control PSC matched tumours (PSC_N and PSC_C, respectively) showed that H3K36me2 levels were higher in cells expressing mNSD3(T1242A), as was the case in tumours (Extended Data Fig. 5h). Furthermore, depletion of NSD3 in PSC_N cells resulted in a reduction in H3K36me2 levels (Fig. 3g). Transcriptomic data show that the effect of mNSD3(T1242A) expression status on the number of differentially expressed genes

(DEGs) is, as expected, far greater when comparing PSCN to PSC tumour biopsies than for PSC_N cells with or without NSD3 depletion, in which NSD3 was depleted for a relatively short time before the analysis (Extended Data Fig. 5i, j; Supplementary Tables 2, 3). Nonetheless, the gene set enrichment analysis (GSEA) profiles of the tumour and cellular data sets show highly significant overlap with the same pathways, including strong correlations with MYC targets and mTOR signalling signatures (Fig. 3h, Extended Data Fig. 5k). We note that regulation of these genes (for example, *Prkaa2*, *Myc* and *Irgm1*) was observed in independent NSD3-depleted cell lines and depended on NSD3 catalytic activity (Extended Data Fig. 5l, m).

Next, we used cleavage under targets and release using nuclease (CUT&RUN)[14] to profile the genomic distribution of H3K36me2 in PSC_N cells with or without NSD3 depletion. We also mapped H3K27me3, as the deposition of this important chromatin modification is directly antagonized by H3K36 methylation[9]. An averaged genome-wide distribution of H3K36me2 in PSC_N cells across all gene bodies showed 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 a gene (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). By contrast, the NSD3-dependent distribution of H3K36me2 specifically at the gene bodies of downregulated DEGs (dDEGs), the most likely direct targets of NSD3, diverged from the overall genome-wide pattern in two ways: (i) the baseline signal was 2–3-fold higher; and (ii) the signal distribution across the gene was distinct, with the most intense peak being 3′ rather than 5′ proximal to the TSS (Fig. 3i, Extended Data Fig. 6a). Furthermore, quantification of the difference between H3K36me2 peak signal intensities with or without NSD3 depletion revealed that the magnitude of peak change at dDEG bodies was significantly greater than the change observed genome-wide (Fig. 3j). Notably, relative to other genes, dDEGs showed 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 was 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 (Fig. 3i, j). The inverse relationship between H3K36me2 and H3K27me3 in relation to NSD3 status extended to intergenic regions (Extended Data Fig. 6c, d). The pattern of diminished H3K36me2 and increased H3K27me3 upon NSD3 depletion was also observed at the level of individual dDEG tracks (Extended Data Fig. 6e). Finally, using chromatin immunoprecipitation (ChIP) assays, we observed a reduction in occupancy of mNSD3(T1242A) and H3K36me2 at target genes (*Prkaa2* and *Irgm1*) upon NSD3 depletion; this was reconstituted to control levels by complementation with catalytically active, but not catalytically dead, NSD3 (Extended Data Fig. 6f–h). Together, these data suggest that aberrant H3K36me2 synthesis by NSD3 directly regulates a 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, and used them to generate xenografts in mice (Fig. 4a, Extended Data Fig. 7a). Depletion of NSD3 impaired the formation of tumours from the $8p11^{AMP}$ or NSD3-overexpressing human LUSC cell lines but did not affect tumour growth from the six control cell lines (Fig. 4a, Extended Data Fig. 4b). The phenotype of the five $8p11^{AMP}$ LUSC-derived tumours is unlikely to result from non-specific DNA damage due to sgNSD3-targeted cutting within amplified regions[15], because: (i) sgRNA-mediated depletion of FGFR1 and PLPP5 had no effect on xenograft growth in $8p11^{AMP}$-positive

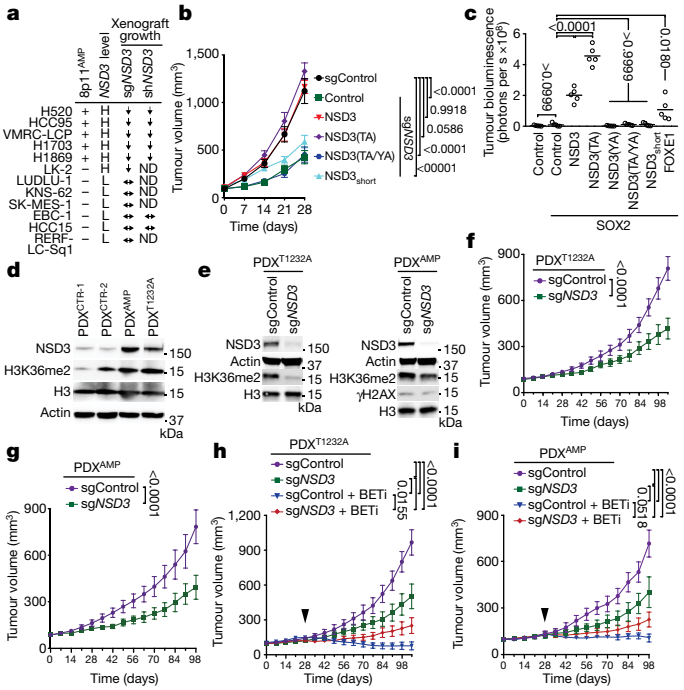


Fig. 4 | NSD3 promotes human lung cell transformation and xenograft tumour growth of LUSC cells and PDXs, and renders PDXs hypersensitive to BETi. a, NSD3 depletion attenuates xenograft tumour growth of $8p11^{AMP}$ and NSD3-overexpressing LUSC cell lines. Summary of xenograft tumour growth for the indicated cell lines treated with sg*NSD3* or a ahort hairpin RNA targeting *NSD3* (sh*NSD3*). *NSD3* levels: H, high mRNA (*Z*-score > 1.0); L, low mRNA (*Z*-score < 1.0). Scoring was consistent with relative protein levels (Extended Data Fig. 7a, b). Xenograft tumour growth: ↓, reduced; ↔, no change (Extended Data Fig. 7b, c); ND, not determined. **b**, NSD3 catalytic activity is required for full H520 xenograft tumour growth. Tumour volume change of H520 xenografts reconstituted with the indicated V5-tagged CRISPR-resistant NSD3 derivatives in immunocompromised mice ($n = 5$ mice, for each group).TA, T1232A; YA, Y1174A. **c**, NSD3 and NSD3(T1232A) transform SOX2-expressing AALE cells in vivo. Quantification of tumour 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 and NSD3(T1232A) versus FOXE1, $P < 0.0001$. Horizontal 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} and PDX^{AMP} cells with or without sg*NSD3* treatment. **f**, **g**, PDX tumour volumes of PDX^{T1232A} (**f**) and PDX^{AMP} tumours (**g**) with or without sg*NSD3* in immunocompromised mice ($n = 5$ for each group). **h**, **i**, Tumour volumes of PDX^{T1232A} (**h**) and PDX^{AMP} tumours (**i**) with or without sgNSD3 and with or without BETi (AZD5153; 2.5 mg kg^{-1}, intraperitoneal) 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**).

H520 cells (Extended Data Fig. 2e); (ii) γH2AX levels were not altered by depletion of NSD3 in $8p11^{AMP}$ cell lines (Extended Data Figs. 2d, 7b); and (iii) depletion of NSD3 in the $8p11^{AMP}$ LUSC cell lines by RNA interference (RNAi) also attenuated xenograft growth (Fig. 4a, Extended Data Fig. 7c). Finally, complementing NSD3-depleted H520 cells with wild-type NSD3 or NSD3(T1232A) reconstituted xenograft tumour growth, whereas catalytically dead NSD3 or a short NSD3 isoform lacking the catalytic domain ($NSD3_{short}$) that has been 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 wild-type NSD3 or NSD3(T1232A) was more efficient in transforming AALE cells than SOX2 with FOXE1, with SOX2– NSD3(T1232A) showing the greatest promotion of anchorage-independent growth in vitro (Extended Data Fig. 7e–g) and tumour growth in vivo (Fig. 4c, Extended Data Fig. 7e, h). By contrast, catalytically dead NSD3 and $NSD3_{short}$ did not promote AALE transformation (Fig. 4c, Extended Data Fig. 7e–h). Together, these data suggest that NSD3, via its catalytic activity, promotes tumour growth of human LUSC cells and oncogenic transformation of human AALE cells.

To further investigate the roles of the NSD3–H3K36me2 axis in human cancer, we screened a collection of 37 LUSC patient-derived xenograft (PDX) samples for NSD3 alterations. Of these, one contained NSD3(T1232A) (PDX^{T1232A}), twelve had amplification of *NSD3* (PDX^{AMP}), and the others had no NSD3 alteration (PDX^{CTR}) (Extended Data Fig. 8 and data not shown). We selected four PDXs (PDX^{T1232A}, $PDX^{AMP\text{-}1}$, and two PDX^{CTR}s), which apart from the NSD3 alterations had similar mutational profiles (for example, PSC signature), for further analysis. In comparison to PDX^{CTR} samples, NSD3 levels were higher in PDX^{T1232A} cells and highest in $PDX^{AMP\text{-}1}$ cells (Fig. 4d). H3K36me2 levels in the PDX^{T1232A} and $PDX^{AMP\text{-}1}$ cells were comparable to one another and elevated relative to controls (Fig. 4d); this pattern of similar H3K36me2 abundance in PDX^{T1232A} and $PDX^{AMP\text{-}1}$ cells, while the former has lower NSD3 levels, is consistent with hyperactivity of NSD3(T1232A). NSD3 depletion, which resulted in loss of H3K36me2, attenuated the growth of PDX^{T1232A} and $PDX^{AMP\text{-}1}$ xenograft tumours (Fig. 4e–g) but did not affect the growth of PDX^{CTR} samples (Extended Data Fig. 9a, b). NSD3 depletion also inhibited the growth of an independent $8p11^{AMP}$ PDX sample (Extended Data Fig. 9c). Thus, a subset of PDXs from primary LUSC human tumours are NSD3-regulated and selectively sensitive to NSD3 depletion.

NSD3 sensitizes LUSC to bromodomain inhibition

Although at present there is no NSD3 catalytic inhibitor that can be used in a physiological setting, we speculated that selection for NSD3-dependency may confer adaptive vulnerability of LUSC to drugs that target known cancer pathways. To test this idea, we screened a library comprising 285 characterized inhibitors covering about 170 cancer targets to identify drugs that are more effective in PSC_N than in PSC_C cells. The four bromodomain inhibitors (BETi) in the library exhibited the highest differential lethality, consistent with BRD4 interacting directly with NSD3 in LUSC cells[17,18] (Supplementary Table 4, Extended Data Fig. 9d). 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, inhibited the expression of NSD3 target genes and the proliferation of PSC_N cells (Extended Data Fig. 9e, f). NSD3-dependent BETi sensitivity was also observed in PDXs; active growth of PDX^{T1232A} and PDX^{AMP} tumours stalled in response to BETi treatment, a phenotype that was weakened by NSD3 depletion (Fig. 4h, i). By contrast, PDX^{CTR} samples were moderately responsive to BETi therapy independent of NSD3 (Extended Data Fig. 9g, h). Thus, the oncogenic advantage provided by NSD3 to LUSC potentially comes with a cost—hypersensitivity to BETi—that represents 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 such as lung adenocarcinoma, in which targeted therapy approaches have been encouraging, translating similar strategies for LUSC has been difficult. Our results show that NSD3 has a key role 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 the expression of oncogenesis-promoting genes (Extended Data Fig. 10). The genes that are most impacted by NSD3 depletion tend to be 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 leukaemia 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, the oncogenic activity of NSD3 is likely to extend beyond LUSC, as the 8p11–12 amplicon is a common molecular signature of breast cancer and other malignancies[7].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-03170-y.

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. USA* **102**, 9625–9630 (2005).
3. Rooney, C. et al. Characterization of FGFR1 locus in sqNSCLC reveals a broad and heterogeneous amplicon. *PLoS One* **11**, e0149628 (2016).
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 (2010).
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 (2016).
6. 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 (2010).
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 (2017).
8. Travis, W. D. Lung cancer pathology: current concepts. *Clin. Chest Med.* **41**, 67–85 (2020).
9. Husmann, D. & Gozani, O. Histone lysine methyltransferases in biology and disease. *Nat. Struct. Mol. Biol.* **26**, 880–889 (2019).
10. Landau, D. A. et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
11. Qiao, Q. et al. The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J. Biol. Chem.* **286**, 8361–8368 (2011).
12. 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 (2016).
13. Yang, S. et al. Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. *Genes Dev.* **30**, 1611–1616 (2016).
14. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
15. 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 (2016).
16. Bass, A. J. et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat. Genet.* **41**, 1238–1242 (2009).
17. Zhang, Q. et al. Structural mechanism of transcriptional regulator NSD3 recognition by the ET domain of BRD4. *Structure* **24**, 1201–1208 (2016).
18. Shen, C. et al. NSD3-short is an adaptor protein that couples BRD4 to the CHD8 chromatin remodeler. *Mol. Cell* **60**, 847–859 (2015).
19. 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]phenoxy]ethyl]-1,3-dimethyl-piperazin-2-one (AZD5153). *J. Med. Chem.* **59**, 7801–7817 (2016).
20. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **69**, 7–34 (2019).
21. Cochran, A. G., Conery, A. R. & Sims, R. J. III. Bromodomains: a new target class for drug development. *Nat. Rev. Drug Discov.* **18**, 609–628 (2019).
22. 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 (2020).

Methods

Plasmids

The cloning of full-length NSD3 and $NSD3_{SET}$ (amino acids (aa) 1021–1320) was based on NCBI sequence NM_023034.2. Full-length NSD3 was cloned into a pCAG vector for transient transfection or into a pENTR3C entry vector for subsequent cloning into a pLenti-CMV-hygro (w117-1) destination vector for expression in HT1080 cells. Different versions of V5–NSD3 (wil-type, T1232A, Y1174A, T1232A/Y1174A or short), SOX2 and V5–FOXE1[16] and AkaLuc (gift from Dr. Michael Z. Lin)[23] were cloned into a pENTR3C entry vector for subsequent cloning into a pPB-CAG-Dest-pA-pgk-bsd (Addgene) destination vector for reconstitution in H520 cells or into a pPB-CAG-Dest-IRES-puro destination vector for transformation in AALE cells[16]. $NSD3_{SET}$ was cloned into a 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)[24] and was independently shown to abrogate NSD3 catalytic activity (data not shown). Wild-type and T1232A mutant versions of $NSD3_{SET}$ (aa 1055–1285) were cloned into a 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, 1:1,000), H3K36me3 (Thermo Fisher Scientific #MA5-24687, 1:1,000), H3K27me3 (Thermo Fisher Scientific #MA5-33081, 1:1,000), V5 (Thermo Fisher Scientific #R960-25, 1:1,000), H3K36me1 (Abclonal #A2364, 1:5,000); H3K4me2 (CST #9725S, 1:1,000), H3K9me2 (CST #4658S, 1:1,000), H3K27me2 (CST #9728S, 1:1,000), γH2AX (CST #9718, 1:1,000), SOX2 (CST #3579, 1:1,000), c-MYC (CST #5605S, 1:1,000), Phospho-4E-BP1 (CST #2855S, 1:1,000), 4E-BP1 (CST #9644S, 1:1,000), FGFR1 (CST #9740S, 1:1,000), H4K20me2 (Abcam #ab9052, 1:1,000), H3 (Abcam #ab1791, 1:5,000), BRD4 (Abcam #ab128874, 1:1,000), β-actin (CST #4970, 1:5,000), β-tubulin (Millipore #05-661, 1:5,000), NSD2[24] (1:500). NSD3 antibody (1:1,000) was generated using a protein fragment corresponding to aa 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, 1:1,000) 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 h 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

Escherichia coli BL21-RIL cells transformed with the respective expression vectors were cultivated in 2× YT or M9 medium at 20 °C for 16–20 h. Cells were lysed using a French press (Constant Systems TS-Series) and lysates were cleared by centrifugation at 14,000 rpm for 1 h and subsequently injected into a HiTrap SP FF 5 ml column followed by elution with a salt gradient. Fractions containing $NSD3_{SET}$ or $NSD3_{SET\text{-}T1232A}$ 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 dialysed 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 flowthrough 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 dialysed 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 $NSD3_{SET\text{-}T1232A}$ versions were prepared from a pQE30 vector by expression in M9 medium. $^{15}NH_4Cl$ at a concentration of 1 g/l served as the sole source of nitrogen, and for U-[^{13}C,^{15}N] labelling, ^{13}C-glucose (CIL) at 3 g/l was added. Recombinant nucleosomes were assembled using 601 Widom DNA[25]. The 601 Widom DNAs with 20 bp overhang (187 bp) and without overhang (147 bp) were prepared by PCR and purified by phenol/chloroform extraction and isopropanol precipitation. The Cy5-labelled dsDNA (147 bp or 187 bp) was prepared by annealing oligonucleotides 187-F Cy5-CGAGGCTGTTCAATACATGC, 187-R GGACCCTATACGCGGCC, 147-F Cy5-CTGGAGAATCCCGGTGCCGAGGC, and 147-R ACAGGATGTATATAT CTGACACGTGCCTG 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 protocols[26]. Histone octamers and nucleosomes were assembled according to methods described previously[27].

Purification of recombinant full-length NSD3

HEK 293T cells were transfected with pCAG-FLAG–NSD3 (wild-type or mutant NSD3) 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, and the pellet was treated with buffer containing 20 mM HEPES (pH 7.9), 1.5 mM $MgCl_2$, 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, and 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[28] using 2 µg of enzyme (unless otherwise noted) and 2 µg of recombinant nucleosomes (EpiCypher, catalogue#s: 16-0009 and 16-2004) or recombinant histone H3 (generated as described[29]) in the presence of 3H-SAM or non-radiolabelled SAM.

Cell lines, primary cell cultures and cell assays

HEK-293T (ATCC #CRL-1573), KNS-62 (JCRB #IFO50358), SK-MES-1 (ATCC #HTB-58), and HT1080 (ATCC #CCL-121) cells were grown in Dulbecco's modified Eagle's medium (DMEM) 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), and 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 tumour biopsies isolated from the indicated mouse models according to methods previously described[30]. Primary mouse cancer cell lines were cultured in DMEM supplemented with 10% fetal calf serum, 2 mM L-glutamine and penicillin–streptomycin (Life Technologies). Primary cultures of mouse lung fibroblast were prepared from the indicated mice using standard methods[31]. In brief, mice were euthanized by CO_2 asphyxiation and perfused with 5 ml normal saline. Lungs were sterilely removed and cut into small 2-mm slices and allowed to adhere on a tissue culture plate. Lung tissue explants were cultured in DMEM 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 h after 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 screens were performed as previously described[32]. In brief, cells were seeded at 5×10^3 cells/ml in triplicate in 96-well plates. The confluency of treated cells was measured using a live cell kinetic imaging system (IncuCyte Zoom) for 120 h with data collection every 4 h. Data are represented as mean ± s.e.m. (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* 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 h 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 the indicated guide RNAs or pLKO.1/puro (Addgene) construct expressing the indicated shRNAs, psPAX2 and pCMV-VSVg in a ratio of 5:3:2 by mass. The following guide RNA or shNRA oligo sequences were used: sgControl CTTCGAAATGTCCGTTCGGT, sg*NSD2* ACTCGTTA ACAAATTCTCCC (made by Saumya Sankaran), sg*NSD3* GGATACTGAT TATATGAC, sg*FGFR1* TCCCCGACCTTGCCTGAACA, sg*PLPP5* CAATAA AACTGATCGTAGGG, shControl CCTAAGGTTAAGTCGCCCTCG, sh*NSD3* GCAGATTGTTTGGGTCAAATT. After 48 h of transfection, target cells were transduced with 0.45 μm filtered viral supernatant and 8 μg/ml polybrene. Cells were selected 24 h after medium 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 h 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$^{LSL\text{-}Pik3ca(H1047R)}$, *Col1a1*$^{LSL\text{-}Sox2}$, *Cdkn2a/b*$^{LoxP/LoxP}$, and *Fgfr1*$^{LoxP/LoxP}$ mice have been described before[33–36]. The reporter-tagged insertion with conditional potential *Nsd3*$^{tm1a(EUCOMM)}$ mouse strain originates from ES clone #HEPD0563-4-G02 obtained from the European Mouse Mutant Archive repository[37]. 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 a *ROSA26*^{FlpO} deleter strain[38] to generate conditional allele *Nsd3*$^{LoxP/LoxP}$. The *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[39]. 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 used for tumour 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 and 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). Tumour size was measured using a digital caliper and tumour volume was calculated using the formula: volume = $(\text{width})^2 \times \text{length}/2$ where length represents the largest tumour diameter and width represents the perpendicular tumour diameter. The endpoint was defined as the time at which a progressively growing tumour 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 tumours in the lungs of *ROSA26*$^{LSL\text{-}Pik3ca(H1047R)}$, *Col1a1*$^{LSL\text{-}Sox2}$ *Cdkn2a/b*$^{LoxP/LoxP}$ (*PSC*), *ROSA26*$^{LSL\text{-}Pik3ca(H1047R)}$; *Col1a1*$^{LSL\text{-}Sox2}$; *Cdkn2a/b*$^{LoxP/LoxP}$; *Fgfr1*$^{LoxP/LoxP}$ (*PSC*$^{FGFR1\text{-}KO}$), *ROSA26*$^{LSL\text{-}Pik3ca(H1047R)}$; *Col1a1*$^{LSL\text{-}Sox2}$; *Cdkn2a/b*$^{LoxP/LoxP}$; *Nsd3*$^{LoxP/LoxP}$ (*PSC*$^{NSD3\text{-}KO}$), *ROSA26*$^{LSL\text{-}Pik3ca(H1047R)}$; *Col1a1*$^{LSL\text{-}Sox2}$; *Cdkn2a/b*$^{LoxP/LoxP}$; *LSL-Nsd3*^{T1242A} (*PSCN*) mutant mice, we used replication-deficient adenoviruses expressing Cre-recombinase (Ad-Cre) as previously described[40]. Briefly, 8-week old mice were anesthetized by continuous gaseous infusion of 2% isoflurane for at least 10 min using a veterinary anaesthesia 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 analysed for tumour 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. In brief, surgically resected tumour specimens were obtained from deidentified patients with histologically confirmed LUSCs. All tumour 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 tumours were generated and propagated by transplanting small tumour fragments isolated directly from surgical specimens subcutaneously into NSG mice as we established previously[32]. Genome-wide copy number analysis was performed for each PDX using an OncoScan CNV Assay Kit (Thermo Fisher Scientific) as previously described[41]. BioDiscovery's Nexus Copy Number software was used to call CNAs using the default parameters. All data were analysed 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 the following primers: *NSD3*ex20-F GACCGTATAATTGATGCCGGC, *NSD3*ex20-R CTGCAGGAATATCACAGAGAGC. For analysis of *NSD3* knockdown growth, collected PDX tumours were minced using a razor blade and digested in collagenase digestion buffer at 37 °C for 1 h. Cells were passed through 100-μm and 40-μm cell strainers and centrifuged at 1,200 rpm for 5 min. Cells were incubated in RBC lysis buffer for 2 min

and then resuspended in 6 ml medium and spun through 0.5 ml of serum layered on the bottom of the tube to remove cellular debris. Contaminating human or mouse haematopoietic and endothelial cells (CD45, Ter119, CD31) were 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 h. Next, the cells were collected, mixed with Matrigel (1:1) and transplanted into 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 into the hind flanks of NSG mice. When tumours became palpable, they were calipered to monitor growth kinetics.

Tumorigenicity assay in vitro and in vivo

Immortalized human tracheobronchial epithelial (AALE) cells have been described before[16]. 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 the sub-renal capsule into immunocompromised NSG mice using a previously described protocol[42]. 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 were discharged slowly from the catheter. For in vivo transformation assays, AALE cells were electroporated with a PiggyBac transposon plasmid expressing AkaLuc, which catalyses the oxidation reaction of a substrate AkaLumine and produces near-infrared bioluminescence that can penetrate most animal tissues[23,43]. 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 analysed using Aura software (Spectral Instruments Imaging) and quantified in radiance units of photons per second per square centimetre 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 tumour-bearing mice at 90, 120 and 150 days after Ad-Cre induction as previously described[44]. In brief, mice were anaesthetized by continuous gaseous infusion of 2% isoflurane for at least 10 min using a veterinary anaesthesia system. The mice were intubated using a 20 gauge × 1 inch catheter and were transferred onto the bed of an 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 were rotated around the subject in roughly 1.0 degree increments. Each animal's breathing was held at a pressure of 20 cm H_20 during the 20 s acquisition. The resulting raw data were reconstructed to a final image volume of 875 × 875 × 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. Tumours 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 tumour 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 were performed by a well-trained researcher in a blinded manner.

Histology and immunohistochemistry

Tissue specimens were fixed in 4% buffered formalin for 24 h and stored in 70% ethanol until paraffin embedding. Three-micrometre sections were stained with haematoxylin 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[40]. The following antibodies were used (at the indicated dilutions): cleaved caspase 3 (CST #9664, 1:100), H3K36me2 (CST #2901, 1:2,000), Ki67 (BD Bioscience #550609, 1:1,000), KRT5 (CST #71536, 1:1,000), NSD3 (described above), p63 (Abcam, #ab124762, 1:500) and TTF1 (Abcam #ab76013, 1:500). Sections were developed with DAB and counterstained with haematoxylin. Pictures were taken using a PreciPoint M8 microscope equipped with PointView software. Analysis of the tumour area and IHC analysis were done using ImageJ software. Quantification of NSD3 IHC chromogen intensity was performed by measuring the reciprocal intensity of the chromogen stain as previously described[45]. In brief, standard RGB colour 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-labelled dsDNA (147 bp or 187 bp) nucleosomes at concentration 100 nM mixed in 1:1 volume ratio with appropriate $NSD3_{SET}$ or $NSD3_{SET\text{-}T1232A}$ 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.

Isothermal titration calorimetry 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 $NSD3_{SET\text{-}T1232A}$ domains at 100–125 µM, and isothermal titration calorimetry (ITC) measurements were performed in a MicroCal PEAQ-ITC instrument (Malvern Panalytical) at 20 °C with 55 injections of 0.7 µl SAM. Data were analysed 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 1 mg/ml) were incubated with SYPRO Orange (ThermoFisher, cat. No S6651) for 1 min 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 eight times and the 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 30-s incubation steps, respectively, with each step including a rise of 0.5 °C from 15 °C to 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-labelled 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 on HNCO, HNcaCO, HNCA, HNcoCA, HNCACB, CBCAcoNH with further support from the TROSY-version of 3D ^{15}N-edited NOESY (100 ms mixing time) recorded on 370 μM U-[^{15}N] samples[46–48]. Data were processed using Topspin 4.07 program and analysed with the program SPARKY (https://www.cgl.ucsf.edu/home/sparky/) and CARA (http://cara.nmr.ch/doku.php). $NSD3_{SET}$ and $NSD3_{SET}$(T1232A) backbone torsion angles were determined with Talos-N software[49]. 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 labelled U-[^{15}N]$NSD3_{SET}$ or $NSD3_{SET}$(T1232A) preparations containing variants[50]. 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: $dNOE_i = |NOE_i| \times [(1/SNR_{NOE,i})^2 + (1/SNR_{noNOE,i})^2]^{0.5}$, where SNR is signal-to-noise ratio for the experiment with (noNOE) and without saturation (NOE) and $|NOE_i|$ is the intensity ratio of saturated and non-saturated signal. The chemical shift perturbation of ^{1}H/^{15}N resonances was determined with the equation[51]: $CSP_i = [(\Delta\delta_{H,i})^2 + 0.25 \times (\Delta\delta_{N,i})^2]^{0.5}$, 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 program[52] was used 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 PDB: 6CEN X-ray structure[53] with the G^{1267}–N^{1268} residues modelled, which are lacking from the experimental electron density in the SwissModel[54]. 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 overlaps nearly completely with the analogical residues from H3.3K36M peptide (residues 29–42) in the $SETD2_{SET}$–H3.3K36M complex[13,55]. The starting peptide pose for the H3.1 peptide for $NSD3_{SET}$ was taken from the $SETD2_{SET}$–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) that drive the docking process. The remaining parameters were treated according to standard docking protocol[52]. $NSD3_{SET}$ with the regulatory loop in the closed conformation (auto-inhibited state) was modelled using SwissModel[54] and the crystal structure of $NSD2_{SET}$ (PDB id: 5LSU:A; fragment 983–1203)[56] 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), and the RNA-seq libraries were constructed and sequenced in BGI. Alternatively, 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

RNA-seq reads from PSC_N cells with or without NSD3 depletion were aligned to mouse genome mm10 by HISAT2 (https://daehwankim-lab.github.io/hisat2/)[57]. Then the transcriptomes were annotated to GENCODE M23 (https://www.gencodegenes.org/)[58] by FeatureCounts (2.0.0)[59]. Differentially expressed genes (DEGs) were detected using the DESeq2 package from Bioconductor (http://bioconductor.org/packages/release/bioc/html/DESeq2.html)[60] using the likelihood ratio test (LRT, adjusted *P* value <0.001) or Wald test. The expression levels of genes were normalized and $\log_2$-transformed, and the variances-stabling transformation was implemented. Bioconductor package EnhancedVolcano (http://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html, 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 (http://www.gsea-msigdb.org/gsea/index.jsp, 4.0.0)[61] and the Bioconductor package GeneAnswers (http://bioconductor.org/packages/release/bioc/html/GeneAnswers.html). 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 dataset analysis

TCGA data from patients with lung squamous cell carcinoma (project ID: TCGA-LUSC, dbGaP study accession: phs000178)[62] were analysed using cBioPortal[63,64] 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 to identify significantly altered regions of amplification or deletion across sets of patients. GISTIC-derived copy number analysis indicates copy-number level per gene, in which 1 and 2 indicates amplification. Basal-normalized transcript expression data (*z* scores) used for this analysis were RNA Seq V2 RSEM.

CUT&RUN

This assay was performed as described previously with modifications[65] and using the H3K36me2 and H3K27me3 antibodies noted above. Buffer exchange was performed by centrifugation at 4 °C, 600*g* for 5 min. After the addition of concanavalin A beads, the buffer exchange was performed with a magnetic stand, allowing at least 5 min 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 min 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. When nuclear aggregates were present, the suspension was filtered through a 40-μm strainer before proceeding to the 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 min, rotating at 4 °C. Successful nuclei–bead binding was confirmed with Trypan Blue staining. Bead-coated nuclei were resuspended in WB at a final concentration of 5×10^6 cells/ml and 100 μl was 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 with rotation overnight at 4 °C. On the next day, nuclei were washed twice with 1 ml ice-cold WB and resuspended in 25 μl ice-cold WB. While gently vortexing, 25 μl of 2× pAG-MNase mix (EpiCypher, Inc. catalogue no.

15-1016) was added to each tube and samples were incubated for 1 h at 4 °C. Nuclei were washed twice with WB and resuspended in 150 µl ice-cold WB. While gently vortexing, 3 µl pre-chilled 100 µM $CaCl_2$ was added and after briefly flicking to mix, tubes were incubated for 30 min in a 0 °C water bath. After 30 min, the MNase was inactivated by adding 150 µl of 2× STOP buffer. To isolate the digested DNA though diffusion, samples were incubated for 20 min at 37 °C with no shaking. Samples were centrifuged at 16,000*g* for 5 min 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 h to overnight without shaking. DNA was purified with a 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 (https://bitbucket.org/qzhudfci/cutruntools/src/master/, version: 49ddd24)[66]. The reads were trimmed by Trimmomatic[67], then aligned to the mouse genome mm10 by Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml)[68]. Duplicate reads at the same locations were removed by MarkDuplicates of Picard package (http://broadinstitute.github.io/picard/, 2.23.8). The enrichment densities of different samples were normalized to read counts per million mapped reads and to have the same scale of the gene body by ngs.plot (https://github.com/shenlab-sinai/ngsplot, 2.63)[69]. 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 (ΔPeaks), the differences between them calculated, and compared to the distribution of ΔPeaks of DEGs and the genome-wide genes with two-sided robust *t*-test at trim level 0.1. The intergenic region annotation was generated by Complement of BedTools[70], and then the average profile of the intergenic regions was visualized by ngs.plot.

ChIP–qPCR

ChIP assays were performed as described[24]. Purified DNA was subject to quantitative real-time PCR analysis. Purified ChIP DNA was analysed on a 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. No statistical methods were used to predetermine sample size.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The CUT&RUN and RNA-seq data from tumour-derived cells have been deposited in the Gene Expression Omnibus (GEO) database under accession number GSE149482, and RNA-seq data from tumour biopsies have been deposited in the GEO database under accession number GSE149272. Full blot data for all the western blots were shown in Supplementary Fig. 1 and all source data are available in Supplementary Data. Source data are provided with this paper.

23. Su, Y. et al. Novel NanoLuc substrates enable bright two-population bioluminescence imaging in animals. *Nat. Methods* **17**, 852–860 (2020).
24. Kuo, A. J. et al. NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell* **44**, 609–620 (2011).
25. 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 (1998).
26. Cheema, M. S. & Ausió, J. Analytical ultracentrifuge analysis of nucleosomes assembled from recombinant, acid-extracted, HPLC-purified histones. *Methods Mol. Biol.* **1528**, 75–95 (2017).
27. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol. Biol.* **119**, 1–16 (1999).
28. Shi, X. et al. Modulation of p53 function by SET8-mediated methylation at lysine 382. *Mol. Cell* **27**, 636–646 (2007).
29. Chen, S. et al. The PZP domain of AF10 senses unmodified H3K27 to regulate DOT1L-mediated methylation of H3K79. *Mol. Cell* **60**, 319–327 (2015).
30. 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 (2015).
31. Edelman, B. L. & Redente, E. F. Isolation and characterization of mouse fibroblasts. *Methods Mol. Biol.* **1809**, 59–67 (2018).
32. Liu, S. et al. METTL13 methylation of eEF1A increases translational output to promote tumorigenesis. *Cell* **176**, 491–504.e21 (2019).
33. Adams, J. R. et al. Cooperation between Pik3ca and p53 mutations in mouse mammary tumor formation. *Cancer Res.* **71**, 2706–2717 (2011).
34. 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 (2016).
35. Krimpenfort, P. et al. p15Ink4b is a critical tumour suppressor in the absence of p16Ink4a. *Nature* **448**, 943–946 (2007).
36. Hoch, R. V. & Soriano, P. Context-specific requirements for Fgfr1 signaling through Frs2 and Frs3 during mouse development. *Development* **133**, 663–673 (2006).
37. Skarnes, W. C. et al. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).
38. Raymond, C. S. & Soriano, P. High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS One* **2**, e162 (2007).
39. Chu, V. T. et al. Efficient generation of Rosa26 knock-in mice using CRISPR/Cas9 in C57BL/6 zygotes. *BMC Biotechnol.* **16**, 4 (2016).
40. Mazur, P. K. et al. SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. *Nature* **510**, 283–287 (2014).
41. Fraser, M. et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359–364 (2017).
42. Shultz, L. D. et al. Subcapsular transplantation of tissue in the kidney. *Cold Spring Harb. Protoc.* **2014**, 737–740 (2014).
43. Iwano, S. et al. Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science* **359**, 935–939 (2018).
44. Fushiki, H. et al. Quantification of mouse pulmonary cancer models by microcomputed tomography imaging. *Cancer Sci.* **100**, 1544–1549 (2009).
45. Wang, Z. et al. SETD5-coordinated chromatin reprogramming regulates adaptive resistance to targeted pancreatic cancer therapy. *Cancer Cell* **37**, 834–849.e13 (2020).
46. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl Acad. Sci. USA* **95**, 13585–13590 (1998).
47. 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).
48. 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 (2017).
49. 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 (2015).
50. Lakomek, N. A., Ying, J. & Bax, A. Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR* **53**, 209–221 (2012).
51. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
52. 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 (2016).
53. Morrison, M. J. et al. Identification of a peptide inhibitor for the histone methyltransferase WHSC1. *PLoS One* **13**, e0197082 (2018).
54. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46** (W1), W296–W303 (2018).
55. Zhang, Y. et al. Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. *Sci. Rep.* **7**, 43906 (2017).

56. Tisi, D. et al. Structure of the epigenetic oncogene MMSET and inhibition by *N*-alkyl sinefungin derivatives. *ACS Chem. Biol*. **11**, 3093–3105 (2016).
57. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
58. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. **47** (D1), D766–D773 (2019).
59. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**, 550 (2014).
61. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
62. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
63. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal*. **6**, pl1 (2013).
64. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. **2**, 401–404 (2012).
65. Hainer, S. J. & Fazzio, T. G. High-resolution chromatin profiling using CUT&RUN. *Curr. Protoc. Mol. Biol*. **126**, e85 (2019).
66. 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 (2019).
67. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
68. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
69. 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 (2014).
70. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

Acknowledgements We thank members of the Gozani and Mazur laboratories for critical reading of the manuscript, and M. Lin for the AkaLuc vector. This work was supported in part by grants from the NIH to O.G. (R01 GM079641 and R35 GM139569), 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., Ł.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), and 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 and FDCT0038/2020/AFJ, N.M.F. by the American Cancer Society – Joe & Jessie Crump Foundation Fellowship (PF-20-108-01-DMC), X.L. by a CPRIT Research Training Grant (RP170067) and S.H. by a Deutsche Forschungsgemeinschaft Fellowship (HA8434/1-1).

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, 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.F.C. supervised. S.M.L. performed bioinformatic analysis of tumour RNA-seq. V.K., I.C., D.A., W.F., M.J. and Ł.J. performed biophysical and NMR experiments. I.I.W. evaluated histopathology. I.I.W., B.F. J.A.R. and J.D.M. provided PDX samples. N.-Y.S. analysed RNA-seq and CUT&RUN data. O.G., P.K.M., Ł.J. and N.Y.S. were equally responsible for supervision of research, data interpretation and manuscript preparation.

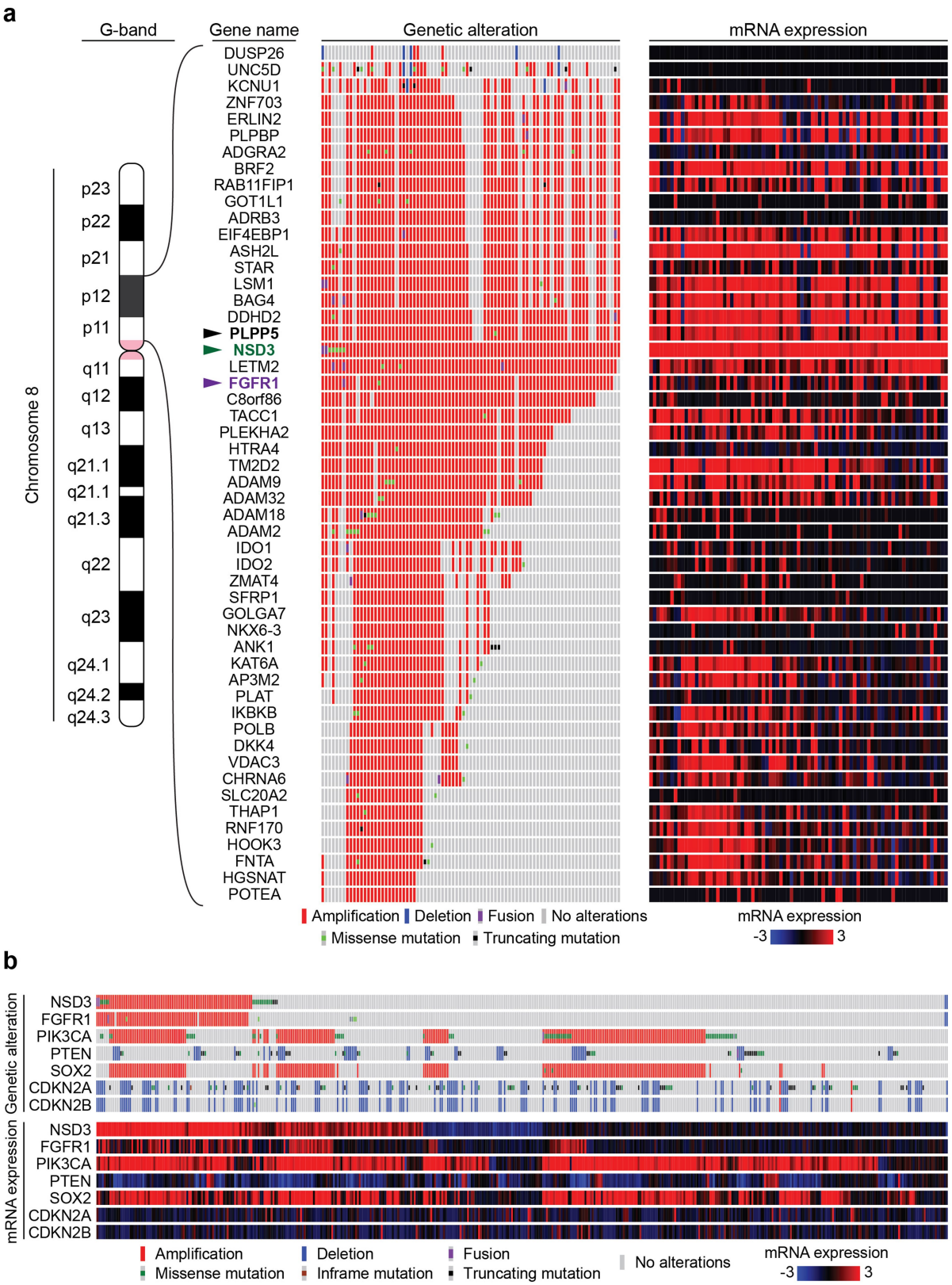
Competing interests O.G. is a co-founder of EpiCypher, Inc. and Athelas Therapeutics, Inc. J.A.R. is a consultant and stock owner (including pending patents) in Genprex, Inc. P.K.M. is a co-founder of Amplified Medicines, Inc. The other authors declare no competing interests.

Additional information

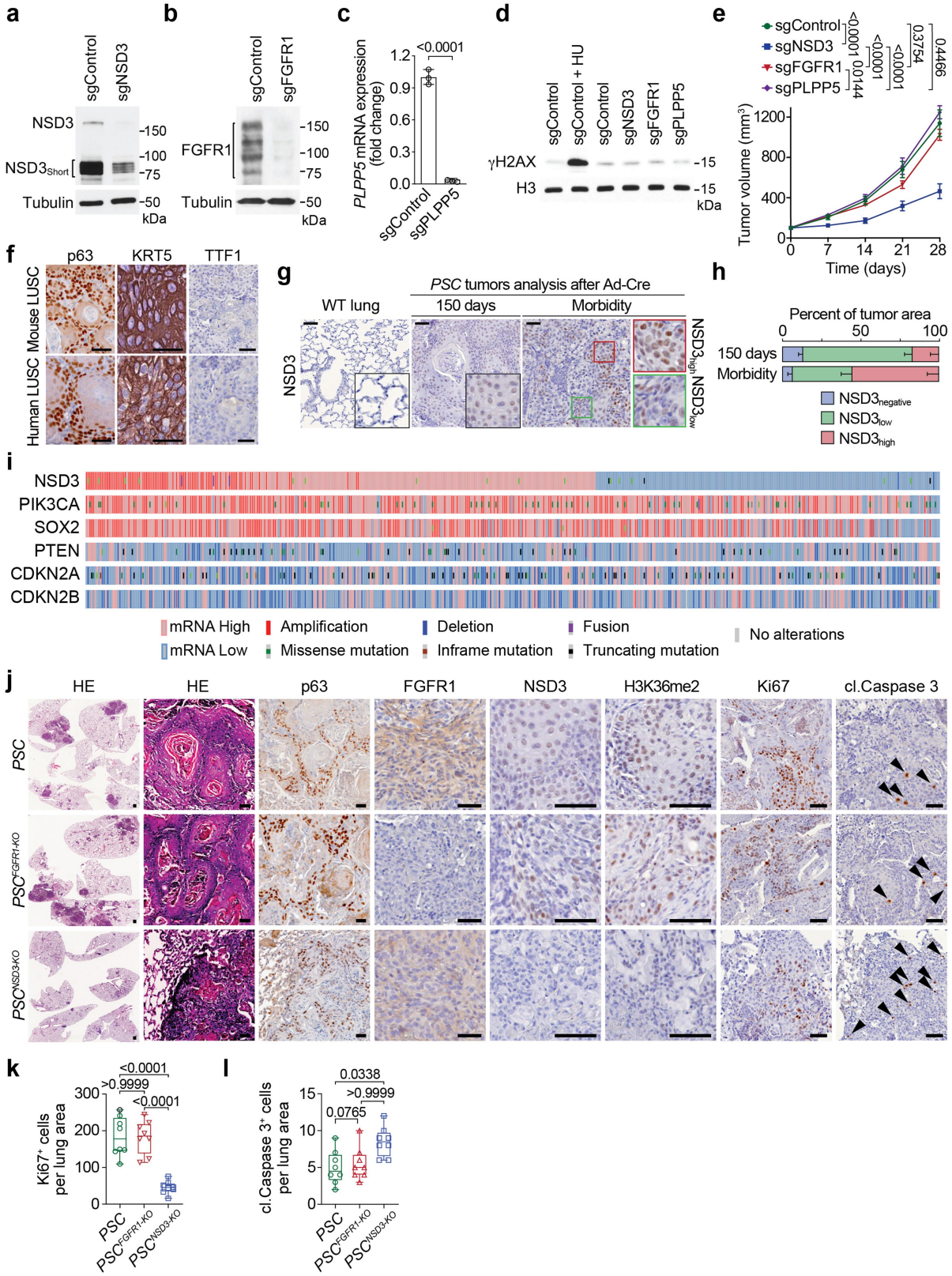
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-020-03170-y.

Correspondence and requests for materials should be addressed to N.-Y.S., Ł.J., P.K.M. or O.G.

Peer review information *Nature* thanks Gang Greg Wang, Jon Wilson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

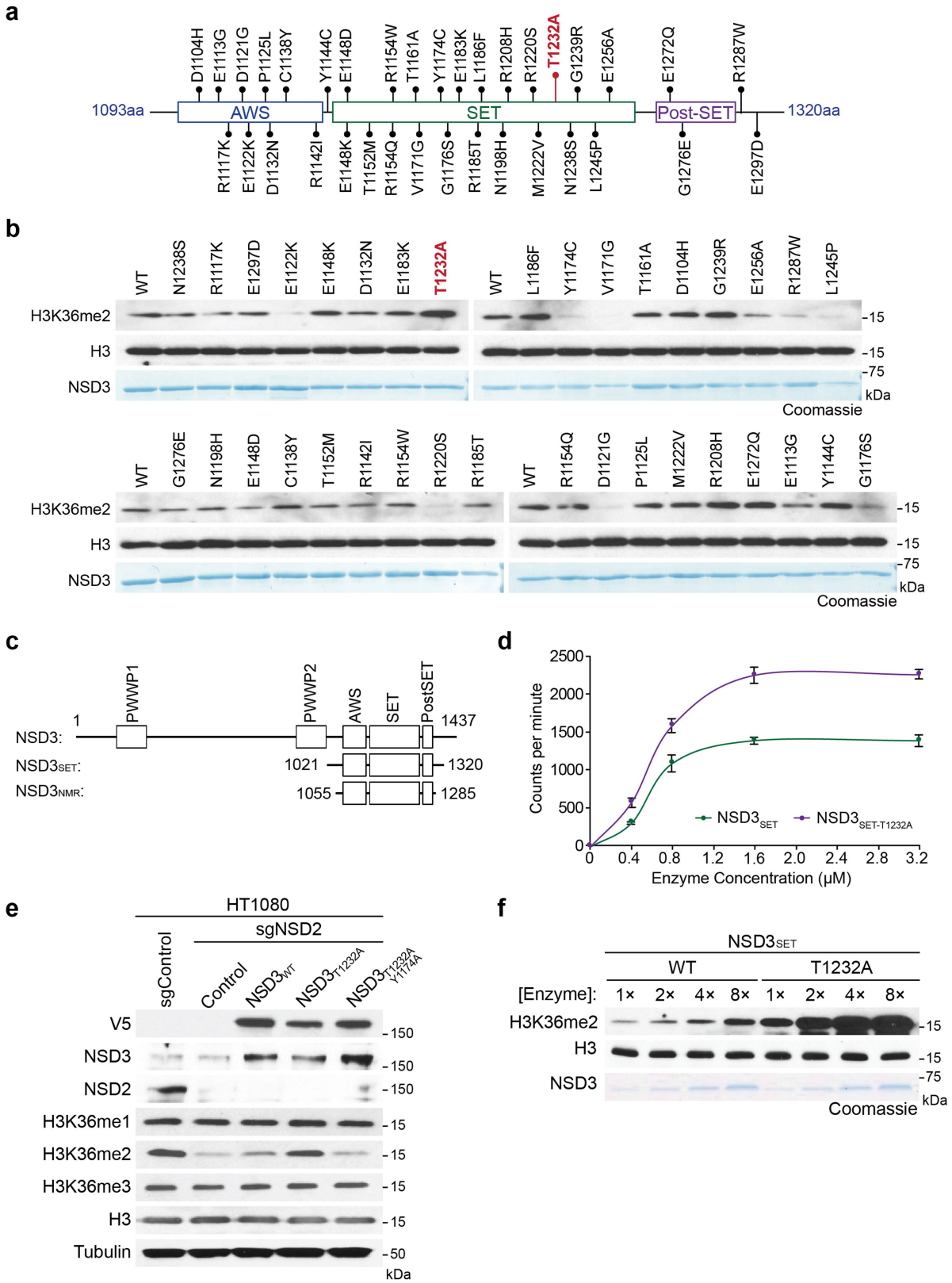


Extended Data Fig. 1 | Characterization of genetic drivers of LUSC tumorigenesis. a, Genetic alterations and mRNA expression of genes in the 8p11–12 amplification genomic region of chromosome 8 from the 85 specimens of 8p11–12-amplified human LUSC in the TCGA dataset. **b**, Genetic alterations and mRNA expression of *NSD3*, *FGFR1*, *PIK3CA*, *PTEN*, *SOX2*, *CDKN2A* and *CDKN2B* from all 464 of the human LUSC specimens in the TCGA dataset (ID: phs000178) analysed using cBioPortal.



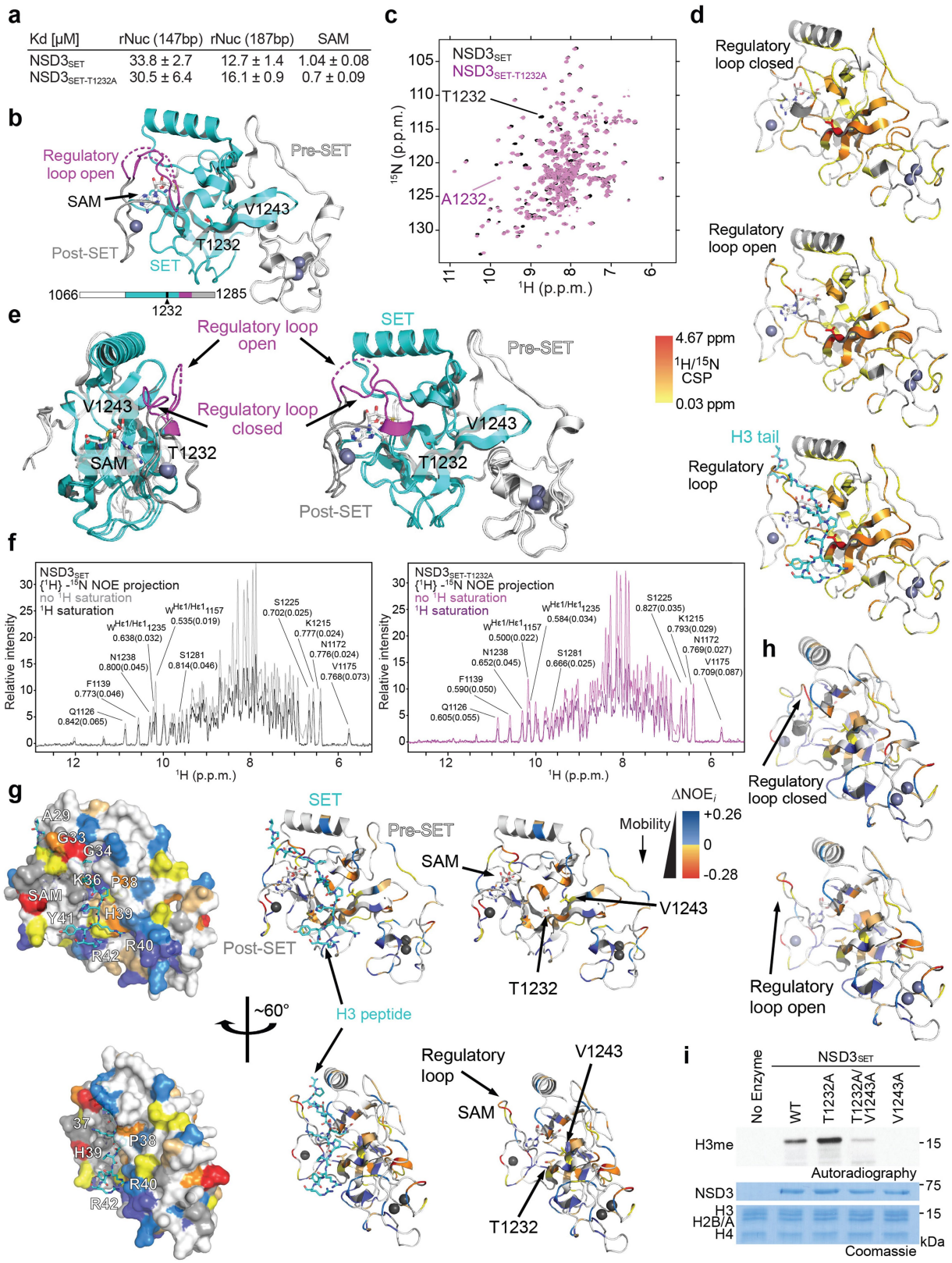
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Depletion of NSD3 but not FGFR1 inhibits LUSC tumorigenesis in vivo. a, Western blot analysis with the indicated antibodies of whole-cell lysates from H520 cells expressing sgControl or sg*NSD3*. 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 sg*FGFR1*. Tubulin was used as a loading control. **c**, Real-time (RT) qPCR analysis of *PLPP5* mRNA expression in H520 cells expressing sgControl or sg*PLPP5*. 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 are presented as fold change relative to the sgControl sample; 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 $8p11^{AMP}$ H520 LUSC cells. Tumour volume quantification of the indicated H520 xenografts generated in immunocompromised mice ($n = 5$ mice, for each treatment group). *P* values determined by two-way ANOVA with Tukey's post hoc test. Data are represented as mean ± s.e.m. Data for sgControl, sg*NSD3* and sg*FGFR1* are the same as from Fig. 1b. **f**, Representative HE-stained sections and IHC staining of lung tissue from PSC mouse model lung tumour 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 tumour progression. IHC analysis of NSD3 levels in wild-type lung and tumour biopsies collected from the PSC model at 150 days and at the clinical endpoint when animals developed substantial morbidity (representative of $n = 8$ samples for each group). Scale bars, 50 μm. **h**, Quantification of NSD3 levels in tumour biopsies as in **g** within tumour 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 tumours. Genetic alterations and mRNA expression of the indicated genes in human LUSC samples from the TCGA dataset showing frequency of overexpression and alternations in *NSD3* and known driver mutations *PIK3CA*, *PTEN*, *SOX2*, *CDKN2A* and *CDKN2B*. **j**, Representative HE-stained sections and IHC staining of lung tissue from PSC (control), $PSC^{FGFR1\text{-}KO}$ and $PSC^{NSD3\text{-}KO}$ 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^+ 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 determined by one-way ANOVA with Tukey's post hoc test (**k**, **l**).



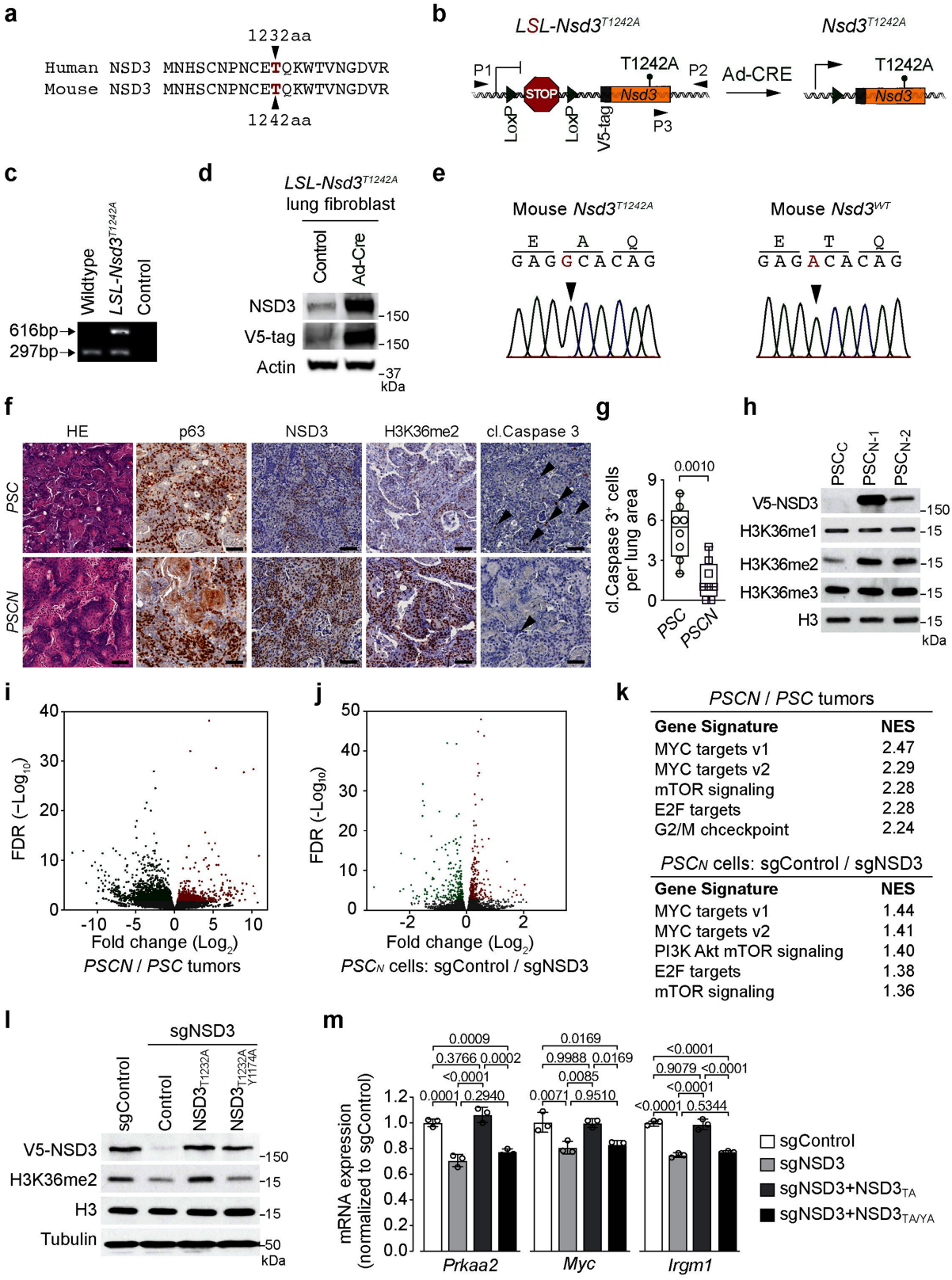
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 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 Fig. 2b on rNuc with recombinant wild-type $NSD3_{SET}$ or the indicated mutants (1–35) with non-radiolabelled SAM (the T1232A mutant is indicated in red). For each screening panel set, the top row shows the western blot and the bottom row shows 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. Data shown as 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 Fig. 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, Coomassie blue stain of NSD3 proteins.



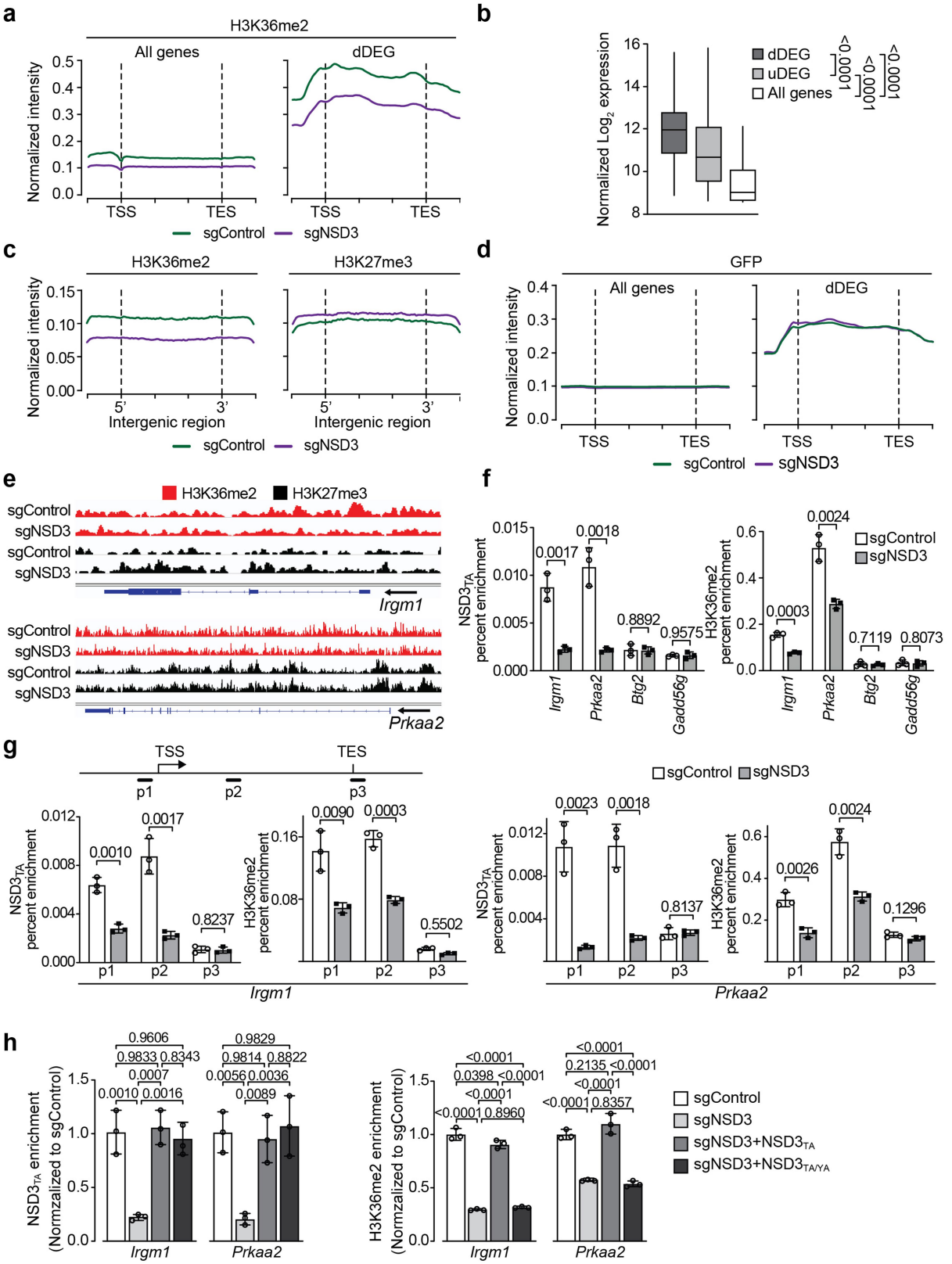
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Biophysical characterization of T1232A substitution on the NSD3$_{SET}$ domain. a, K_d values (± s.d.) for binding studies of NSD3$_{SET}$ and NSD3$_{SET}$(T1232A) to recombinant nucleosomes (rNuc) reconstituted on 147 bp and 187 bp 601 Widom DNA as indicated (determined by microscale thermophoresis (MST)) and the cofactor SAM (determined by ITC) (Supplementary Table 1; see Methods). Mean ± s.d. from three independent experiments. **b**, The topology of NSD3$_{SET}$ domain segments of pre-SET (white-grey), SET (cyan), post-SET (dark grey) and regulatory loop (magenta) marked onto the ribbon representations of existing NSD3$_{SET}$ 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 grey spheres and the SAM cofactor with sticks. Residues T1232 and V1243 are marked with their side-chains. Bottom, schematic representation of the primary sequence with indicated domains and the location of the T1232A substitution. **c**, Overlay of the 2D [^{1}H,^{15}N] TROSY–HSQC (transverse relaxation optimized spectroscopy–heteronuclear single quantum coherence) spectra of 250 µM wild-type U-[^{15}N]NSD3$_{SET}$ (dark grey) and U-[^{15}N] NSD3$_{SET}$(T1232A) (magenta) at pH 7.5 and 25 °C. The ^{1}H/^{15}N backbone amide cross-peaks of T1232 in wild-type NSD3$_{SET}$ and A1232 of NSD3$_{SET}$(T1232A) are marked. **d**, The backbone amide chemical shift perturbations ($CSP_i = [(\Delta\delta_{H,i})^2 + 0.25 \times (\Delta\delta_{N,i})^2]^{0.5}$) between NSD3$_{SET}$ and NSD3$_{SET}$(T1232A) are mapped onto the ribbon representations of protein structures; left, NSD3$_{SET}$ with regulatory loop in closed conformation (PDB 5UPD); centre, in open conformation (based on the NSD2$_{SET}$ crystal structure (PDB 5LSU) modelled with SwissModel); and right, the open conformation with the H3.1 A^{28}–R^{41} substrate docked (see Methods). The T1232 and V1243 residues are marked with their side-chains. Residues are coloured as follows: $0.03 \le CSP_i < 0.05$ (yellow), $0.05 \le CSP_i < 0.11$ (light orange), $0.11 \le CSP_i < 0.75$ ppm (orange), $CSP_i > 4.60$ (red), prolines and residues missing amide assignments are grey. **e**, 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. Protein sequence graph and colouring as in Fig. 4h. **f**, Projections of 2D TROSY-type heteronuclear {^{1}H}-^{15}N nuclear Overhauser effect (NOE) experiments collected for 370 µM wild-type U-[^{15}N]NSD3$_{SET}$ (left) and NSD3$_{SET}$(T1232A) (right) at pH 7.5 and 25 °C with several residues marked (see Methods). The higher the NOE ratio value, the more rigid the N–H vector, meaning less dynamic motion, whereas a 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, $H^{\varepsilon 1}/N^{\varepsilon 1}$ from W1235 and W1157. **g**, The observed changes in {^{1}H}-^{15}N NOE values (ΔNOEi) reporting on ps-ns polypeptide main-chain mobility differences between NSD3$_{SET}$ and NSD3$_{SET}$(T1232A) are plotted on the 3D static structures of NSD3$_{SET}$ (PDB: 6CEN, left column represented as protein surface, middle column ribbon with H3 peptide, right column ribbon with no substrate). Key features are indicated. Light blue to dark blue indicates the decrease in fast dynamics after T1232A substitution, and yellow to red indicates 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 NOE mapped on the ribbon representations of the open and closed states of the regulatory loop within the NSD3$_{SET}$ and in Fig. 2e. The structures are the same as in **e**. **i**, Methylation assays as in Fig. 2a with rNuc (187bp) and the indicated NSD3 variants.



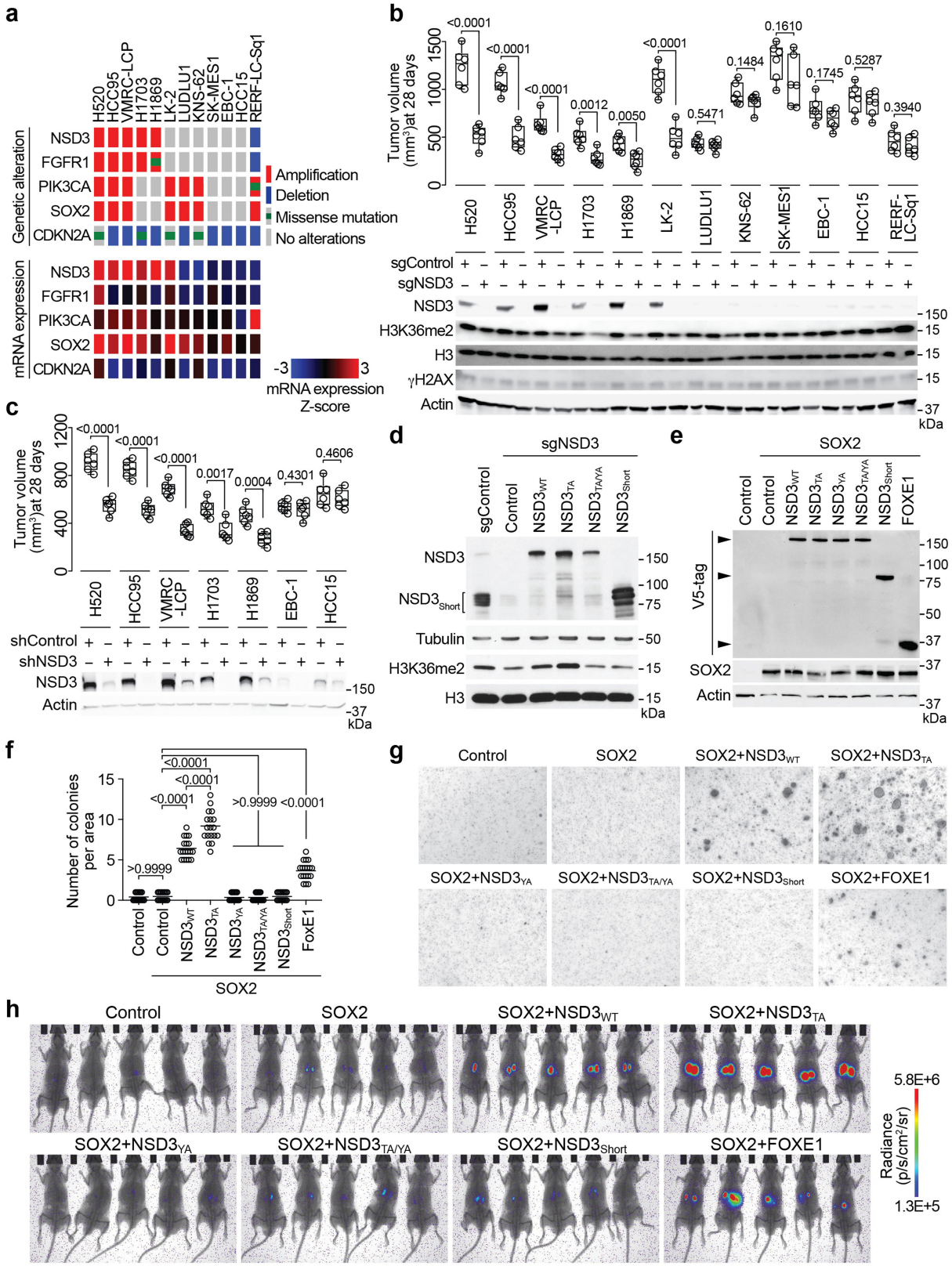
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 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 mouse 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 locations 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 product sizes are marked. **d**, Western blots with the indicated antibodies (V5 antibody detects V5-tagged NSD3(T1242A)) of lysates from *LSL-V5-Nsd3*^{T1242A} lung fibroblasts transduced ex vivo with Ad-Cre or vehicle (control). Actin is shown as a loading control. **e**, Left, Sanger sequencing confirmation that the *Nsd3*^{T1242A} mutation is present in generated conditional *LSL-V5-Nsd3*^{T1242A} mutant mouse. Right, wild-type NSD3 sequence from control animals. **f**, Representative lung tissue sections from *PSC* and *PSCN* mutant mice stained with HE or IHC with indicated antibodies (representative of $n = 8$ mice for each group). Scale bars, 50 µm; arrowheads, cleaved caspase 3-positive cells. **g**, Quantification of cleaved caspase 3, a marker of apoptosis in samples as in **f**. *P* values 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 lines. Total H3 was used as a loading control. **i**, Volcano plot of RNA-seq comparison between PSCN and PSC tumour biopsies (three independent biological replicates for each condition) shows higher expression of 891 genes shown in red (fold change $\log_2 \leq -0.5$ and $P < 0.05$ by Wald test) and decreased expression of 1,839 genes shown in green (fold change $\log_2 \geq 0.5$ 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 with or without sg*NSD3* (three biological replicates for each condition) shows that sg*NSD3* caused decreased expression of 234 genes shown in red (adjusted $P < 0.001$ by likelihood ratio test) and increased 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 (see 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 with or without sg*NSD3* 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 determined by one-way ANOVA with Tukey's post hoc test.



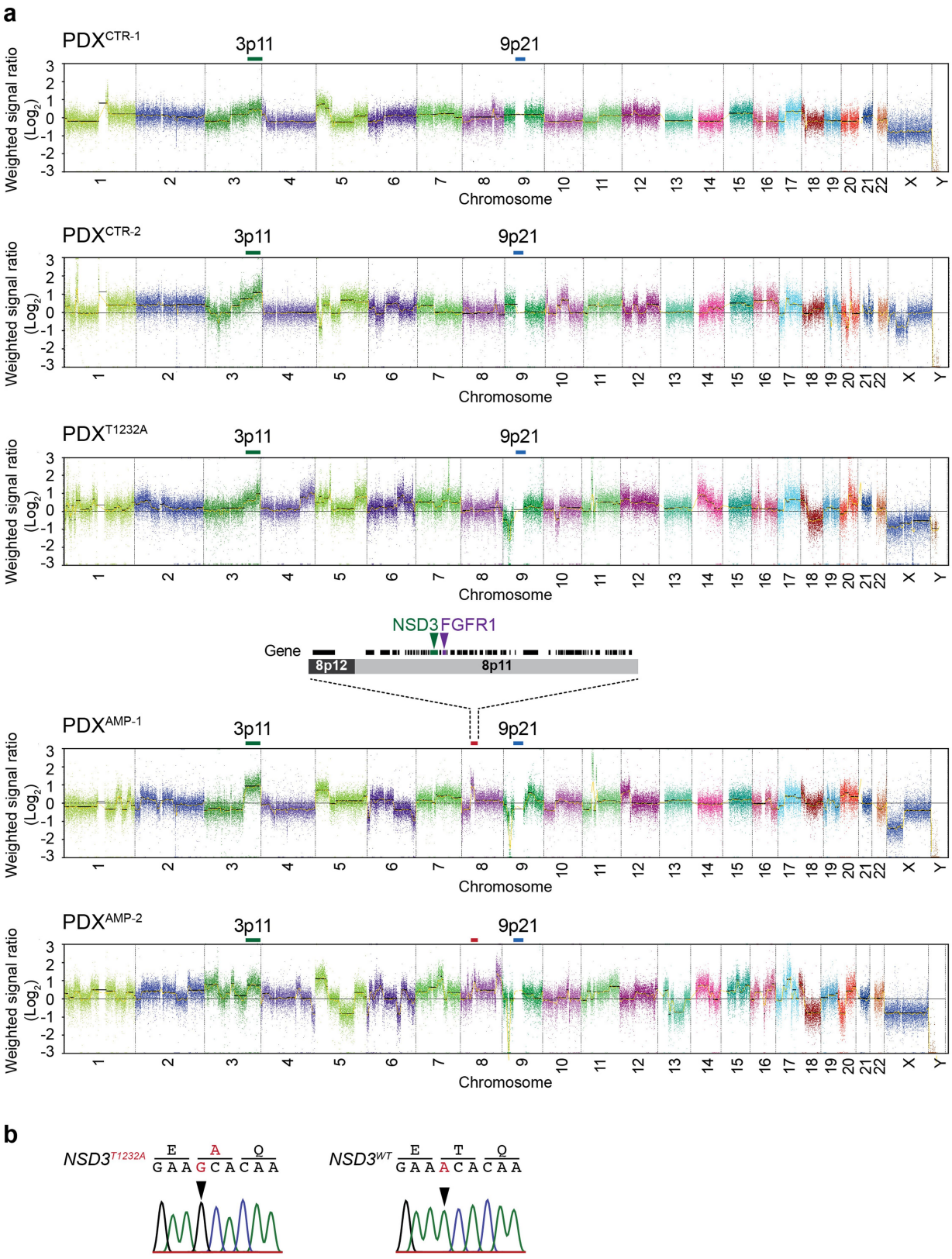
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | NSD3(T1242A)-mediated H3K36me2 in LUSC oncogenic reprogramming. a, Independent CUT&RUN replicate for H3K36me2 as in Fig. 3i. **b**, Normalized gene expression levels from the indicated groups from PSC_N cell line RNA-seq datasets. dDEGs, downregulated DEGs; uDEGs, upregulated DEGs, from Extended Data Fig. 5j. *P* values determined by two-tailed robust *t*-test (see Methods). dDEGs $n = 234$, uDEGs $n = 229$, all genes $n = 16{,}091$. Boxes, 25th to 75th percentiles; whiskers, minimum to maximum; centre line, median. **c**, CUT&RUN profile for H3K36me2 and H3K27me3 in PSC_N cells with or without sg*NSD3* over average of intergenic regions on a genome-wide scale. **d**, CUT&RUN profile for a monoclonal IgG (against GFP) as in Fig. 3i for all genes and dDEGs as indicated in PSCN cells with or without sg*NSD3* 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 are shown as control regions that do not change in response to NSD3 depletion. ChIP–qPCR analysis of V5–NSD3(T1242A) (left) and H3K36me2 (right) in the bodies of the indicated genes. The data were plotted as per cent enrichment relative to input. Mean ± s.d. from three independent experiments. *P* values 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, schematic of general gene structure and site of primers for *Irgm1* (left) and *Prkaa2* (right) gene loci. ChIP–qPCR analysis of V5–NSD3(T1242A) (left) and H3K36me2 (right) 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 per cent enrichment relative to input. Mean ± s.d. from three independent experiments. *P* values determined by two-tailed unpaired *t*-test. **h**, ChIP–qPCR analysis of NSD3(T1242A) (left) and H3K36me2 (right) 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. Mean ± s.d. from three independent experiments. *P* values determined by one-way ANOVA with Tukey's post hoc test.

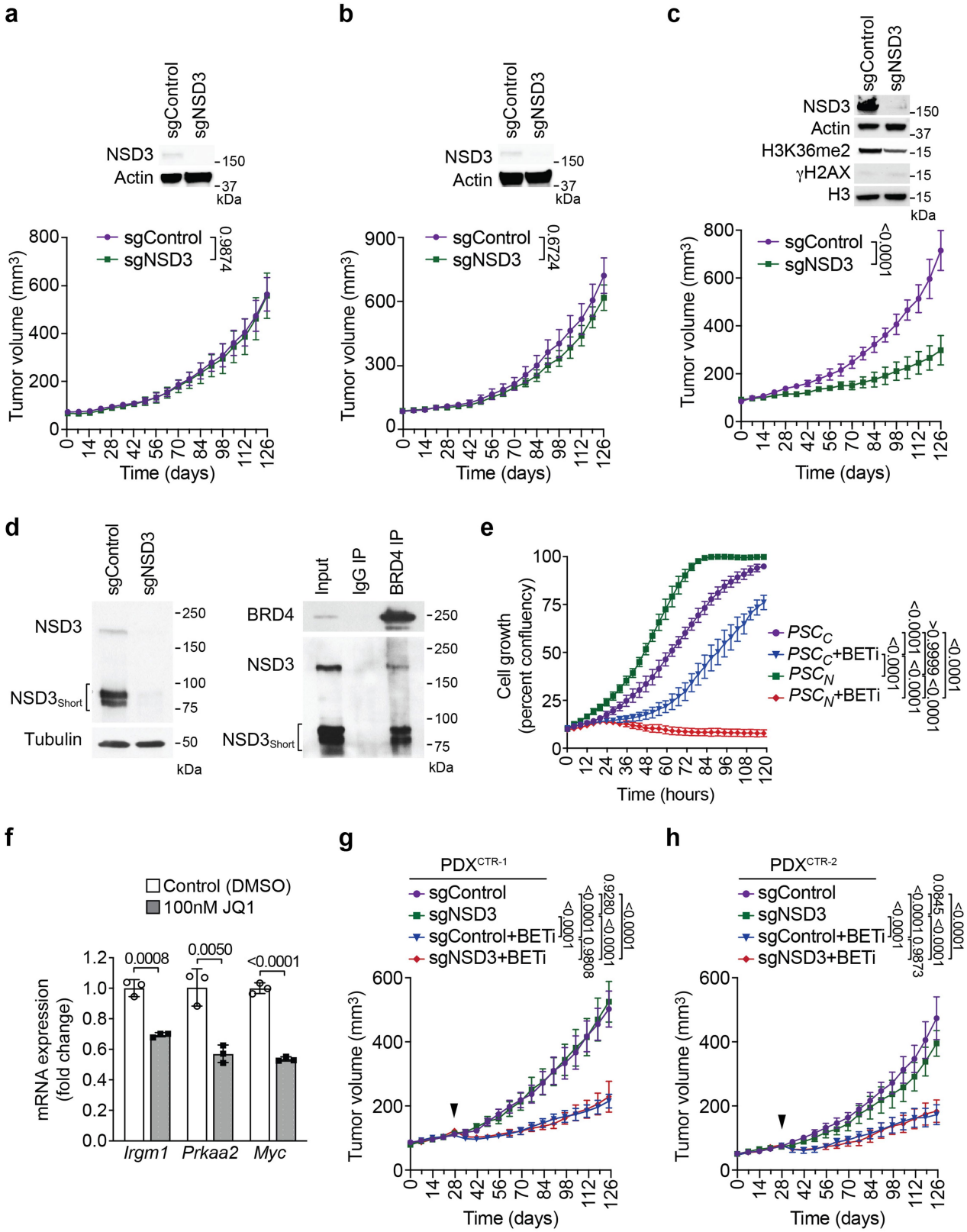


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 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 tumour growth of $8p11^{AMP}$ and NSD3-overexpressing LUSC cell lines. Top, tumour volume quantification of human LUSC cell line xenografts described in Fig. 4a and Extended Data Fig. 7a treated with sgControl or sg*NSD3* and grown in immunocompromised mice at 28 days after implantation ($n = 5$ mice for each treatment group). *P* values determined by one-way ANOVA with Tukey's post hoc test. Data are represented as mean ± s.e.m. Bottom, western blot analysis with the indicated antibodies of whole-cell lysates from cells in top panel. γH2AX levels are shown to assess whether sg*NSD3* expression induces non-specific DNA damage. H3 and actin are used as loading controls. **c**, NSD3 depletion by shRNA attenuates xenograft tumour growth of $8p11^{AMP}$-positive LUSC cell lines. Top, tumour volume quantification of the indicated $8p11^{AMP}$-positive and two control $8p11^{AMP}$-negative human LUSC cell line xenografts as in **a** treated with shControl or sh*NSD3* as indicated and grown in immunocompromised mice at 28 days after implantation ($n = 5$ mice for each treatment group). *P* values determined by one-way ANOVA with Tukey's post hoc test. Data are represented as mean ± s.e.m. Bottom, western blot analysis with the indicated antibodies of whole-cell lysates from cells in top 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) ($NSD3_{TA}$) and NSD3(T1232A/Y1174A) ($NSD3_{TA/YA}$). Western blots of H520 lysates with indicated antibodies H3 and tubulin were used as loading controls (Fig. 4b). **e**, Western blots of whole-cell lysate of AALE cells used in transformation assays for Fig. 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 Fig. 4c. Optical overlay of bioluminescent signal with X-ray images 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 colour bar indicates the total bioluminescence radiance (photons/s/cm^2/sr).

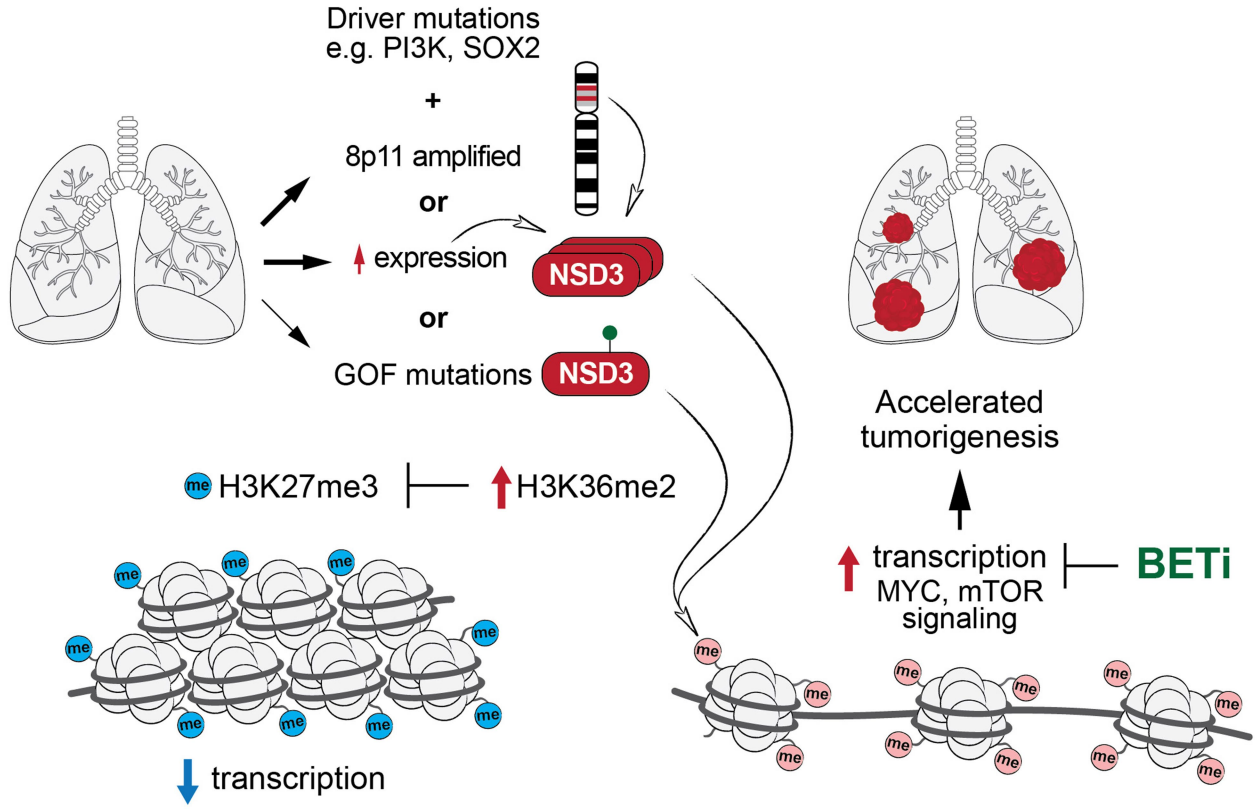


Extended Data Fig. 8 | Genomic characterization of PDXs from primary LUSC. a, Distribution of whole-genome copy number alterations in LUSC PDXs used in this study by OncoScan (see Methods). Green lines, 3p11 amplification (including *PIK3CA* and *SOX2*); red lines, marks, 8p11–12 (for example, *NSD3* 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 PDXs identified *NSD3*^{T1232A} mutation in one of the PDX samples (PDX^{T1232A}) and wild-type *NSD3* in other analysed PDXs.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | NSD3 dependency renders PDXs from patients with LUSC therapeutically vulnerable to BETi. a–**c**, Bottom, tumour volumes of $PDX^{CTR\text{-}1}$ (**a**), $PDX^{CTR\text{-}2}$ (**b**) and $PDX^{AMP\text{-}2}$ (**c**) growing in immunocompromised mice ($n = 5$ mice, for each treatment group). Top, western blots with the indicated antibodies of lysates from PDXs with or without sg*NSD3*. Actin is shown as a loading control. **d**, BRD4 interacts with NSD3 in PSC_N cells. Left, western blot of whole-cell lysates from PSC_N cells with or without sg*NSD3* to show the relative position of NSD3 and $NSD3_{short}$ in PSC_N cells. Tubulin used as a loading control. Right, 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 cells treated with vehicle control or 20 nM 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. RT–qPCR analysis of the indicated mRNAs from PSC_N cells treated with DMSO or 100 nM JQ1. The RT–qPCR data for each target gene were normalized to *Actb* and presented as fold change relative to DMSO-treated sample. Mean ± s.d. from three independent experiments. *P* values determined by two-tailed unpaired *t*-test. **g**, **h**, Tumour volumes of $PDX^{CTR\text{-}1}$ (**g**) and $PDX^{CTR\text{-}2}$ (**h**) treated with BETi or vehicle (control). Arrowhead indicates start of the treatment. *P* values 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**–**h**).



Extended Data Fig. 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 (about 20% of patients with LUSC). While increased expression of *FGFR1* is postulated to be the causative mutation of the 8p11–12 amplicon, our work implicates amplification of the neighbouring gene *NSD3* as the main driving alteration. As shown, *NSD3* amplification leads to increased NSD3 expression and hence increased synthesis of H3K36me2, which works 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 patients with LUSC; thus, NSD3 is frequently highly expressed in LUSCs that do not harbour 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 present in human LUSC. The NSD3(T1232A) variant, because of its enhanced catalytic behaviour, 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 signalling and MYC-associated pathways. We speculate that NSD3-regulated tumours become addicted to H3K36me2-driven transcriptional activation, rendering these tumours particularly vulnerable to BETi as NSD3 and BRD4 interact and cooperate in transcription. This hypersensitivity could potentially be exploited clinically owing 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 owing to amplification.

Author Queries

Journal: **Nature**
Paper: **s41586-020-03170-y**
Title: **Elevated NSD3 histone methylation activity drives squamous cell lung cancer**

AUTHOR:
The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query Reference	Reference
Q1	Please check that the display items are as follows (ms no: 2020-04-06988D): Figs 0 (black & white); 1–4 (colour); Tables: None; Boxes: None; Extended Data display items: Figs 1-10; SI: yes. The eproof contains the main-text figures edited by us and (if present) the Extended Data items (unedited except for legends) and the Supplementary Information (unedited). Please note that the eproof should be amended in only one browser window at any one time, otherwise changes will be overwritten. Please check the edits to all main-text figures (and tables, if any) very carefully, and ensure that any error bars in the figures are defined in the figure legends. Extended Data items may be revised only if there are errors in the original submissions. If you need to revise any Extended Data items please upload these files when you submit your corrections to this eproof .
Q2	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article.
Q3	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.
Q4	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly. Note that changes here will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding email address(es).
Q5	You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes. If these are not considered scientific changes, any altered Supplementary files will not be used, only the originally accepted version will be published.

Author Queries

Journal: **Nature**
Paper: **s41586-020-03170-y**
Title:**Elevated NSD3 histone methylation activity drives squamous cell lung cancer**

AUTHOR:
The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query Reference	Reference
Q6	If applicable, please ensure that any accession codes and datasets whose DOIs or other identifiers are mentioned in the paper are scheduled for public release as soon as possible, we recommend within a few days of submitting your proof, and update the database record with publication details from this article once available.

nature research

Corresponding author(s): Or Gozani, Pawel K. Mazur, Lukasz Jaremko, Ning-Yi Shao

Last updated by author(s): 11/22/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

HISAT2 (ccb.jhu.edu/software/hisat2).
DESeq2 package from Bioconductor (bioconductor.org/packages/release/bioc/html/DESeq2).
GSEA (gsea-msigdb.org/gsea, 4.0.0).
GeneAnswers package from Bioconductor (bioconductor.org/packages/release/bioc/html/GeneAnswers, 2.32.0).
Cut&RunTools (bitbucket.org/qzhudfci/cutruntools/src/master/, version: 49ddd24).
Picard package (broadinstitute.github.io/picard, 2.23.8).
ngs.plot (github.com/shenlab-sinai/ngsplot, 2.63).

Data analysis

RNA-seq analysis.
The PSCN cells ±NSD3 depletion RNA-seq reads were aligned to mouse genome mm10 by HISAT2 (ccb.jhu.edu/software/hisat2). Then the transcriptomes were annotated to GENCODE M23 (gencodegenes.org)83 by FeatureCounts (2.0.0). Differentially expressed genes (DEGs) were detected by DESeq2 package from Bioconductor (bioconductor.org/packages/release/bioc/html/DESeq2) using likelihood ratio test (LRT, adjusted P-value < 0.001). 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) 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 robust t-test, and the trim level was 0.1.

CUT&RUN sequencing data analysis.
The CUT&RUN sequencing data were aligned by Cut&RunTools (bitbucket.org/qzhudfci/cutruntools/src/master/, version: 49ddd24). The reads were trimmed by Trimmomatic, then aligned to the mouse genome mm10 by Bowtie2 (bowtie-bio.sourceforge.net/bowtie2). 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). 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 PSCN cells with control knockdown over NSD3 knockdown (dPeaks), the differences between them calculated, and compared to the distribution of dPeaks of DEGs and the genome-wide genes with robust two-sided t-test at the trim level 0.1. The intergenic region annotation was generated by Complement of BedTools, and then the average profile of the intergenic regions was visualized by ngs.plot.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

GEO accession number: GSE149482, GSE149272. Figure 3 h, i, j, Extended Data Figure 5 i, j, k, Extended Data Figure 6 a, b, c, d, e are associated with these data. The data are shared under the restrictions of Nature Press and Stanford University. All data were already released to the public.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For in vivo PDX studies, n=6 mice per group were used. For AALE transformation in vivo assays n=5 mice per group were used. For genetically defined animal models studies, n=8 to 12 mice per group were used. Animal numbers for in vivo studies were selected using power analysis alpha 0.05 and 80% power such that a minimum change of 32-49% could be detected on the observed data scale. No sample size calculation was performed for in vitro studies. The sample size (n) of each experiment is provided in the figure captions. Each quantitative experiment was performed in at least 3 replicates, each qualitative experiment was performed in at least two replicates. The expected effect sizes were large therefore 3 replicates were fully sufficient.
Data exclusions	No data were excluded. None of the mice with the appropriate genotype were excluded from this study or used in any other experiments.
Replication	In vitro experiments were repeated the indicated number of times or at least 3 independent times (e.g. methylation assays). Drug screening study was repeated twice. The tumor growth measurements were repeated twice with similar results. All other in vivo studies were performed once with indicated number of animals. Independent repeats and sample sizes, as well as statistical analyses and significance levels, are also indicated in the Figure legends or in the Methods section.
Randomization	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 PDX treatment in vivo studies, animals were evenly distributed such that each group had a similar mean and SEM at the start of the study.
Blinding	Treatment groups for the in vivo studies were blinded to the investigator. For in vitro experiments, investigators were not blinded during data collection or analysis. However, controls and samples were analyzed in exactly the same way.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
[ ]	[x] Antibodies
[ ]	[x] Eukaryotic cell lines
[x]	[ ] Palaeontology and archaeology
[ ]	[x] Animals and other organisms
[ ]	[x] Human research participants
[x]	[ ] Clinical data
[x]	[ ] Dual use research of concern

Methods

n/a	Involved in the study
[ ]	[x] ChIP-seq
[x]	[ ] Flow cytometry
[x]	[ ] MRI-based neuroimaging

Antibodies

Antibodies used

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), yH2AX (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-Actin (CST #4970), β-Tubulin (Millipore #05-661), KRT5 (CST #71536), p63 (Abcam #ab124762), TTF1 (Abcam #ab76013), Ki67 (BD Biosciences, #550609), Cleaved Caspase 3 (CST #9664), H3K36me2 (CST #2901, for immunofluorescence), GFP (Abcam #ab183734), NSD2 (PMID: 19808676), NSD3 (CST #92056S, for detecting both NSD3 and NSD3Short), in-house NSD3 antibody.

Validation

The antibodies were validated by firms or the following publication:

H3K36me2 Thermo Fisher Scientific 701767, validated by western blot with human cancer cell acid extracts; H3K36me3 Thermo Fisher Scientific MA5-24687, validated by western blot with human cancer cell acid extracts; H3K27me3 Thermo Fisher Scientific MA5-33081, validated by western blot with human cancer cell acid extracts; V5 Thermo Fisher Scientific R960-25, validated by western blot with whole cell lysates from 293T cells transiently overexpressing a V5 tag fusion protein; H3K36me1 Abclonal #A2364, validated by western blot with human cancer cell acid extracts; H3K4me2 CST #9725S, validated by western blot with human cancer cell whole cell lysates; H3K9me2 CST #4658S, validated by western blot with human cancer cell whole cell lysates; H3K27me2 CST #9728S, validated by western blot with human cancer cell whole cell lysates;H2AX CST #9718, validated by western blot with whole cell lysates from human cancer cells treated with DNA damage inducing reagent; SOX2 CST #3579, validated by western blot with whole cell lysates from NTERA2 and NCCIT cells; c-Myc CST #5605S, validated by western blot with whole cell lysates from control or c-Myc knockout HEK293 cells; Phospho-4E-BP1 CST #2855S, validated by western blot with whole cell lysates from HEK293 cells which were starved for 24 hours in serum-free medium and underwent a 1hr amino acid deprivation; 4E-BP1 CST #9644S, validated by western blot with whole cell lysates from control HeLa cells or HeLa cells with a targeted mutation in the gene encoding 4E-BP1; FGFR1 CST #9740S, validated by western blot with whole cell lysates from A-204 (FGFR1 positive), KG-1a (FGFR1 oncogenic partner-FGFR1 fusion), A172 (FGFR1 low), and HT-29 (FGFR1 negative) cells; H4K20me2 Abcam #ab9052, validated by western blot with human cancer cell whole cell lysates; H3 Abcam #ab1791, validated by western blot with human cancer cell extracts; BRD4 Abcam #ab128874, validated by western blot with whole cell lysates from control or BRD4 knockout HeLa cells; β-Tubulin Millipore 05-661, validated by western blot with human cancer cell whole cell lysates; KRT5 CST #71536, validated by immunohistochemical analysis of paraffin-embedded human mucoepidermoid carcinoma of the larynx; p63 Abcam #ab124762, validated by immunohistochemistry (Formalin/PFA-fixed paraffin-embedded sections) analysis of human breast tissue sections; TTF1 Abcam #ab76013, validated by immunohistochemistry (Formalin/PFA-fixed paraffin-embedded sections) analysis of human lung carcinoma tissue sections; Ki67 BD Biosciences #550609, validated by immunohistochemical staining of formalin-fixed human spleen and tonsil tissues; Cleaved Caspase 3 CST #9664, validated by immunohistochemical analysis of paraffin-embedded mouse embryo; H3K36me2 CST #2901, validated by immunohistochemical analysis of paraffin-embedded human gastric carcinoma; GFP Abcam #ab183734, validated by western blot, flow Cytometry analysis, immunofluorescence with Control and GFP-transfected 293T cells; NSD2 antibody, validated as described in PMID: 19808676; NSD3 CST#92056S, for detecting both NSD3 and NSD3Short, validated by western blot with human cancer cell whole cell lysates from A549 and COS-7 cell lines; in-house NSD3 antibody generated in Genemed Synthesis Inc., validated by western blot in several knout or knockdown cell lines including here Figure 1c, 3f, 3g and Extended Data Figure 7b, 7c, 9c.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

HEK-293T (ATCC #CRL-1573), KNS-62 (JCRB #IFO50358), SK-MES-1 (ATCC #HTB-58), HT1080 (ATCC #CCL-121), 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), H1869 (ATCC #CRL-5900), AALE were described before PMID: 12085236, mouse tumor derived cells were generated in Dr. Pawel Mazur's lab.

Authentication

Authenticated human cell lines were obtained directly from cell repositories ATCC, RIKEN, JCRB, ECACC, DSMZ. Primary mouse cancer lines were obtained from animals with indicated genotypes an confirmed by PCR genotyping.

Mycoplasma contamination

mycoplasma free

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory 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.
Wild animals	Studies did not involve wild animals.
Field-collected samples	Studies did not involve samples collected in the field.
Ethics oversight	All animal procedures were reviewed and approved by the MDACC Institutional Animal Care and Use Committee (IACUC 00001636, PI: Mazur)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	This study utilized previously collected in tissue repositories patient-derived xenografts and human formalin-fixed paraffin embed (FFPE) cancer samples that were completely de-identified to the investigators. 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).
Recruitment	No subjects were recruited for this retrospective study of patient-derived cancer xenograft tissue and formalin-fixed paraffin embed (FFPE) cancer samples.
Ethics oversight	This study was approved by the institutional review board of the University of Texas M.D. Anderson Cancer Center (protocol PA19-0435, PI: Mazur).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149482
Files in database submission	GSM4503568 H3K36me2 sgControl Sample 1 GSM4503569 H3K36me2 sgControl Sample 2 GSM4503570 H3K27me3 sgControl GSM4503571 H3K36me2 sgNSD3 Sample 1 GSM4503572 H3K36me2 sgNSD3 Sample 2 GSM4503573 H3K27me3 sgNSD3 GSM4912939 SgControl_IgG GSM4912940 SgNSD3_IgG GSM4503574 Total RNA-seq sgControl Sample 1 GSM4503575 Total RNA-seq sgControl Sample 2 GSM4503576 Total RNA-seq sgControl Sample 3 GSM4503577 Total RNA-seq sgNSD3 Sample 1 GSM4503578 Total RNA-seq sgNSD3 Sample 2 GSM4503579 Total RNA-seq sgNSD3 Sample 3
Genome browser session (e.g. UCSC)	N/A.

Methodology

Replicates	Two biological replicates for H3K36me2.
Sequencing depth	Pair end. Length: 101bp. H3K36me2 sgControl Sample 1: 45M/98.5% aligned H3K36me2 sgControl Sample 2: 46M/98.4% aligned

	H3K27me3 sgControl: 64M/97.1% aligned H3K36me2 sgNSD3 Sample 1: 50M/98.2% aligned H3K36me2 sgNSD3 Sample 2: 47M/98.6% aligned H3K27me3 sgNSD3: 61M/97.6% aligned SgControl_IgG: 60M/98.3% aligned SgNSD3_IgG: 44M/98.1% aligned
Antibodies	H3K36me2 (Thermo Fisher Scientific #701767), H3K27me3 (Thermo Fisher Scientific #MA5-33081), GFP (Abcam #ab183734, used as IgG control).
Peak calling parameters	$mac2bin/macs2 callpeak -t $workdir/$dir/"$base_file".bam -g mm -f BAMPE -n $base_file --$outdirbroad2 –broad –broad-cutoff 0.1 -B –SPMR 2> $logdir/"$base_file".broad.all.frag.dedup.macs2
Data quality	H3K36me2 sgControl Sample 1: 993 H3K36me2 sgControl Sample 2: 1018 H3K27me3 sgControl: 51602 H3K36me2 sgNSD3 Sample 1: 1431 H3K36me2 sgNSD3 Sample 2: 1295 H3K27me3 sgNSD3: 21891 IgG samples are negative controls, no peak calling implemented. SgControl_IgG: N/A SgNSD3_IgG: N/A
Software	Macs2 2.1.3.3 (in Cut&RunTools, version: 49ddd24)