Supporting Information (SI Appendix)

The RNA binding protein FXR1 is a new driver in the 3q26-29 amplicon and predicts poor prognosis in human cancers

Jun Qian, Mohamed Hassanein, Megan D. Hoeksema , Bradford K. Harris , Yong Zou, Heidi Chen, Pengcheng Lu, Rosana Eisenberg, Jing Wang, Allan V Espinosa, Xiangming Ji, Fredrick T. Harris, S.M. Jamshedur Rahman and Pierre P. Massion

Supplementary methods

Supplementary references

Supplementary figures 1-7 and legends

Supplementary tables 1-34

Supplementary methods

Array Comparative Genomic Hybridization (array CGH)

Digestion, labeling, and hybridization were performed at the Vanderbilt Genome Sciences Resource Core by following Agilent's protocol version 4.0 for Agilent Human Genome CGH 244A oligo microarrays. The hybridized arrays were washed and scanned using an Agilent scanner. The array images were then analyzed using Agilent Feature Extraction Software (version 9.5.3.1), which also performs dye normalization for the data. The array CGH data have been deposited in NCBI's Gene Expression Omnibus (GEO)(1) and are accessible through GEO Series accession number GSE40048 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=vtgbnumkuqmggng&acc=GSE40048).

Gene Expression Microarray

Evaluation of gene expression was performed at Vanderbilt Genome Sciences Resource Core using the Agilent Human Gene Expression 4 × 44K Microarray platform using manufacturer-recommended procedures for microarray-based one-color assay. The array was scanned and then analyzed using Agilent Feature Extraction Software (version 10.7.1.1). The expression data have been deposited in NCBI's Gene Expression Omnibus(1) and are accessible through GEO Series accession number GSE40074 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=rlezrggikiiaqry&acc=GSE40074).

Gene Copy Number and Expression Data Analysis

Array CGH data obtained (n=24) in our laboratory and from dataset GSE20393 (n=45) were analyzed using Agilent DNA Analytics Software (version 4.0) with ADM-2 algorithm. The average log2 ratio of 0.8 was defined as the cut-off for amplification or 0.3 for low level gain. Genomic positions are mapped to the hg18 genome build.

For mRNA microarray expression analysis on 24 SCCs and the samples in NCBI GEO database GSE31552, the raw data and associated sample information were loaded and processed by GeneSpring11 (Agilent Technologies). The normalized data of FXR1, FMR1, FXR2, ECT2 and PRKCI were used in GraphPad Prism 5.0 to calculate statistical significance (t test).

For TCGA datasets, copy number and mRNA expression data (IllumninaHiSeq_RNASeqV2) in 18 types of human cancers were downloaded using the UCSC Cancer Genomics Browser (2). Copy-number alteration determined using GISTIC was used for analysis. Total tumor samples of each type of cancer were summarized in *SI Appendix* Table S10 (updated May 2014). Gene copy number and mRNA expression in lung cancer cell lines were directly downloaded from the Cancer Cell line Encyclopedia database (3).

For other expression datasets publically available, total five lung adenocarcinoma datasets (caArray, GSE31210, GSE3141, GSE37745 and Bhattacharjee 2001, see detailed description in *SI Appendix,* Table S9)(4-8), seven breast cancer datasets (GSE3494, GSE4922, GSE2034, GSE2603, GSE11121, GSE7390 and METABRIC, see detailed description in *SI Appendix*, Table S11-18)(9-15), three ovarian cancer datasets (GSE14764, GSE26712 and GSE9891, see detailed description in *SI Appendix*, Table S27)(16-18) and one HNSC dataset (E-TABM-302, *SI Appendix*, Table S29-30)(19) were analyzed. Expression data from Gene Expression Omnibus (GEO) database were collected as CEL files, and normalization was done using the fRMA function of Bioconductor (20) or directly downloaded from Insilico Database (21) or KM plotter database (22). Head and neck SCC cohort E-TABM-302 is available at Arrayexpress database (http://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-302/). The METABRIC breast cancer cohort (EGAS00000000083) is available at the European Genome-phenome Archive (EGA).

Dual-Color Interphase FISH Assay

One TMA including 49 lung SCC tumors was used for assessment of FXR1 gene CN using dual color fluorescence in situ hybridization (FISH). BAC clones for specific genes were selected from different libraries: FXR1 (RP11-905N10, RP11-282H19) and centromeric chromosome 3 (CEP3) probes were purchased from Vysis. Tissue hybridization, washing, color detection and scoring were performed as described previously (23).

Cell Cultures, Plasmids, and Reagents

Five human squamous carcinoma cell lines (H520, HCC95, HCC15, H157, and SW900), five human lung adenocarcinoma cell lines (H1819, H1648, H2882, A549 and H23), two large-cell carcinoma cell lines (H1299 and H460), and one immortalized epithelial cell line (BEAS-2B) were purchased from the American Type Culture Collection. Dr. Jennifer A. Pietenpol kindly provided seven breast cancer and five head&neck

squamous cancer cell lines. Dr. Dineo Khabele kindly provided two ovarian cancer cell lines. HBEC3KT cells were a generous gift from Dr. John D. Minna (24, 25). HEK293T cells were a gift from Dr. David Carbone. All cancer cell lines were maintained in RPMI or DMEM with 10% fetal bovine serum (FBS). Beas-2b and HEK293T cells were maintained in DMEM with 10% FBS. HBEC3KT cells were plated on porcine collagen I-coated tissue culture dishes containing K-SFM media (Invitrogen). All cells were grown in 1 mM penicillin/streptomycin. The PKC inhibitor chelerythrine chloride and propidium iodide (PI) were purchased from Sigma-Aldrich. Chelerythrine was dissolved in DMSO and PI was dissolved in sterile water for in vitro experiments.

The expression plasmid wild type pHACE-PRKCI-WT tagged human influenza hemagglutinin (HA) was obtained from Addgene (plasmid# 21252). Dominant negative pHACE-PRKCI-DN was generated by ligating full-length open reading frames of PRKCI with a dominant negative K3M point mutation at the ATP binding site and then subcloned into pHACE digested with EcoRI (26). For overexpression of FXR1, full-length cDNA of FXR1 was PCR based on the plasmid pCMV-SPORT6-FXR1 (Openbiosystems), subcloned into BamhI and XhoI sites of pBabe-hygro vector (Addgene plasmid#1756), Flag tagged pCMV-Tag4A vector, or Flag tagged pBabe-hygro vector and then confirmed by double-strand sequencing.

RNAi Knockdown Assay

For transient siRNA transfection, siPORT™ NeoFX™ Transfection Agent , Silencer Negative Control siRNA , Opti-MEM I, Silencer Select pre-designed siRNA products were purchased from Invitrogen. The Two small interfering RNAs sequences are as follows: for siRNA-FXR1, (siRNA ID# s15612, 5’ CGAGCUGAGUGAUUGGUCAtt; antisense, UGACCUCACUCAGCUCGtc; siRNA ID# s15612, sense, CGAGCUGAGUGAUUGGUCAtt, antisense, UGACCAAUCACUCAGCUCGtc); for siRNA-PRKCI (siRNA ID# S11111, sense, GGAUAUGAUGGAGCAAAAAtt, antisense, UUUUUGCUCCAUCAUAUCCca; siRNA ID# s11112, sense, GUAAUUCCAUAUAAUCCUUtt, antisense, AAGGAUUAUAUGGAAUUACtg). siRNA were suspended in water at a concentration of 20 nmol/L. The transfections were performed according to the manufacturer’s instructions or as previously described (27). For stably silencing FXR1 in H520 cells, three individual pGIPZ lentiviral shRNA-FXR1 and one pGIPZ non-silencing shRNA lentiviral control vector was purchased from Open Biosystems. Transfection and transduction were conducted according to manufacturer’s

instructions or as previously described (27). Briefly, H520 cells were plated the day before infection and subsequently incubated with virus in 8 μg/ml polybrene for 6 h. Puromycin (1.5 μg) was added the following day. After selection, various cell colonies were tested for FXR1 expression using immunoblotting. The colonies showing knockdown of FXR1 were plated for soft agar colony formation assay, cell invasion assay and flow cytometry assay.

Retroviral Introduction of FXR1

Stable transfection of pBabe-hygro-FXR1 or pBabe-Flag-hygro-FXR1 and empty vector into HBEC3KT or BEAS-2B cells was conducted as previously described (27). Hygromycin was added 48 h post-transfection at a final concentration of 4 ug/mL to obtain stable colonies overexpressing FXR1. Protein expression was confirmed via immunoblotting with antibodies against FXR1 (Sigma-Aldrich).

Immunoblot Analysis

Prestige rabbit Anti-FXR1 antibody, rabbit anti β-actin was purchased from Sigma-Aldrich. Rabbit polyclonal ECT2 (H-300), normal rabbit IgG, normal mouse IgG were purchased from Santa Cruz Biotechnology. PKC iota (phospho T555 + T563) antibody (ab5813) was obtained from Abcam. Anti-PRKCI mouse monoclonal antibody was purchased from BD Biosciences. Rabbit polyclonal p-Erk1/2, polyclonal antibodies to poly (ADPribose) polymerase (PARP), and cleaved caspase-3 were obtained from Cell Signaling. Mouse anti-phosphoserine (clone 4A4) and rabbit anti-phosphothreonine were purchased from Millipore. Immunoblot analysis was done using standard procedures as described in our previous study (27), with detection using the enhanced chemiluminescence system (Pierce Biotechnology). Antibody dilutions for immunoblotting were 1:1,000. The blots were re-probed with an anti-β-actin antibody to correct for protein loading differences. Anti-rabbit and anti-mouse secondary antibodies were purchased from Promega.

Immunoprecipitation (IP) and IP of FXR1 RNA Protein Complexes

For regular immunoprecipitation (IP) or co-IP, isolated lung cancer cells were suspended in Ripa buffer supplemented with 1mM PMSF and protease inhibitor cocktail and kept on ice for 15 min. Soluble extracts were separated by centrifugation at 10,000 x g for 10 min. Cell extracts (300–500 ug of total proteins) were precleared for 1 h on Protein A/G PLUS-Agarose beads (Santa Cruz Biotechnology) and collected as

supernatant after centrifugation for 1 min at 1000 x g. Supernatants were then incubated with 1-5 ug of various primary antibodies for 3 h at 4°C under constant shaking. Protein A-Dyna beads (Invitrogen) were added into the immunocomplexes for 1 additional hour. Hence, beads were washed three times with lysis buffer and adsorbed proteins were eluted in SDS-sample buffer for immunoblot analysis. In some experiments, HEK293T cells were transiently transfected with Flag-tagged pCMV-Tag4A-FXR1 and either HA-tagged pHACE-PRKCI-WT or pHACE-PRKCI-DN for 48h. Cell lysates were then subjected to IP with anti-HA or anti-Flag using protein A-Dyna beads.

Immunoprecipitation (IP) of endogenous RNA protein complexes (RNP) was done according to previously described methods (28, 29). Briefly, cytoplasmic lysates, prepared from H520 cells using lysis buffer (100 mM NaCl, 10 mM MgCl2, 30 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol [DTT], protease inhibitor cocktail, 100U/ml RNasin), supplemented with 0.1% Nonidet P40, were divided into two equal parts and incubated (1 h, 4°C) with 100 ul of a 50% (v/v) suspension of Protein-A Dyna beads precoated with 10 ug each of rabbit IgG1 (Santa Cruz Biotechnology) or anti-FXR1 (Sigma-Aldrich). The beads were washed five times with NT2 buffer (50mM Tris–Cl,150mM NaCl, 1mM MgCl2 and 0.05% NP-40). Contaminating DNA was removed from the sample by incubating the beads in 100 ml of NT2 buffer supplemented with 2 ml DNase I (Invitrogen) for 30 min at 37 °C. The beads were washed twice in NT2 buffer and bound proteins were digested by incubating the beads in 100 ml NT2 buffer containing 0.1% SDS and supplemented with 0.5 mg/ml Proteinase K for 15 min at 55 °C.

mRNA Stability Assay

Actinomycin D (20 μg/mL, Sigma-Aldrich) was added to H520 shFXR1 knockdown cells or non-silencing shRNA control. Total RNA was extracted at 0, 2, 4, 6, 17, or 24 hours after the addition of actinomycin D, and quantitative real-time polymerase chain reaction (PCR) was performed (30). The mRNA decay curves were constructed and half-lives were calculated from the curves.

Cell Proliferation Assays

Lung cancer cells were transiently transfected in 6-well plates with either FXR1 siRNA or negative control

siRNA as described above. The direct CyQUANT assay (Invitrogen) was then performed following the manufacturer's instructions on the wells post-transfection at indicated times up to 6 days. A representative viability experiment is shown with average standard deviation (s.d.).

Soft Agar Colony Formation Assay

Anchorage-independent growth was assayed by the ability of cells to form colonies in soft agar. The bottom agar consisted of growth medium containing 10% fetal bovine serum and 0.8% agarose in 6-well plates. H520-shFXR1 and non-silencing (NT) cells were plated on top of the bottom agar at 20,000/well with growth medium containing 10% fetal bovine serum and 0.4% agarose. The cells were incubated at 37 °C in 5% CO2. Cell colonies were stained with MTT and visualized and quantified under a dissecting microscope (Olympus) after three weeks in culture.

Matrigel Cell Invasion Assay

We used 6.5-mm diameter Transwell inserts (Costar) with the 8-um pore membranes coated with matrigel (BD Biosciences) to assess the invasive potential of FXR1 knockdown H520 cells. A total of 5×10^4 H520-NT and H520-shFXR1 cells in 0.2 mL were prepared and seeded onto the upper chamber of a Transwell and the lower chamber was filled with 400 uL of RPMI with 10% fetal bovine serum. After a 24 h incubation period, the cells in the upper chamber that didn’t migrate were gently scraped away and adherent cells present on the lower surface of the insert were stained with MTT, photographed and counted.

RT-PCR

To measure ECT2 and GAPDH mRNA after IP of FXR1 RNP complex, RNA post IP was then extracted using phenol:chloroform and precipitated in the presence of Glycoblue (Invitrogen). Reverse transcription was carried out using the First-Strand cDNA Synthesis kit (Promega) with oligo d(T)18 primers (Applied Biosystems). Precipitated mRNA was then detected by RT-PCR and visualized on a 1.5% agarose gel by ethidium bromide staining. Alternatively, real-time PCRs were performed in triplicate with SYBR Green Mix (Bio-rad) on an iCycler apparatus (Bio-rad) with results normalized to BIM expression.

Flow Cytometry

Analysis of DNA content and cell cycle of H520 and H520-FXR1 knockdown cells was performed by flow cytometry using propidium iodide (PI; Sigma-Aldrich) staining as previous described (27). A total of 10,000 to 20,000 stained nuclei were subjected to flow cytometry analysis. Data were collected on a Becton Dickinson FACSCalibur flow cytometer using CellQuest Pro software (BD Biosciences). Cell cycle analysis was done using the ModFit LT software (Verity Software House). The percentage of cells in sub-G1 was considered apoptotic.

Mouse Xenograft Study

Various number (2-5x10^6 cells) of four cancer cell lines including H520, H1299, A2780 and SCC12 stably transfected with FXR1-shRNA knockdown (KD) and non-target (NT) shRNA control constructs were injected into the flanks of thirty-five athymic nude mice-foxn1^{nu} (Harlan Laboratories, Indianapolis, IN, USA). Tumor growth was monitored up to 7 weeks and measured with calipers. Immunohistochemistry for FXR1 and Ki-67 was performed. Images were taken and analyzed with a Olympus BX41TF microscope and Olympus cellSens software.

Immunohistochemistry study on FXR1 protein in NSCLC

Immunohistochemical staining was performed as previously described (31). Cytoplasmic staining of FXR1 was scored using the Allred system with slight modifications (32). Briefly, the staining index was considered as the sum of the intensity score (0, no staining; 1+, weak; 2+, moderate; 3+, strong) and the distribution score (0, no staining; 0.1, staining of 1%-9% of cells; 0.5, 10%-49% and 1 if >50% of cells). The final immunoreactivity H score was determined by multiplying the intensity and extent of positivity scores of stained cells, with the minimum score of 0 and a maximum score of 3. The median value of all the H scores was a priori chosen as the cutoff point for separating FXR1-high tumors from FXR1-low tumors.

Supplementary References:

1. Edgar R, Domrachev M, & Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research* 30(1):207-210.
2. Zhu J, *et al.* (2009) The UCSC Cancer Genomics Browser. *Nature methods* 6(4):239-240.
3. Barretina J, *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483(7391):603-607.
4. Director's Challenge Consortium for the Molecular Classification of Lung A, *et al.* (2008) Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nature medicine* 14(8):822-827.
5. Okayama H, *et al.* (2012) Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer research* 72(1):100-111.
6. Bild AH, *et al.* (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439(7074):353-357.
7. Botling J, *et al.* (2013) Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clinical cancer research : an official journal of the American Association for Cancer Research* 19(1):194-204.
8. Bhattacharjee A, *et al.* (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of the National Academy of Sciences of the United States of America* 98(24):13790-13795.
9. Miller LD, *et al.* (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proceedings of the National Academy of Sciences of the United States of America* 102(38):13550-13555.
10. Ivshina AV, *et al.* (2006) Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer research* 66(21):10292-10301.
11. Wang Y, *et al.* (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365(9460):671-679.
12. Minn AJ, *et al.* (2005) Genes that mediate breast cancer metastasis to lung. *Nature* 436(7050):518-524.
13. Schmidt M, *et al.* (2008) The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer research* 68(13):5405-5413.
14. Desmedt C, *et al.* (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clinical cancer research : an official journal of the American Association for Cancer Research* 13(11):3207-3214.
15. Curtis C, *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486(7403):346-352.
16. Denkert C, *et al.* (2009) A prognostic gene expression index in ovarian cancer - validation across different independent data sets. *The Journal of pathology* 218(2):273-280.
17. Bonome T, *et al.* (2008) A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer research* 68(13):5478-5486.
18. Tothill RW, *et al.* (2008) Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical cancer research : an official journal of the American Association for Cancer Research* 14(16):5198-5208.
19. Rickman DS, *et al.* (2008) Prediction of future metastasis and molecular characterization of head and neck squamous-cell carcinoma based on transcriptome and genome analysis by microarrays. *Oncogene* 27(51):6607-6622.
20. McCall MN, Bolstad BM, & Irizarry RA (2010) Frozen robust multiarray analysis (fRMA). *Biostatistics* 11(2):242-253.
21. Coletta A, *et al.* (2012) InSilico DB genomic datasets hub: an efficient starting point for analyzing genome-wide studies in GenePattern, Integrative Genomics Viewer, and R/Bioconductor. *Genome biology* 13(11):R104.
22. Gyorffy B, Lanczky A, & Szallasi Z (2012) Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. *Endocrine-related cancer* 19(2):197-208.
23. Massion PP, *et al.* (2002) Genomic copy number analysis of non-small cell lung cancer using array comparative genomic hybridization: implications of the phosphatidylinositol 3-kinase pathway. *Cancer research* 62(13):3636-3640.
24. Ramirez RD, *et al.* (2004) Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. *Cancer research* 64(24):9027-9034.
25. Sato M, *et al.* (2006) Multiple oncogenic changes (K-RAS(V12), p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer research* 66(4):2116-2128.
26. Soh JW & Weinstein IB (2003) Roles of specific isoforms of protein kinase C in the transcriptional control of cyclin D1 and related genes. *J Biol Chem* 278(36):34709-34716.
27. Qian J, Zou Y, Rahman JS, Lu B, & Massion PP (2009) Synergy between phosphatidylinositol 3-kinase/Akt pathway and Bcl-xL in the control of apoptosis in adenocarcinoma cells of the lung. *Mol Cancer Ther* 8(1):101-109.
28. Paronetto MP, *et al.* (2006) The nuclear RNA-binding protein Sam68 translocates to the cytoplasm and associates with the polysomes in mouse spermatocytes. *Mol Biol Cell* 17(1):14-24.
29. Lewis SM, *et al.* (2008) The eIF4G homolog DAP5/p97 supports the translation of select mRNAs during endoplasmic reticulum stress. *Nucleic acids research* 36(1):168-178.
30. Leclerc GJ, Leclerc GM, & Barredo JC (2002) Real-time RT-PCR analysis of mRNA decay: half-life of Beta-actin mRNA in human leukemia CCRF-CEM and Nalm-6 cell lines. *Cancer Cell Int* 2(1):1.
31. Hassanein M, *et al.* (2013) SLC1A5 mediates glutamine transport required for lung cancer cell growth and survival. *Clinical cancer research : an official journal of the American Association for Cancer Research* 19(3):560-570.
32. Olaussen KA, *et al.* (2006) DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *N Engl J Med* 355(10):983-991.

Supplementary Figure 1-7 legends:

Supplementary Figure 1. *FXR1* is co-amplified and co-overexpressed with *ECT2* and *PRKCI* in lung SCC. (A) Copy number alteration of *FXR1*, *ECT2* and *PRKCI* in 24 lung SCCs (GSE40048). "2" is a high-level amplification (log2 ratio>0.8), "1" indicates a low-level gain (log2 ratio>0.3), "0" is diploid, "-1" is a single-copy loss (heterozygous deletion). Frequencies of high-level amplification are shown as a percentage of all cases. (B) Tumors harboring *FXR1* amplification (n=13) express higher *FXR1* mRNA than tumors without *FXR1* amplification (n=11) (* p< 0.0001). (C) Positive correlation between CN of *FXR1* and mRNA expression in TCGA lung SCCs (n=257). Pearson correlation P value is shown. (D) Lung SCCs (T) possess higher *FXR1*, *ECT2* and *PRKCI* mRNA level than matched normal tissues (N) in GSE31552 dataset.(E) *FXR1* copy number gain is significantly higher than *FMR1* or *FXR2* in lung SCC. Data are obtained from our study (n=24, GSE40048), Kan, Z. et al 's study (n=45, GSE20393) and TCGA (n=257). (F) Only *FXR1* is overexpressed in 18 matched SCC tumors (T) compared to non-tumor tissues (N). Data is obtained from GSE31552. (G) Only *FXR1* is overexpressed in 267 lung SCCs (T) versus 35 normal lung (N). UP: upregulated in tumor. NS: not significant. DOWN: downregulated in tumor. Data is obtained from TCGA. (H) Copy number alteration of *FXR1*, *FMR1* and *FXR2* in cell lines of lung squamous carcinoma (LUSC), adenocarcinoma (LUAD), large cell carcinoma (LUL) and small cell carcinoma (LUS). Data is obtained from the Cancer Cell line Encyclopedia database. (I) Highest coexpression correlation is observed in lung SCC cell lines compared to other types of lung cancer cell lines. Spearman correlation r value is shown. ** $P < 0.001$. Data is obtained from the Cancer Cell line Encyclopedia database.

Supplementary Figure 2. FXR1 regulates lung cancer cell growth *in vitro* and *vivo*. (A) Downregulation of FXR1 protein in A549 cells treated with siRNA against *FXR1*. SC: scrambled siRNA control. si#1, si#2: two individual *FXR1* siRNA. (B) Effect of *FXR1*-shRNA on H520 anchorage-independent soft agar colony formation. Representative cell colonies stained with MTT in H520 NT and two stable FXR1 knockdown cell lines are shown. (C) Effect of *FXR1*-shRNA on H520 cell cycle distribution. Representative flow cytometry profiles and corresponding percentage of cells in G1, S, G2 phase and subG1 for each group are shown. (D) Effect of *FXR1*-shRNA on H520 cell invasion. Significant differences between H520 NT and H520/FXR1 KD cells are indicated; * $p< 0.05$, ** $p<0.01$. (E) Immunoblot analysis of HBEC3KT or BEAS-2B cells ectopically overexpressing FXR1 by using the indicated antibodies. (F) Effects of *FXR1* knockdown on tumorigenicity in nude mice. Tumors (mm^3) from four mice on indicated days are shown on left. Scale bars, 5 mm. Five mice with tumors (mm^3) at 6 weeks are shown on right . The P values (determined by Student's t test) are relative to tumors formed by shRNA control cells. Scale bars, 5 mm. (G) Representative H&E staining, immunohistochemistry staining of FXR1 and Ki67 in tumors formed by H520 NT and H520 KD cells in mice #246 and #247. Magnification in all images is 200x. Scale bars, 100 um. (H) Effects of FXR1 knockdown on tumorigenicity of H1299 cancer cells in nude mice. Five mice with tumors (mm^3) at four weeks are shown. The P values (determined by Student's t test) are relative to tumors formed by shRNA control (NT) cells.

Supplementary Figure 3. FXR1 regulates ECT2 and PRKCI expression via distinct mechanisms. (A) Downregulation of ECT2, PRKCI and phospho-ERK1/2 in *FXR1* knockdown H1299 and H520 cells using siRNA against *FXR1*. (B) Immunoprecipitation of FXR1 or PRKCI co-precipitates PRKCI or FXR1 but not ECT2 in HCC95, A549 and H520 cells. (C) Immunoprecipitation of PRKCI or PRDA6A reveals that FXR1 binds to PRKCI but not PARD6A

in H520 cells. (D) Immunoprecipitation of PRKCI or PRDA6A reveals that FXR1 binds to PRKCI but not PARD6A in HCC95 or A549 cells. (E) Immunoprecipitation of endogenous ECT2 does not co-precipitate either FXR1 or PRKCI in H520 and A549 cells. (F) Immunoprecipitation of FXR1 reveals that FXR1 binds to PRKCI in a panel of NSCLC cells. (G) Immunoprecipitation with PRKCI co-precipitates FXR1 in H520 cells but not normal immortalized lung epithelial cell lines such as Beas2b, HBEC3KT or HBEC3KTR. Expression of FXR1 is confirmed in total cell lysates (input). (H) Immunoprecipitation of phospho-PRKCI T555 or FXR1 reveals that FXR1 binds to phosphorylated PRKCI T555 in H520 cells. (I) Immunoprecipitation of phosphothreonine or phosphoserine reveals that FXR1 is bound to phosphothreonine and phosphoserine in H520 cells, respectively. Co-precipitated PRKCI bound to phosphothreonine or phosphoserine was used as a positive control for IP. (J) FXR1 is a threonine phosphorylated protein in H520 cells. H520 cell lysates were precipitated using anti-FXR1 or anti-rabbit IgG and then blotted with anti-phophothorenine. Co-precipitated phospho-serine/ threonine PRKCI was used as a positive control. (K) Immunoprecipitation of phosphoserine co-precipitated FXR1 or PRKCI in HCC95, H1648 and H1819 cells. (L) Immunoprecipitation of phosphorylated PRKCI-T555 co-precipitates FXR1 in indicated NSCLC cell lines. (M) HEK293T cells were co-transfected with Flag-tagged pCMV-Tag4A-FXR1 and either HA-tagged pHACE-PRKCI-WT or pHACE-PRKCI-DN for 48h. Cell lysates were then subjected to IP with anti-HA, anti-Flag, anti-FXR1 or anti-PRKCI. The IP experiments reveal that active PRKCI but not the PRKCI DN mutant binds to FXR1. (N) H520 cells were treated with PKC inhibitor chelerythrine at the indicated dose or time course and total cell lysates were subjected to immunoblot analysis. The inhibition of PKC activity leads to the downregulation of phosphorylated PRKCI T555, FXR1 and apoptosis as indicated by cleaved PARP in H520 cells. (O) Chelerythrine reduces levels of phosphorylated PRKCI T555 and decreases both FXR1 and ECT2 expression in H520, HCC95 and H1299 cells. (P) Quantitative RT-PCR reveals significant fold enrichment of *ECT2* mRNA in lysate prepared from H520 cells subjected to immunoprecipitation using a normal IgG or anti-

FXR1. (Q) Half-life of *ECT2* mRNA is decreased in H520/FXR1 KD cells (10.8±2.4h) compared to H520/NT cells (24.2±3.5 h), $p<0.01$.

Supplementary Figure 4. FXR1 expression is correlated with non-small cell lung cancer squamous histology, disease stage and smoking status. *P* value was calculated using Kruskal-Wallis test. Characteristics of the patients are summarized in *SI Appendix,* Table S5.

Supplementary Figure 5. FXR1 overexpression correlates with poor clinical outcomes in multiple cancers. (A) Copy number alteration of *FXR1* in TCGA datasets including lung squamous cell carcinoma (LUSC, n=343), ovarian cancer (OV, n=559), cervical cancer (CESC, n=102) and head & neck squamous carcinoma (HNSC, n=306). The value are derived from the copy number analysis algorithms GISTIC. "2" is a high-level amplification, "1" indicates a low-level gain, "0" is diploid, "-1" is a single-copy loss (heterozygous deletion). Frequencies of high-level amplification were shown as a percentage of all cases. (B) FXR1 is overexpressed in basal-like (n=331) and luminal B (Lu-B, n=492) subtype of breast cancer (METABRIC cohort, n=1992). P value was calculated using Kruskal-Wallis Test. FXR1 is overexpressed in TCGA ovarian tumors (n=568) compared with in normal samples (n=8). FXR1 is overexpressed in TCGA HNSC (n=497) compared with in normal samples (n=43). (C) Immunoblot analysis of a panel of 14 cancer cell lines including seven breast cancer cell lines, five HNSC cell lines and two ovarian cancer cell lines for FXR1 and β-actin. FXR1 is overexpressed in 11 out of 14 examined cancer cell line. (D) Downregulation of FXR1 protein in A2780 cells treated with shRNA against FXR1. NT: non-target shRNA control. KD: shRNA FXR1 knockdown. (E) Effects of FXR1 knockdown on tumorigenicity in nude mice. A2780 NT and A2780 FXR1 KD cells were injected subcutaneously into the flanks of nude mice (n=5). Tumor volume was measured twice a week in all experiments by caliper and calculated by the formula: V = 3.14 (smaller

diameter)^2(larger diameter)/6. The quantification of tumor volume over a 4-week period is shown; * $p<0.05$. (F) Representative H&E staining, immunohistochemistry staining of FXR1 and Ki-67 in tumors formed by A2780 NT and A2780 KD in one mouse was shown on left. Scale bars, 50 um. (G) Effects of FXR1 knockdown on tumorigenicity of A2780 and SCC12 cancer cells in nude mice. Five mice with tumors (mm^3) per cell line at four weeks are shown. The P values (determined by Student's t test) are relative to tumors formed by shRNA control (NT) cells.

Supplementary Figure 6. Higher *FXR1* mRNA level in both tumor (T) and normal lung tissues (N) compared to *FMR1* or *FXR2* in lung squamous cell carcinoma (LUSC) or head and neck squamous cell carcinoma (HNSC). Data is obtained from TCGA.

Supplementary Figure 7. FXR1 mRNA level is correlated with smoking history. In contrast to mRNA levels of FMR1 and FXR2, FXR1 mRNA level is significantly higher in cancer patients who are smokers (Student's t test) in TCGA lung (n=971) and head&neck squamous carcinoma (HNSC, n=497) cohorts.

Figure. S1

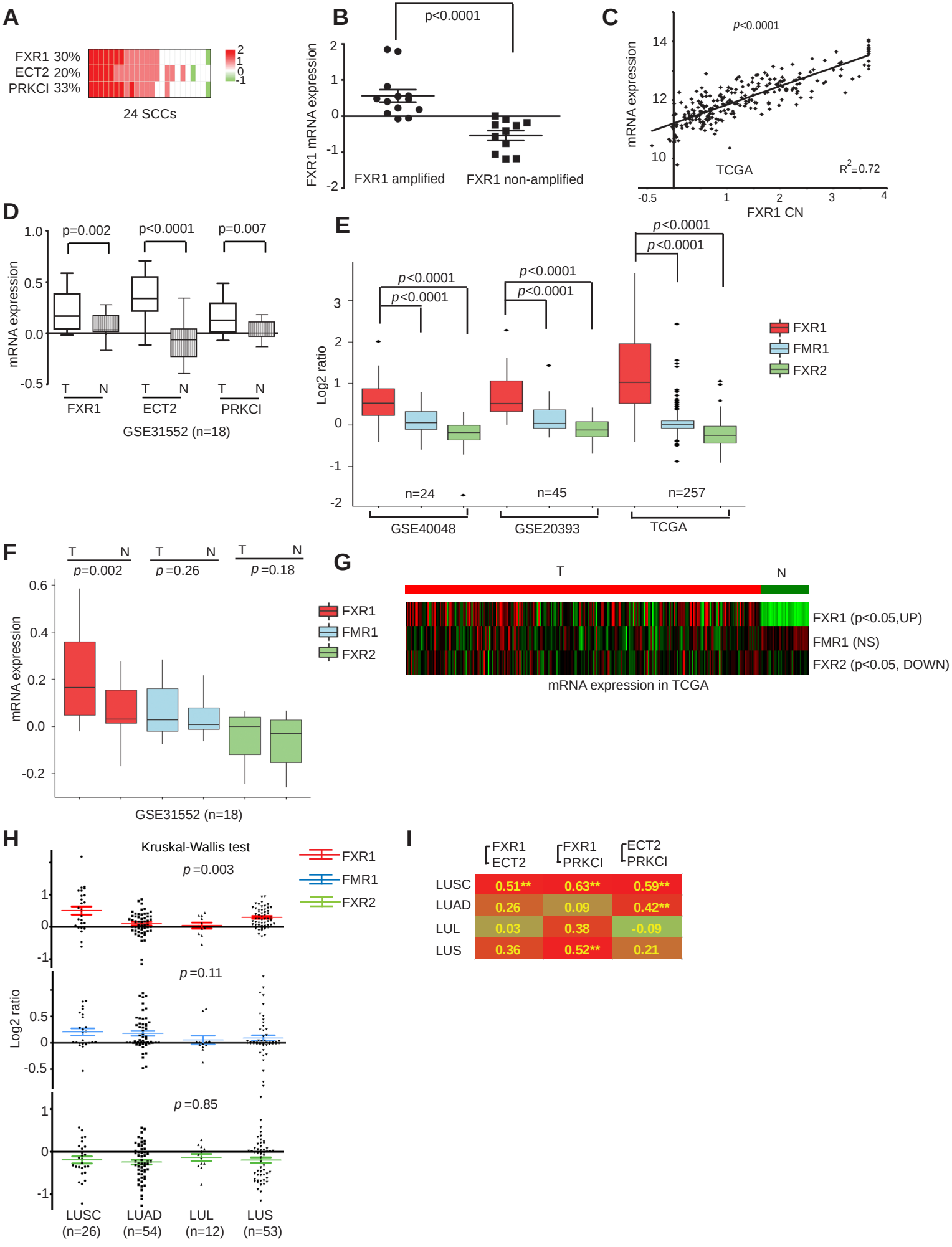
A
FXR1 30%
ECT2 20%
PRKCI 33%
24 SCCs
B
p<0.0001
FXR1 mRNA expression
FXR1 amplified
FXR1 non-amplified
C
p<0.0001
mRNA expression
TCGA
R2=0.72
FXR1 CN
D
p=0.002
p<0.0001
p=0.007
mRNA expression
T N
FXR1
ECT2
PRKCI
GSE31552 (n=18)
E
p<0.0001
p<0.0001
p<0.0001
p<0.0001
p<0.0001
p<0.0001
Log2 ratio
n=24
n=45
n=257
GSE40048
GSE20393
TCGA
FXR1
FMR1
FXR2
F
T N
p=0.002
p =0.26
p =0.18
mRNA expression
GSE31552 (n=18)
G
T
N
FXR1
FMR1
FXR2
FXR1 (p<0.05,UP)
FMR1 (NS)
FXR2 (p<0.05, DOWN)
mRNA expression in TCGA
H
Kruskal-Wallis test
p =0.003
p =0.11
p =0.85
Log2 ratio
FXR1
FMR1
FXR2
LUSC (n=26)
LUAD (n=54)
LUL (n=12)
LUS (n=53)
I
FXR1 ECT2
FXR1 PRKCI
ECT2 PRKCI
LUSC 0.51** 0.63** 0.59**
LUAD 0.26 0.09 0.42**
LUL 0.03 0.38 -0.09
LUS 0.36 0.52** 0.21

Figure. S2

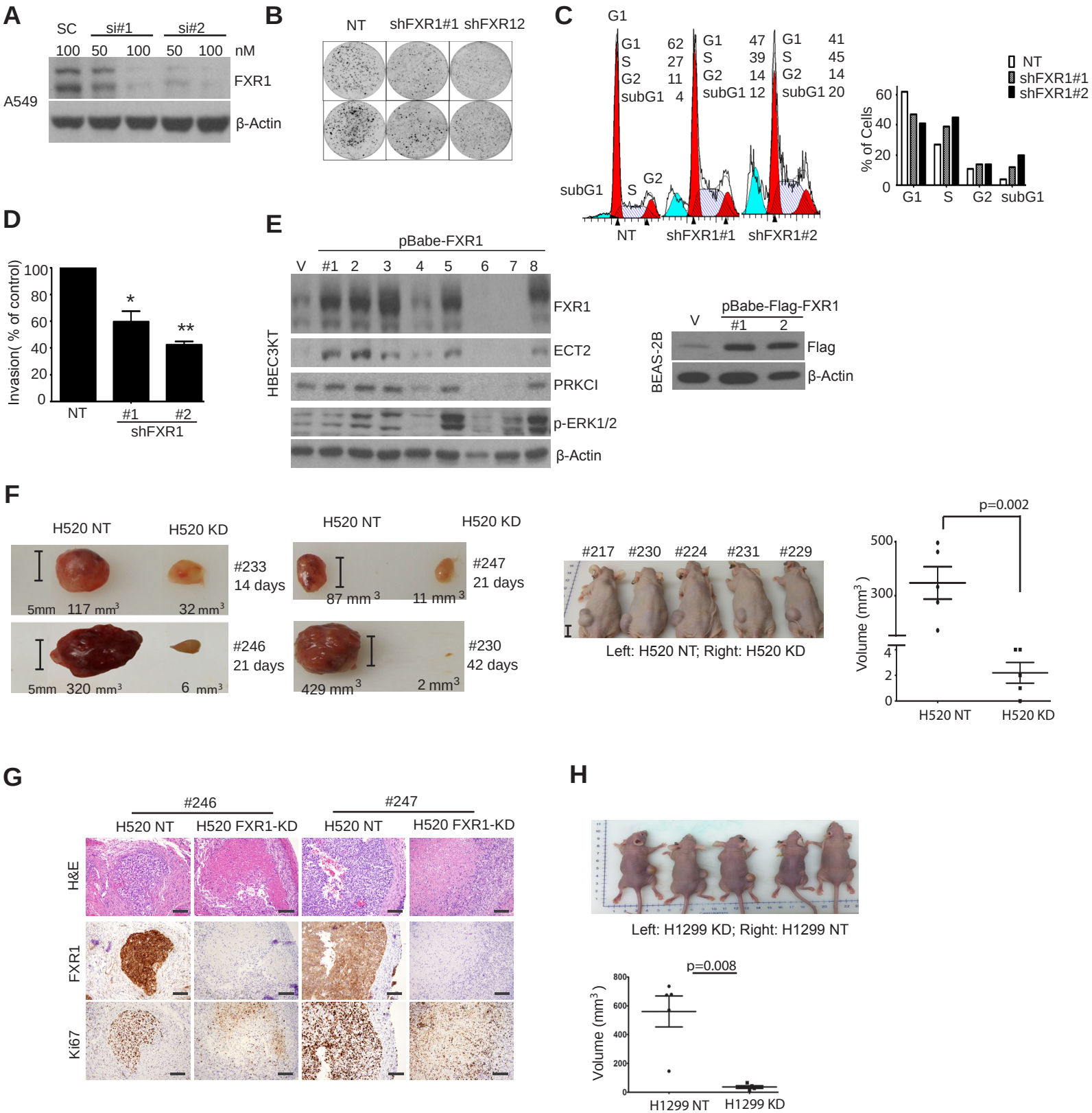
A
SC
si#1
si#2
100 50 100 50 100 nM
A549
FXR1
β-Actin
B
NT
shFXR1#1
shFXR12
C
G1
S
G2
subG1
G1 62
S 27
G2 11
subG1 4
G1 47
S 39
G2 14
subG1 12
G1 41
S 45
G2 14
subG1 20
NT
shFXR1#1
shFXR1#2
% of Cells
NT
shFXR1#1
shFXR1#2
D
Invasion(% of control)
NT
#1
#2
shFXR1
E
pBabe-FXR1
V #1 2 3 4 5 6 7 8
HBEC3KT
FXR1
ECT2
PRKCI
p-ERK1/2
β-Actin
pBabe-Flag-FXR1
V
#1 2
BEAS-2B
Flag
β-Actin
F
H520 NT
H520 KD
#233
14 days
5mm 117 mm3
32 mm3
#246
21 days
5mm 320 mm3
6 mm3
#247
21 days
87 mm3
11 mm3
#230
42 days
429 mm3
2 mm3
#217 #230 #224 #231 #229
Left: H520 NT; Right: H520 KD
p=0.002
Volume (mm3)
H520 NT
H520 KD
G
#246
#247
H520 NT
H520 FXR1-KD
H520 NT
H520 FXR1-KD
H&E
FXR1
Ki67
H
Left: H1299 KD; Right: H1299 NT
p=0.008
Volume (mm3)
H1299 NT
H1299 KD

Figure S3

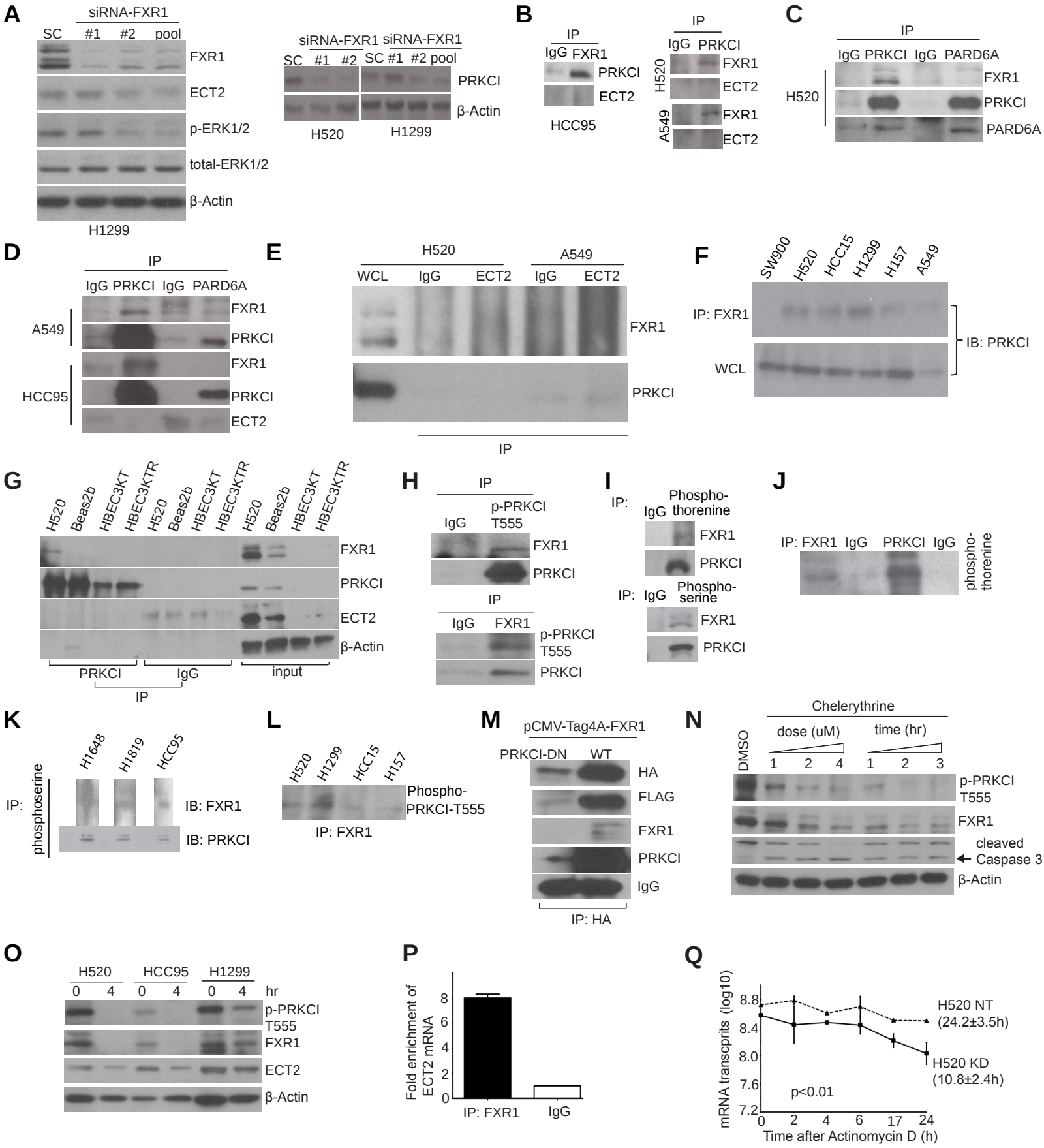
A
siRNA-FXR1
SC #1 #2 pool
FXR1
ECT2
p-ERK1/2
total-ERK1/2
β-Actin
H1299
siRNA-FXR1 siRNA-FXR1
SC #1 #2 SC #1 #2 pool
PRKCI
β-Actin
H520 H1299
B
IP
IgG FXR1
PRKCI
ECT2
HCC95
IP
IgG PRKCI
H520
FXR1
ECT2
A549
FXR1
ECT2
C
IP
IgG PRKCI IgG PARD6A
H520
FXR1
PRKCI
PARD6A
D
IP
IgG PRKCI IgG PARD6A
A549
FXR1
PRKCI
HCC95
FXR1
PRKCI
ECT2
E
H520
WCL IgG ECT2
A549
IgG ECT2
FXR1
PRKCI
IP
F
SW900 H520 HCC15 H1299 H157 A549
IP: FXR1
WCL
IB: PRKCI
G
H520 Beas2b HBEC3KT HBEC3KTR H520 Beas2b HBEC3KT HBEC3KTR H520 Beas2b HBEC3KT HBEC3KTR
FXR1
PRKCI
ECT2
β-Actin
PRKCI IgG input
IP
H
IP
p-PRKCI T555
IgG
FXR1
PRKCI
IP
IgG FXR1
p-PRKCI T555
PRKCI
I
IP:
Phospho-thorenine
IgG
FXR1
PRKCI
IP:
IgG
Phospho-serine
FXR1
PRKCI
J
IP: FXR1 IgG PRKCI IgG
phospho-thorenine
K
H1648 H1819 HCC95
IP:
phosphoserine
IB: FXR1
IB: PRKCI
L
H520 H1299 HCC15 H157
Phospho-PRKCI-T555
IP: FXR1
M
pCMV-Tag4A-FXR1
PRKCI-DN WT
HA
FLAG
FXR1
PRKCI
IgG
IP: HA
N
Chelerythrine
DMSO
dose (uM)
1 2 4
time (hr)
1 2 3
p-PRKCI T555
FXR1
cleaved Caspase 3
β-Actin
O
H520 HCC95 H1299
0 4 0 4 0 4 hr
p-PRKCI T555
FXR1
ECT2
β-Actin
P
Fold enrichment of ECT2 mRNA
8
6
4
2
0
IP: FXR1
IgG
Q
mRNA transcprits (log10)
8.8
8.4
8.0
7.6
7.2
H520 NT (24.2±3.5h)
H520 KD (10.8±2.4h)
p<0.01
0 2 4 6 17 24
Time after Actinomycin D (h)

Figure. S4

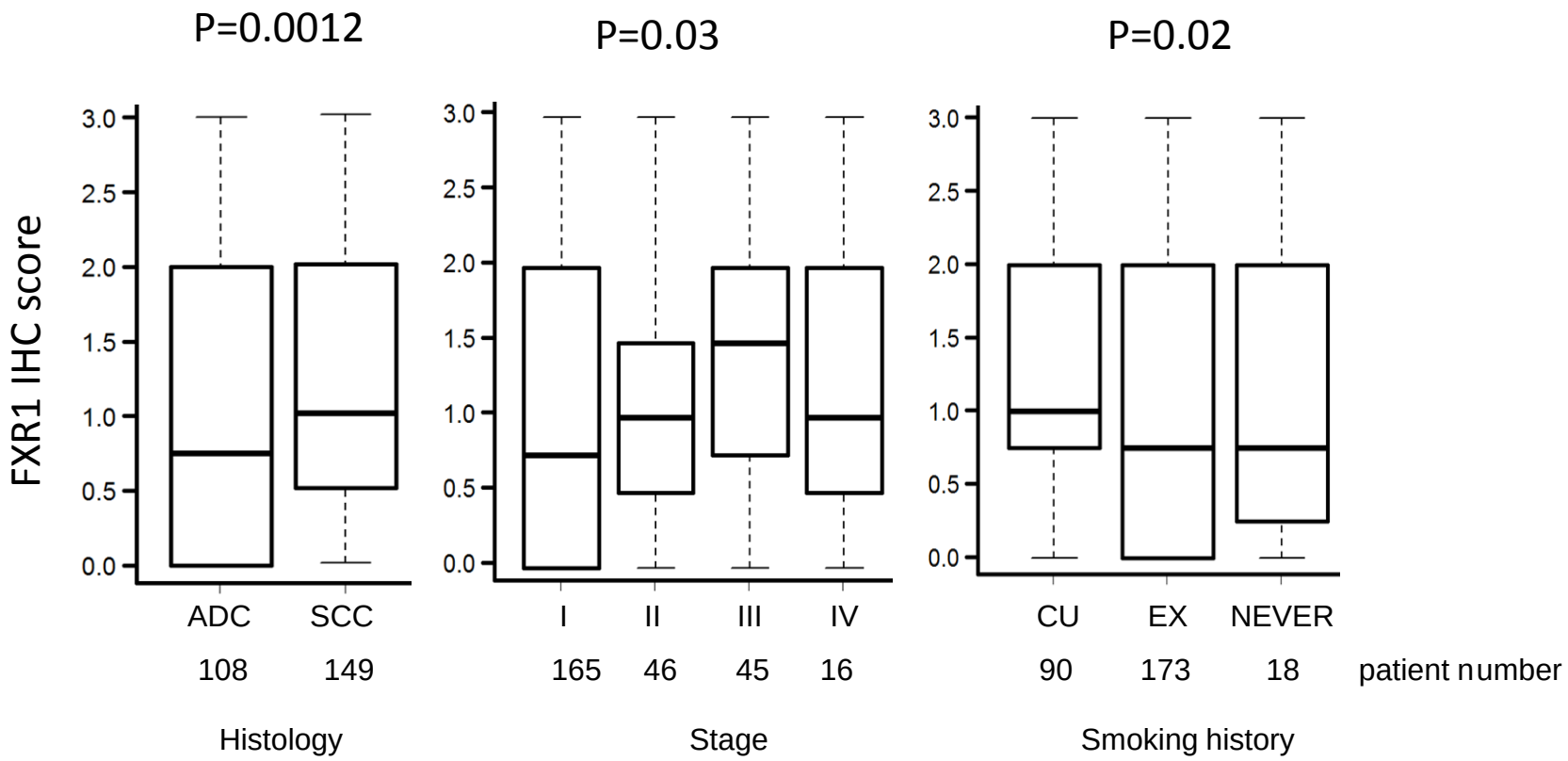
P=0.0012
P=0.03
P=0.02
FXR1 IHC score
3.0
2.5
2.0
1.5
1.0
0.5
0.0
ADC
SCC
108
149
Histology
I
II
III
IV
165
46
45
16
Stage
CU
EX
NEVER
90
173
18
Smoking history
patient number

Figure. S5

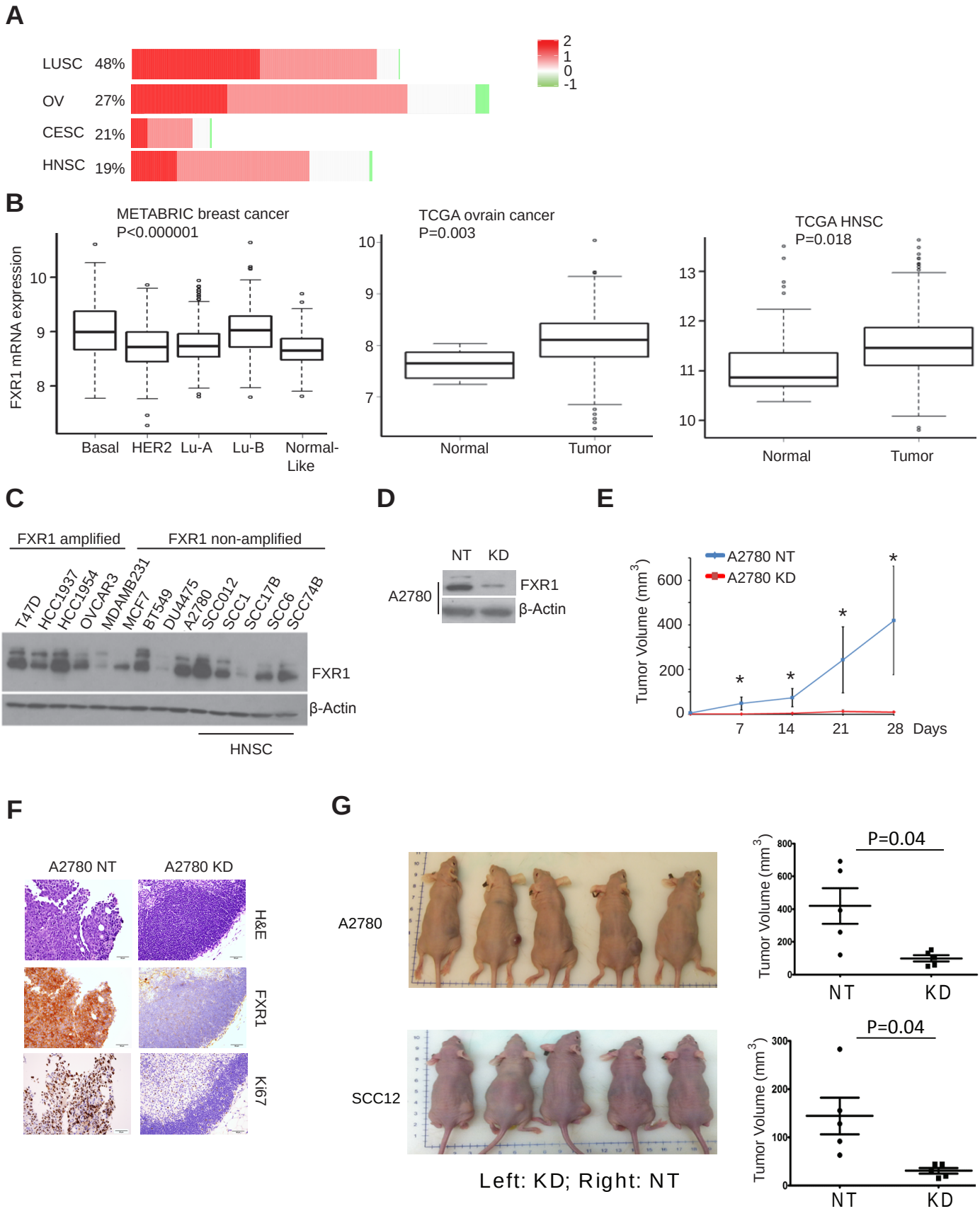
A
LUSC 48%
OV 27%
CESC 21%
HNSC 19%
2
1
0
-1
B
METABRIC breast cancer
P<0.000001
FXR1 mRNA expression
Basal
HER2
Lu-A
Lu-B
Normal-Like
TCGA ovrain cancer
P=0.003
Normal
Tumor
TCGA HNSC
P=0.018
Normal
Tumor
C
FXR1 amplified
FXR1 non-amplified
T47D
HCC1937
HCC1954
OVCAR3
MDAMB231
MCF7
BT549
DU4475
A2780
SCC012
SCC1
SCC17B
SCC6
SCC74B
FXR1
β-Actin
HNSC
D
NT
KD
A2780
FXR1
β-Actin
E
A2780 NT
A2780 KD
Tumor Volume (mm3)
Days
F
A2780 NT
A2780 KD
H&E
FXR1
Ki67
G
A2780
SCC12
Left: KD; Right: NT
P=0.04
Tumor Volume (mm3)
NT
KD
P=0.04
Tumor Volume (mm3)
NT
KD

Figure. S6

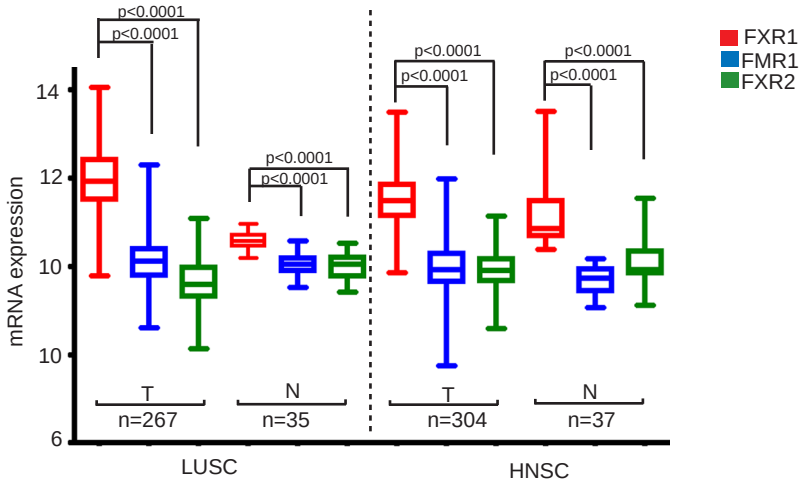
p<0.0001
p<0.0001
p<0.0001
p<0.0001
p<0.0001
p<0.0001
p<0.0001
p<0.0001
FXR1
FMR1
FXR2
mRNA expression
14
12
10
10
6
T
N
T
N
n=267
n=35
n=304
n=37
LUSC
HNSC

Figure S7

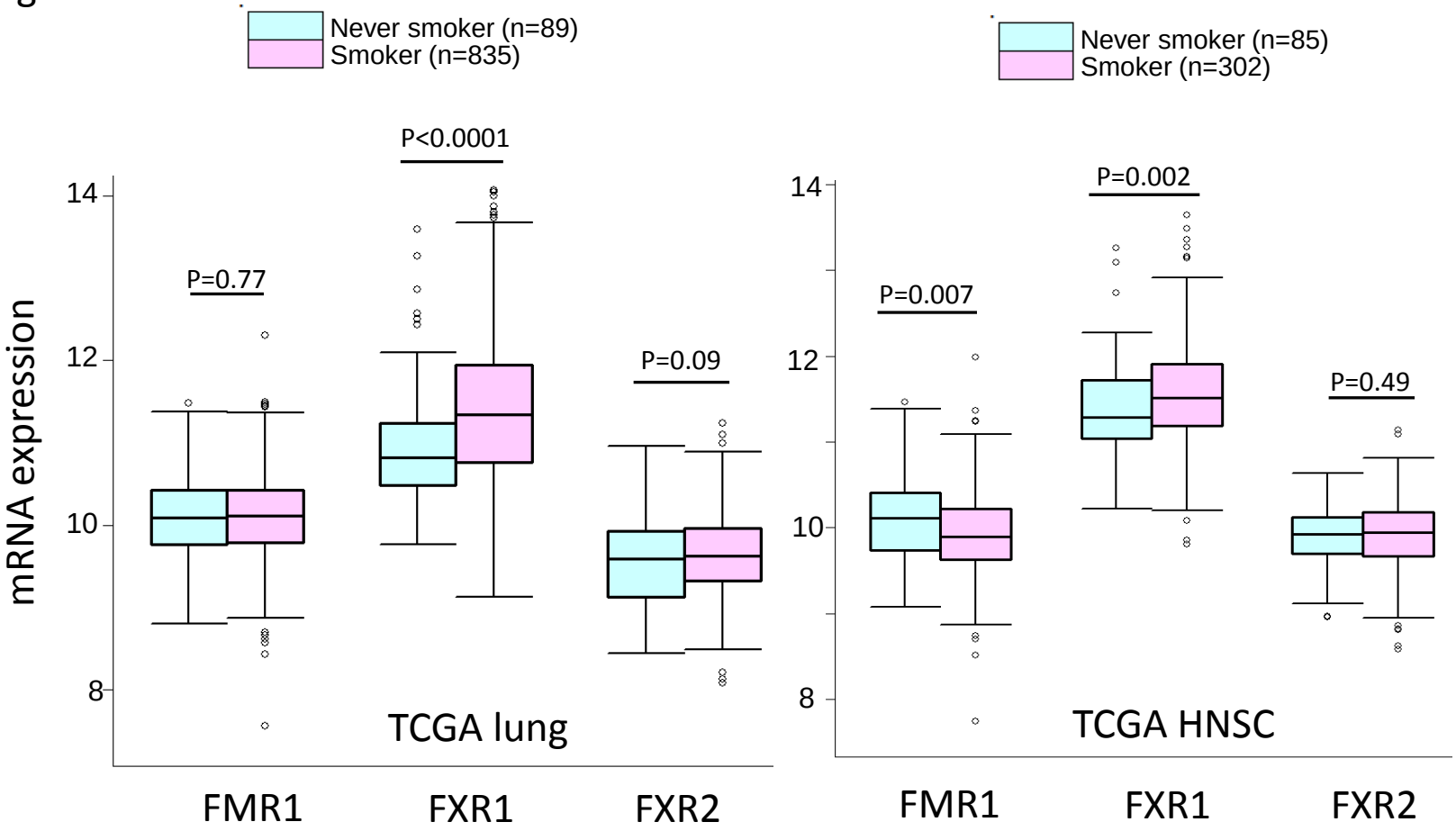
Never smoker (n=89)
Smoker (n=835)
mRNA expression
14
12
10
8
P=0.77
P<0.0001
P=0.09
TCGA lung
FMR1
FXR1
FXR2
Never smoker (n=85)
Smoker (n=302)
14
12
10
8
P=0.007
P=0.002
P=0.49
TCGA HNSC
FMR1
FXR1
FXR2

Table S1. Most highly correlated genes to FXR1 on 3q amplicon in 19 SCC tumors

Gene Symbol	Pearson Correlation
FXR1	1
ECT2	0.93850122
PSMD2	0.923744576
POLR2H	0.9173129
ABCF3	0.9071726
EIF4A2	0.903446895
ABCC5	0.898564126
RFC4	0.888882401
PARL	0.8847963
SENP2	0.8779619
SFRS10	0.8714842
RPL39L	0.868759553
MFN1	0.8675731
EIF2B5	0.8660367
PRKCI	0.860323008
MRPL47	0.8501934
DVL3	0.8288864
MYNN	0.8239362
PDCD10	0.82119603
CLCN2	0.8138976
MAGEF1	0.8001603
AP2M1	0.7947156
EIF4G1	0.790344
DNAJB11	0.7899947
SOX2	0.771896225
PLD1	0.7580384
PIK3CA	0.718942096
EHHADH	0.7161602
FETUB	0.7020755
SKIL	0.6934548

Table S2. Positive correlation between FXR1 and ECT2, PRKCI mRNA expression in lung SCCs.

Datasets	Genes	Pearson Correlation			
		SCC (n=50)		ADC (n=55)	
TCGA	FXR1_ECT2	0.78	p <0.0001	0.68	p <0.0001
	FXR1_PRKCI	0.77	p <0.0001	0.38	p =0.004
		SCC (n=18)		ADC (n=35)	
GSE31552	FXR1_ECT2	0.69	p =0.001	0.20	p =0.25
	FXR1_PRKCI	0.82	p <0.0001	0.49	p =0.003

Table S3. Most highly correlated genes to FXR1 on 3q amplicon in TCGA lung SCC (n=257)

Gene Symbol	Pearson Correlation	p Value
ZNF639	0.787925247	0
PIK3CA	0.761063701	0
PSMD2	0.741544455	0
DVL3	0.731518907	0
ATP11B	0.720195633	0
SENP2	0.714989202	0
DNAJC19	0.704312391	0
YEATS2	0.686002512	0
ACTL6A	0.685794741	0
DCUN1D1	0.67809138	0
MFN1	0.676019592	0
ALG3	0.675245529	0
* PRKCI	0.674432227	0
MRPL47	0.669660818	0
TBL1XR1	0.668826344	0
VPS8	0.668466414	0
MAGEF1	0.667226559	0
ABCF3	0.666559205	0
EIF2B5	0.666384745	0
* ECT2	0.65647635	0
ABCC5	0.642872105	0
PHC3	0.642526007	0
CLCN2	0.638330205	0
PARL	0.628753793	0
DNAJB11	0.621220659	0
OPA1	0.617025421	0
POLR2H	0.612882691	0
SOX2	0.609279715	0
MCCC1	0.606010135	0

Table S4. Positive correlation among FXR1 and ECT2, PRKCI gene CN in lung SCCs.

	TCGA (n=257)		GSE20393 (n=45)		GSE40048 (n=24)	
	r	*p*	r	*p*	r	*p*
FXR1_ECT2	0.88	*p* <0.0001	0.80	*p*<0.0001	0.87	*p*<0.0001
FXR1_PRKCI	0.86	*p* <0.0001	0.86	*p*<0.0001	0.94	*p*<0.0001
ECT2_PRKCI	0.96	*p*<0.0001	0.97	*p*<0.0001	0.89	*p*<0.0001

Table S5. Clinical characteristics of NSCLC patients (TMA, n=292)

Age	65±10.78
Gender	
Female	110 (39%)
Male	172 (61%)
Stage	
I	165 (60%)
II	46 (17%)
III	45 (17%)
IV	16 (6%)
Subtype	
ADC	108 (37%)
LCC	11 (4%)
SCC	149 (51%)
NSCLC	24 (8%)
Smoking history	
Current smoker	90 (32%)
Ex-smoker	173 (62%)
Never smoker	18 (6%)

Table S6. Multivariable Cox proportional hazards analysis of overall survival in 292 NSCLC patients.

Covariate	HR	95% C.I.		P
FXR1	1.27	1.09	1.49	0.003
Age	1.01	0.99	1.02	0.270
Gender (Male vs Female)	1.13	0.85	1.49	0.430
Stage				
II vs I	2.20	1.45	3.33	<0.0001
III vs I	2.55	1.85	3.52	<0.0001
IV vs I	3.92	2.60	5.90	<0.0001
Smoking history*				
EX vs CU	0.73	0.54	0.97	0.030
Never vs CU	0.48	0.25	0.91	0.024

*EX: ex-smoker, CU: current smoker

Table S7. Multivariable Cox proportional hazards analysis of overall survival in 161 stage I NSCLC patients.

Covariate	HR	95% C.I.		P
FXR1	1.32	1.05	1.65	0.016*
Age	1.01	0.99	1.03	0.499
Gender (Male vs Female)	1.13	0.73	1.74	0.579
Smoking history (EX vs CU)	1.11	0.71	1.76	0.642

Table S8. Multivariable Cox proportional hazards analysis of overall survival in 78 stage I lung SCC patients.

Covariate	HR	95% C.I.		P
FXR1	1.42	1.02	1.99	0.037*
Age	1.01	0.98	1.04	0.46
Gender (Male vs Female)	1.03	0.50	2.11	0.94
Smoking history (EX vs CU)	1.09	0.54	2.22	0.8

Table S9. Clinical characteristics of six lung ADC datasets

Dataset	Platform	Sample size	Median OS	NO.of deaths	Meidan RFS	No. of Relapse	Age	Gender (% male)	% of never smokers	Histrology (% A/S/L)	Stage (% 1/2/3/4)	Grade (% poor/ moderate/well)	% Adjuvant chemotherapy
caArray	GPL96	462	3.88	236	2.75	205	64±10	50	11	100/0/0	32/54/6/3	36/45/13	19
GSE31210	GPL570	226	4.85	35	4.54	64	59±7.4	46	51	100/0/0	74/26/0/0	-	7
GSE3141	GPL570	111	2.58	58	-	-		-	-	52/48/0	30/-	-	-
GSE37745	GPL570	196	3.44	145	1.68	49	63±9.2	55	-	54/34/12	66/18/14/2	-	15
TCGA ADC	RNAseq	488	1.08	117	0.92	96	65±9.7	42	15	100/0/0	55/23/17/5		1
Bhattacharjee		126	2.95	71	-	-	63±10	42	-	100/0/0	67/21/9/3	-	-
Total		1609	3.27	662	2.47	414	65±9.8	45% (n=724)	25% (n=294)	80/16/3	56/30/11/4	36/45/13	10% (n=141)

Table S10. Copy number and mRNA expression of FXR1 in multiple human cancers

Cancer Types		Amplification	Sample numbers[a]	mRNA expression[b]
Lung squamous cell carcinoma	(LUSC)	48.33%	490	**UP**
Ovarian serous cystadenocarcinoma	(OV)	27.07%	569	**UP**
Head and Neck squamous cell carcinoma	(HNSC)	21.35%	452	**UP**
Cervical squamous cell carcinoma	(CESC)	19.30%	171	NS
Uterine Corpus Endometrioid Carcinoma	(UCEC)	8.17%	514	NS
Stomach adenocarcinoma	(STAD)	6.65%	331	**UP**
Bladder Urothelial Carcinoma	(BLCA)	4.07%	221	NS
Breast invasive carcinoma	(BRCA)	3.57%	1007	**UP**[c]
Glioblastoma multiforme	(GBM)	3.04%	560	**UP**
Pancreatic adenocarcinoma	(PAAD)	2.47%	81	NS
Lung adenocarcinoma	(LUAD)	2.23%	493	**UP**
Prostate adenocarcinoma	(PRAD)	2.16%	278	NS
Liver hepatocellular carcinoma	(LIHC)	1.82%	165	**UP**
Kidney renal clear cell carcinoma	(KIRC)	1.79%	504	**UP**
Brain Lower Grade Glioma	(LGG)	1.10%	365	NA
Kidney renal papillary cell carcinoma	(KIRP)	0.00%	172	NS
Colon and Rectum adenocarcinoma	(COADREAD)	0.00%	589	**UP**
Thyroid carcinoma	(THCA)	0.00%	494	NS

a. The sample numbers are number of samples that only have been measured for copy number.
b. mRNA expression was calculated by the comparison between FXR1 level in tumors and normal tissues.
$p<0.05$ means significant change. UP: upregulated in tumors. NA: not available (due to a lack of the data from the normal tissues. NS: not significant. c. FXR1 mRNA is upregulated in basal-like subtype of breast cancer.

Table S11. Clinical characteristics of breast cancer patients (GSE3494, n=250)

Covariates	Patient number	(%)
Age (mean±SD)	62±14	
Tumor size		
<=20mm	128	(51%)
>20mm	123	(49%)
Estrogen receptor		(%)
Positive	213	(86%)
Negative	34	(14%)
Progesterone receptor		(%)
Positive	190	(76%)
Negative	61	(24%)
Grade		(%)
I	67	(27%)
II	128	(51%)
III	54	(22%)
Lymph Node status		(%)
Positive	84	(35%)
Negative	158	(65%)
p53 status		(%)
Mutant	58	(23%)
Wild type	193	(77%)
PAM50 subtype		(%)
Basal-like	34	(15%)
Her2	28	(12%)
Luminal A	83	(36%)
Luminal B	30	(13%)
Normal Like	53	(23%)

Table S12. Clinical characteristics of breast cancer patients (GSE4922, n=249)

Covariates	Patient number	(%)
Age (mean±SD)	62±14	
Tumor size		
<=20mm	126	(51%)
>20mm	123	(49%)
Estrogen receptor		(%)
Positive	211	(86%)
Negative	34	(14%)
Grade		(%)
I	68	(24%)
II	166	(57%)
III	55	(19%)
Lymph Node status		(%)
Positive	81	(34%)
Negative	159	(66%)
p53 status		(%)
Mutant	58	(23%)
Wild type	189	(77%)
PAM50 subtype		(%)
Basal-like	24	(12%)
Her2	36	(17%)
Luminal A	65	(31%)
Luminal B	85	(40%)

Table S13. Clinical characteristics of breast cancer patients (GSE2034, n=286)

Covariates	Patient number	(%)
Age (mean±SD)	52±12	
Tumor size		
<=20mm	8	(3%)
>20mm	278	(97%)
Estrogen receptor		(%)
Positive	209	(73%)
Negative	77	(27%)
Grade		(%)
I	7	(4%)
II	42	(21%)
III	148	(75%)
Lymph Node status		(%)
Negative	286	(100%)
PAM50 subtype		(%)
Basal-like	49	(17%)
HER2-enriched	9	(3%)
Luminal A	104	(36%)
Luminal B	124	(43%)

Table S14. Clinical characteristics of breast cancer patients (GSE2603, n=121)

Covariates	Patient number	(%)
Age (mean±SD)	56±14	
Tumor size		
<=20mm	15	(15%)
>20mm	84	(85%)
Estrogen receptor		(%)
Positive	57	(58%)
Negative	42	(42%)
Progesterone receptor		(%)
Positive	43	(44%)
Negative	55	(56%)
HER2		(%)
Positive	26	(30%)
Negative	62	(70%)
		(%)
Lymph Node status		(%)
Positive	64	(65%)
Negative	35	(35%)
PAM50 subtype		(%)
Basal-like	24	(24%)
Her2	15	(15%)
Luminal A	33	(33%)
Luminal B	7	(7%)
Normal-like	20	(20%)
van't Veer Signature		(%)
Good	37	(45%)
Bad	45	(55%)

Table S15. Clinical characteristics of breast cancer patients (GSE11121, n=200)

Covariates	Patient number	(%)
Age (mean±SD)	60±12	
Tumor size		
<=20mm	112	(56%)
>20mm	88	(44%)
Estrogen receptor		(%)
Positive	148	(80%)
Negative	36	(20%)
Grade		(%)
I	29	(15%)
II	136	(68%)
III	35	(17%)
Lymph Node status		(%)
Negative	200	(100%)
PAM50 subtype		(%)
Basal-like	38	(15%)
Her2	25	(12%)
Luminal A	17	(36%)
Luminal B	57	(13%)
Normal-like	63	(23%)

Table S16. Clinical characteristics of the TCGA breast cancer cohort of patients with invasive ductal carcinoma (n=453)

Covariates	Patient number	Percentage
Age (mean±SD)	57±13	
T stage		
T1	118	(26%)
T2	275	(61%)
T3	38	(9%)
T4	19	(4%)
M stage		
M1	12	(97%)
M0	437	(3%)
Lymph Node status		
Positive	221	(49%)
Negative	231	(51%)
Estrogen receptor		
Positive	339	(76%)
Negative	107	(24%)
Progesterone receptor		
Positive	288	(65%)
Negative	157	(35%)
HER2		
Positive	71	(16%)
Negative	364	(84%)
PAM50 subtype		
Basal-like	89	(20%)
HER2-enriched	54	(12%)
Luminal A	185	(43%)
Luminal B	111	(25%)
Death event	50	(12%)
Censeoring	377	(88%)

Table S17. Clinical characteristics of breast cancer patients (METABRIC, n=1992)

Covariates	Patient number (%)
Age (mean±SD)	61±13
stage	
I	372 (35%)
II	579 (55%)
III	90 (9%)
IV	10 (1%)
Tumor size	
<=20mm	862 (44%)
>20mm	1110 (56%)
Lymph Node status	
Positive	1042 (52%)
Negative	943 (48%)
Estrogen receptor	
Positive	1518 (76%)
Negative	474 (24%)
Progesterone receptor	
Positive	1049 (53%)
Negative	943 (47%)
HER2	
Positive	249 (12%)
Negative	1743 (88%)
PAM50 subtype	
Basal-like	331 (17%)
HER2	240 (12%)
Luminal A	721 (36%)
Luminal B	492 (25%)
Normal-like	202 (10%)

Table S18. Clinical characteristics of breast cancer patients (TRANSBIG, n=198)

Covariates	Patient number (%)
Age (mean±SD)	46±7.2
Tumor size	
<=20mm	102 (52%)
>20mm	96 (48%)
Grade	(%)
I	30 (15%)
II	83 (42%)
III	83 (42%)
Estrogen receptor	(%)
Negative	64 (32%)
Positive	132 (68%)
PAM50 subtype	(%)
Basal-like	30 (15%)
HER2	11 (6%)
Luminal A	64 (32%)
Luminal B	90 (45%)
Normal Like	3 (2%)

Table S19. Multivariate Cox proportional-hazards analysis of the risk of distant recurrence free survival in breast cancer patients (GSE3494, n=250)

	HR	95% C.I		P
FXR1	2.23	1.36	3.66	0.002
Age	1.01	0.99	1.03	0.23
Tumor size (>20mm vs <20mm)	1.65	0.97	2.83	0.07
Estrogen receptor (positive vs negative)	1.66	0.67	4.08	0.27
Progesterone receptor(positive vs negative)	0.89	0.42	1.89	0.76
Grade				
II vs I	1.11	0.58	2.15	0.75
III vs I	1.81	0.72	4.54	0.2
Lymph Node status (positive vs negative)	1.55	0.94	2.57	0.09
P53 status (wildtype vs mutant)	0.79	0.45	1.39	0.4
PAM50 subtype				
HER2-enriched vs basal-like	3.79	1.63	8.78	0.002
Luminal A vs basal-like	1.52	0.56	4.09	0.41
Luminal B vs basal-like	2.18	0.86	5.50	0.1
Normal-lIke vs basal-like	2.14	0.74	6.22	0.16

Table S20. Multivariate Cox proportional-hazards analysis of the risk of distant recurrence free survival in breast cancer patients (GSE4922, n=249)

	HR	95% C.I		P
FXR1	2.62	1.50	4.58	0.0006
Age	1.02	1.00	1.04	0.12
Tumor size (>20mm vs <20mm)	1.90	1.11	3.27	0.02
Estrogen receptor (positive vs negative)	1.37	0.64	2.96	0.41
Grade				
II vs I	1.11	0.57	2.18	0.75
III vs I	1.91	0.78	4.63	0.15
Node (positive vs negative)	1.35	0.82	2.24	0.24
p53 (mutant vs WT)	1.38	0.80	2.38	0.25
PAM50 subtype				
Her2-enriched vs basal-like	5.68	1.83	17.59	0.003
Luminal A vs basal-like	3.14	0.86	11.49	0.082
Luminal B vs basal-like	2.81	0.82	9.61	0.098

Table S21. Multivariate Cox proportional-hazards analysis of the risk of distant recurrence free survival in breast cancer patients (GSE2034, n=286)

Covariates	HR	95% C.I		P
FXR1	2.06	1.30	3.24	0.002
Age	0.99	0.97	1.00	0.14
Tumor size (>20mm vs <20mm)	0.63	0.09	4.55	0.64
Grade (III vs I and II)	2.22	1.12	4.39	0.02
Estrogen receptor (positive vs negative)	1.17	0.58	2.35	0.67
PAM50 subtype				
Basal-like vs other*	0.86	0.39	1.86	0.7

*other includes HER2-enriched (n=9), Luminal A (n=104) and Luminal B (n=124)

Table S22. Multivariate Cox proportional-hazards analysis of the risk of distant recurrence free survival in breast cancer patients (GSE2603, n=121)

	HR	95% C.I		P
FXR1	2.86	1.26	6.48	0.01
Age	1.03	1.00	1.07	0.06
Tumor size (>20mm vs <20mm)	0.62	0.23	1.71	0.35
Estrogen receptor (positive vs negative)	1.29	0.29	5.73	0.74
Progesterone receptor (positive vs negative)	1.30	0.40	4.28	0.66
HER2(positive vs negative)	0.21	0.04	1.03	0.05
Lymph node (positive vs negative)	0.95	0.39	2.35	0.91
PAM50 subtype				
HER2-enriched vs basal-like	2.75	0.70	10.86	0.14
Luminal A, B vs basal-like	0.49	0.08	3.02	0.44
Normal-IIke vs basal-like	1.13	0.29	4.48	0.86
van't Veer Signature (good vs bad)	0.74	0.24	2.26	0.59

Table S23. Multivariate Cox proportional-hazards analysis of the risk of distant recurrence free survival in breast cancer patients (GSE11121, n=200)

	HR	95% C.I		P
FXR1	1.21	1.44	7.75	0.005
Age	-0.01	0.96	1.03	0.71
Tumor size (>20mm vs <20mm)	0.20	0.61	2.46	0.57
Estrogen receptor (positive vs negative)	-0.21	0.32	2.04	0.66
Grade				
II vs I	0.24	0.41	3.89	0.68
III vs I	0.84	0.60	9.05	0.22
PAM50 subtype				
Her2-enriched vs basal-like	0.71	0.71	5.81	0.18
Luminal A vs basal-like	-0.03	0.17	5.63	0.97
Luminal B vs basal-like	-0.07	0.29	3.04	0.91
Normal-lIke vs basal-like	0.15	0.36	3.75	0.8

Table S24. Multivariate Cox proportional-hazards analysis of the risk of overall survival in the TCGA breast invasive ductal carcinoma cohort (n=453)

	HR	95% C.I		P
FXR1	2.75	1.11	6.84	0.03
Age	1.04	1.01	1.07	0.01
M stage (M1 vs M0)	2.26	0.76	6.76	0.14
T stage				
T2 vs T1	0.81	0.30	2.17	0.67
T3 vs T1	0.60	0.14	2.51	0.48
T4 vs T1	0.76	0.16	3.68	0.73
Lymph node status (positive vs negative)	1.73	0.77	3.91	0.19
Estrogen receptor (positive vs negative)	1.26	0.38	4.15	0.7
Progesterone receptor (positive vs negative)	0.32	0.12	0.81	0.02
HER2 status (positive vs negative)	0.35	0.09	1.37	0.13
PAM50 subtype				
HER2-enriched vs basal-like	3.29	0.83	13.01	0.09
Luminal A vs basal-like	1.30	0.37	4.57	0.68
Luminal B vs basal-like	2.48	0.71	8.61	0.15

Table S25. Multivariate Cox proportional-hazards analysis of the risk of overall survival in breast cancer patients (TRANSBIG, n=198)

	HR	95% C.I		P
FXR1	2.00	1.17	3.42	0.01
Age	1.01	0.98	1.05	0.45
Tumor size (>20mm vs <20mm)	1.39	0.76	2.55	0.29
Estrogen receptor (positive vs negative)	0.50	0.25	0.99	0.05
Grade				
II vs I	1.03	0.43	2.45	0.95
III vs I	0.68	0.26	1.79	0.43
PAM50 subtype				
Her2-enriched vs basal-like	3.41	1.04	11.19	0.04
Luminal A vs basal-like	0.90	0.28	2.89	0.86
Luminal B vs basal-like	1.83	0.72	4.63	0.21

Table S26. Multivariate Cox proportional-hazards analysis of the risk of disease specific free survival in breast cancer patients (METABRIC, n=1992)

	HR	95% C.I		P
FXR1	1.43	1.04	1.97	0.03
Age	1.01	1.00	1.02	0.25
Tumor size (>20mm vs <20mm)	1.56	1.12	2.18	0.01
Lymoph node status (positive vs negtive)	1.81	1.17	2.81	0.01
Grade				
II vs I	1.48	0.63	3.48	0.37
III vs I	1.76	0.74	4.16	0.2
Stage				
II vs I	0.90	0.54	1.52	0.7
III/IV vs I	2.47	1.29	4.73	0.01
Estrogen receptor (positive vs negative)	1.14	0.70	1.87	0.59
Progesterone receptor (positive vs negative)	0.95	0.66	1.35	0.77
HER2 (positive vs negative)	1.62	1.11	2.37	0.01
PAM50 subtype				
HER2-enriched vs basal-like	1.03	0.63	1.67	0.91
LuminalA vs basal-like	0.32	0.17	0.61	0.001
LuminalB vs basal-lie	0.80	0.46	1.42	0.45
Normal-like vs basal-like	0.68	0.36	1.30	0.25

Table S27. Clinical properties of the ovarian cancer patients used in the analysis.

Dataset	Reference	Samples	Death event /censor	Meidan OS (months)	Relapse event /censor	Meidan RFS (months)	Stage (1/2/3/4)	Grade (1/2/3)	Serous samples	Debulk optimal /suboptimal
GSE14764	Denkert et al.2009	80	21/59	35	58/22	16.5	8/1/69/25	24/18/217	73	27/2
GSE26712	Bonome et al.2008	195	129/56	38.8	153/32	38.8	0/0/146/36	all high grade	185	90/95
GSE9891	Tothill et al.2008	285	110/175	28.5	193/90	15	24/18/218/22	20/99/161	265	160/70
TCGA	TCGA 2011	512	277/228	30.6	286/219	13.7	15/24/386/81	5/62/431	512	333/118

Table S28. Univariate Cox proportional hazards analysis of FXR1 in early stage of ovarian cancer patients

Dataset	HR	95% C.I		P	Patient numbers
GSE9891	10.9	1.39	85.35	0.023	42
TCGA	8.98	2.13	37.78	0.0027	39
Combined	5.38	1.95	14.82	0.0011	81

Table S29. Clinical characteristics of TCGA HNSC dataset (n=497)

Age	61±12
Gender	
Male	285 (72%)
Female	112 (28%)
Stage	
I	24 (7%)
II	61 (18%)
III	65 (19%)
IV	193 (56%)
Grade	
well	43 (11%)
moderate	236 (62%)
poor	99 (26%)
Lymph node status	
Negative	105 (43%)
Positive	142 (57%)
Perineural invasion presence	
NO	142 (51%)
YES	136 (49%)
HPV status by ISH testing	
Negative	52 (86%)
Positive	8 (14%)
HPV status by p16 testing	
Negative	61 (76%)
Positive	19 (24%)
Radiation therapy	
NO	33 (47%)
YES	37 (53%)
Neoadjuvant treatment	
No	387 (97%)
Yes	9 (3%)
Death event	137 (28%)
Relapse event	46 (9%)

Table S30. Clinical characteristics of E-TABM-302 HNSC dataset (n=81)

Age	
>58	42 (52%)
<58	39 (48%)
Gender	
Male	76 (94%)
Female	5 (6%)
Stage	
II	7 (9%)
III	23 (28%)
IV	51 (63%)
Lymph node status	
Negative	17 (21%)
Positive	64 (79%)
Grade	
poor	19 (23%)
modertate	38 (47%)
well	24 (30%)
HPV status	
No	75 (93%)
NA	6 (7%)
Four-gene predictor	
High risk	32 (44%)
Low risk	41 (56%)
Metastasis status	
M	41 (51%)
NM	40 (49%)
Metastasis-free survival (years, mean±SD)	
M	1.4±0.95
NM	8.31±2.87

Table S31. Multivariate Cox analysis of overall survival in TCGA HNSC dataset

Covariates	HR	95% C.I		P
FXR1	1.56	1.01	2.40	0.04*
Age	1.03	1.00	1.06	0.02*
Gender (Male vs Female)	1.00	0.58	1.74	1.00
Stage (III,IV vs I,II)	1.85	0.75	4.58	0.18
Grade (Well vs Moderate-poor)	0.71	0.27	1.86	0.48
Lymph node (positive vs negative)	1.30	0.71	2.37	0.40
Smoking history				
Former vs Current	0.57	0.34	0.95	0.03*
Never vs Current	0.61	0.28	1.32	0.21

Table S32. Multivariate Cox analysis of metastasis-free survival in E-TABM-302 HNSC cohort (n=81)

	HR	95% C.I		P
FXR1	2.25	1.21	4.18	0.01*
Age	0.96	0.93	1.00	0.05*
Gender(male vs female)	1.21	0.28	5.16	0.80
Grade(moderate-poor vs well)	0.61	0.29	1.28	0.19
Stage (III vs II)	0.11	0.02	0.69	0.02*
Stage (IV vs II)	0.26	0.05	1.55	0.14
Lymph node (positive vs negative)	3.23	0.75	14.01	0.12

Table S33. Multivariable Cox analysis of OS in TCGA stage I lung ADC patients (n=255)

	HR	95% C.I.		P[a]
FXR1	2.46	1.25	4.86	0.009*
FXR2	0.85	0.43	1.68	0.64
FMR1	0.75	0.37	1.50	0.41

[a] Adjusted for age, gender and smoking history

Table S34. Multivariable Cox analysis of OS in TCGA HNSC dataset

	HR	95% C.I.		P
FXR1	1.56	1.01	2.40	0.04*
FMR1	0.96	0.55	1.68	0.88
FXR2	0.87	0.49	1.55	0.64

HR: adjusted for age, gender, stage, grade, lymph node status and smoking history.