Supporting Information

O'Donnell et al. 10.1073/pnas.1115433109

SI Methods

Western Blotting. Antibodies for immunoblotting were as follows: anti–c-Myc XP rabbit monoclonal antibody (D84C12; Cell Signaling), anti-TIF2 mouse monoclonal antibody (610985; BD Transduction Laboratories), and anti–α-tubulin mouse monoclonal antibody (clone DM1A; Calbiochem).

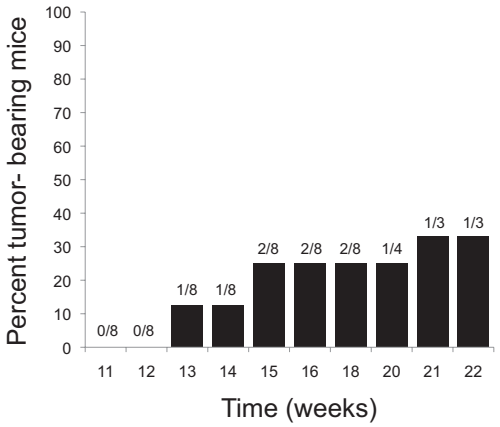

Fig. S1. Control experiment to determine tumor latency in the MYC liver tumor model on a mixed genetic background. Quantification of percentage of tet-*o-MYC*; LAPtTA animals expressing the tetracycline-repressible *MYC* transgene that developed liver tumors over time. Double-transgenic males were bred to C57BL/6J WT females to obtain tet-*o-MYC*; LAPtTA mice on a mixed background. *MYC* was induced at 6 wk of age, and three to eight mice were dissected each week from 11–22 wk of age.

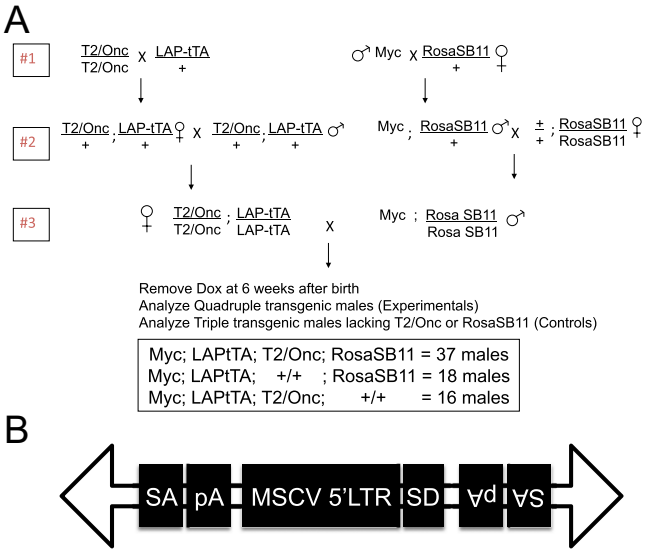

Fig. S2. Crossing scheme for *Sleeping Beauty* (*SB*) liver tumorigenesis screen and structure of the T2/Onc transgene. (*A*) Crossing scheme for generation of experimental and control cohorts. Individual transgenic mice were bred to obtain double-transgenic animals. These animals then were intercrossed to obtain homozygose transgenes. *T2/Onc; LAPtTA* females then were bred to *MYC; Rosa26-SB11* males to obtain quadruple transgenics (experimental) and triple-transgenic (control) mice that lacked the transposon or transposase. (*B*) The T2/Onc mutagenic transposon can alter gene function in two ways. In both the sense and antisense orientations, a splice acceptor (SA) is followed by a polyadenylation signal (pA). When the transposon inserts into a gene, the gene trap may be spliced to the transcript, and the pA signal will truncate the mRNA prematurely, disrupting expression of candidate tumor-suppressor genes. Additionally, the murine stem cell virus (MSCV) 5′ LTR followed by a splice donor (SD) is present in only one orientation. Transposon insertions that use the MSCV-5′ LTR/SD may therefore drive expression of candidate oncogenes.

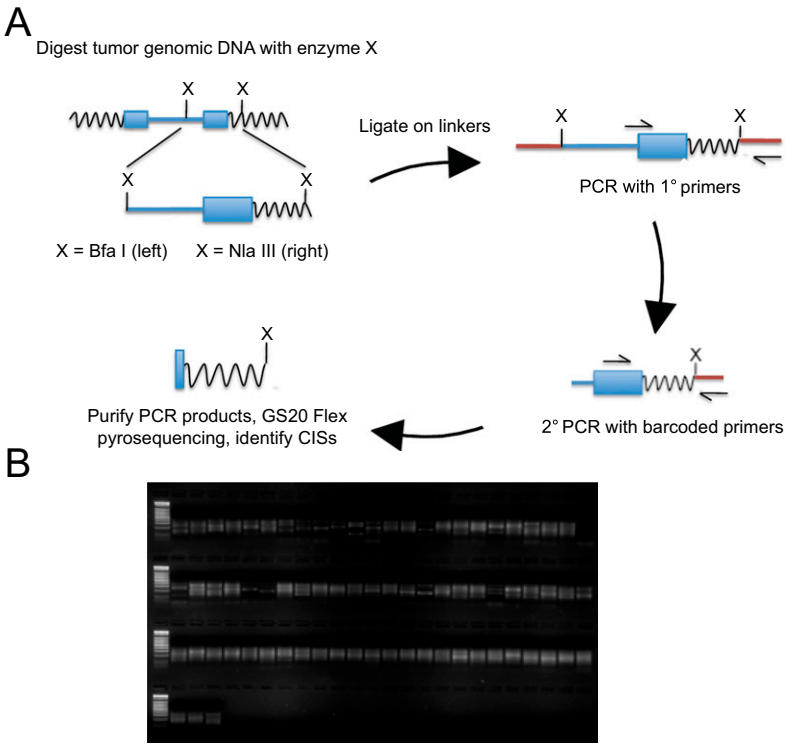

Fig. S3. Overview of ligation-mediated PCR and sequencing of transposon insertions. (*A*) Schematic of ligation-mediated PCR. Liver tumor genomic DNA (gDNA) from quadruple-transgenic animals was digested with *BfaI* and *NlaIII* enzymes. After ligation of linkers onto digested gDNA, two rounds of PCR were performed. In the second PCR, barcoded primers were used to allow pooling of samples. After purification of PCR products, the ligation-mediated PCR products were sequenced using GS20 Flex pyrosequencing. (*B*) Gel image of ligation-mediated PCR products. A smear of products ranging from 200–600 bp is observed.

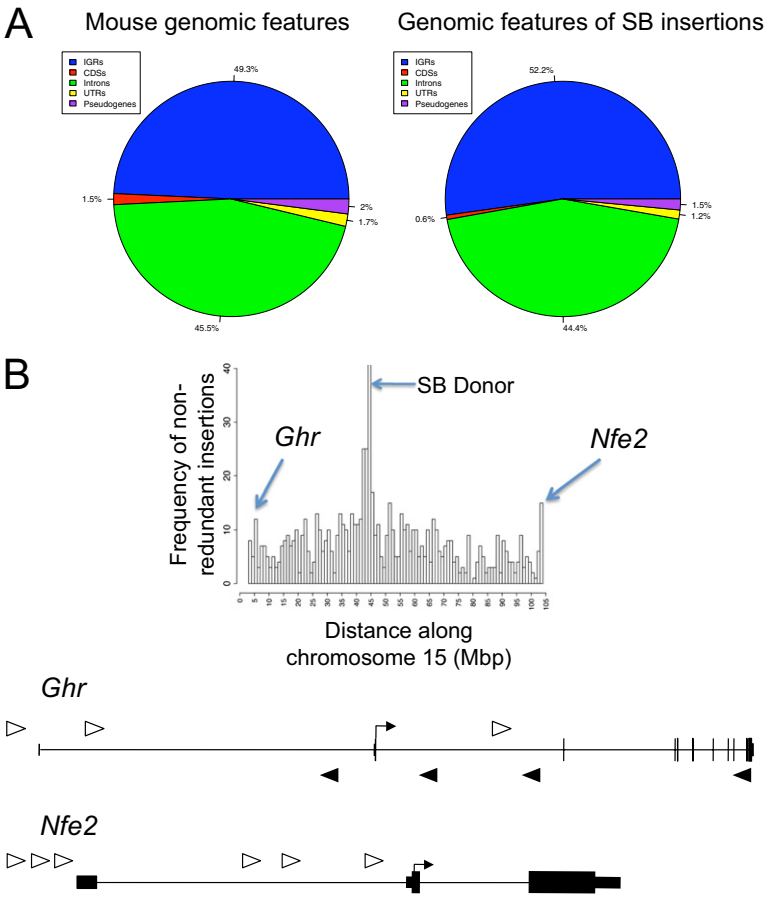

Fig. S4. Analysis of transposon insertions in the mouse genome. (*A*) (*Left*) Pie chart of the portion of the mouse genome occupied by different genomic features: intergenic regions (IGRs), CDS (coding regions), introns, UTRs, and pseudogenes. (*Right*) Pie chart of the genomic features that harbor the *SB* insertions inside the mouse genome. (*B*) (*Upper*) The frequency of nonredundant insertions on mouse chromosome 15. The peak at 45 Mbp corresponds to the position of the T2/Onc transgene array. Two additional peaks at 4 Mbps and 103 Mbps correspond to the *Ghr* and *Nfe2* genes, respectively. (*Lower*) Schematic representation of *SB* transposon insertions into *Ghr* and *Nfe2*. White arrowheads represent T2/Onc insertions in the sense orientation relative to target genes; black arrowheads represent antisense-oriented insertions. The height of protein-coding exons is taller than the untranslated sequences for each gene.

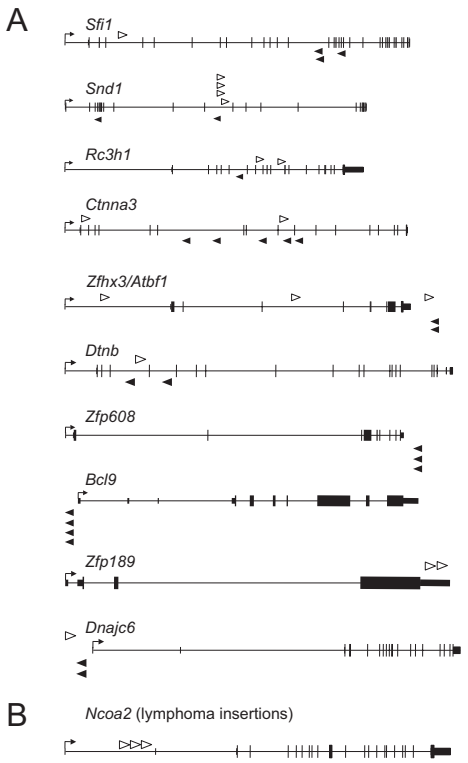

Fig. S5. *SB* insertions in candidate common-insertion sites (CIS) genes. (*A*) Schematic representation of transposon insertions in candidate liver CIS genes identified in the *SB* mutagenesis screen. White arrowheads represent T2/Onc insertions in the sense orientation relative to target genes; black arrowheads represent antisense-oriented insertions. (*B*) Transposon insertions in *Ncoa2* in three independent lymphoma samples.

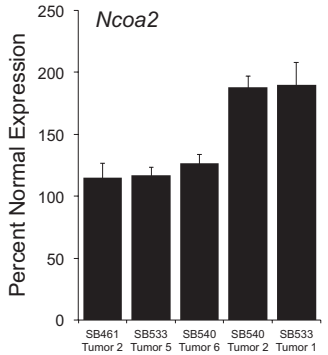

Fig. S6. Real-time PCR quantitation of *Ncoa2* mRNA levels in five liver tumors without SB insertions in *Ncoa2*. For each tumor, mRNA expression is normalized to *Ncoa2* expression of the surrounding normal liver. Bar graphs represent mean *Ncoa2* mRNA levels relative to 18S rRNA control. Error bars represent SDs from three independent measurements.

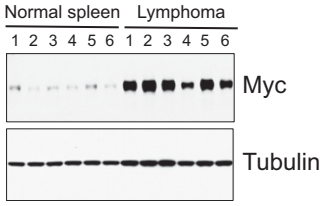

Fig. S7. Western blots documenting MYC protein levels in (*Left*) normal spleen from tet-*o-MYC*; LAPtTA mice (lanes 1–4) and tet-*o-MYC*; +/+ mice (lanes 5 and 6) and in (*Right*) lymphomas from tet-*o-MYC*; LAPtTA;T2/Onc; *Rosa26-SB11* mice (lanes 1–6). Tubulin served as a loading control.

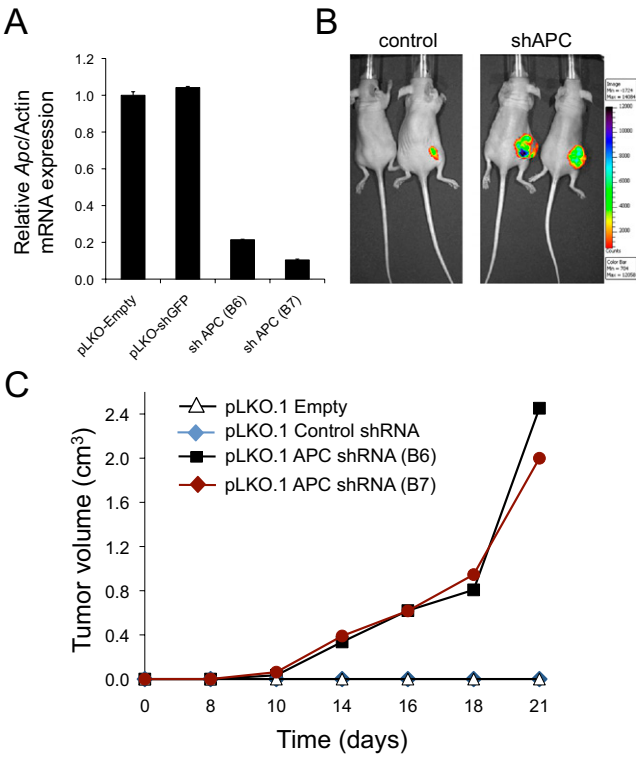

Fig. S8. Knockdown of *Apc* is a positive control in functional studies of CIS genes. (*A*) Real-time PCR quantitation of mRNA expression after knockdown of *Apc* in *Trp53*$^{-/-}$; *Myc* hepatoblasts. Expression was normalized to pLKO-Empty–infected cells. Bar graphs represent mean expression levels relative to Actin. Error bars represent SDs from three independent measurements. (*B*) Fluorescence imaging of nude mice injected with liver progenitor cells expressing shRNAs corresponding to *GFP* (*Left*) and *Apc* (*Right*). (*C*) Quantification of tumor volumes in nude mice injected with liver progenitor cells expressing *Apc* and control shRNAs. Line graphs depict mean tumor volumes from a total of four or five mice per shRNA tested. Independent tumorigenesis experiments yielded similar results. *Apc* was used as a positive control in every tumorigenesis assay performed.

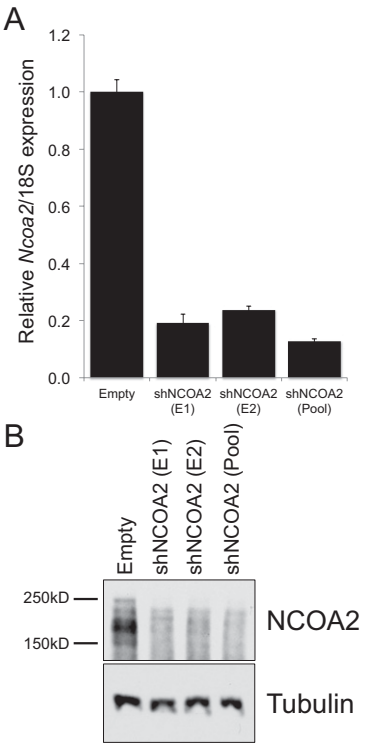

Fig. S9. Confirmation of *Ncoa2* knockdown in mouse hepatoblasts. (*A*) Real-time PCR quantitation of *Ncoa2* expression normalized to pLKO-Empty cells. Bar graphs represent mean expression relative to 18SrRNA. Error bars represent SDs from three independent measurements. (*B*) Western blots documenting knockdown of NCOA2 protein in hepatoblasts infected with *Ncoa2* shRNAs. Tubulin was used as a loading control.

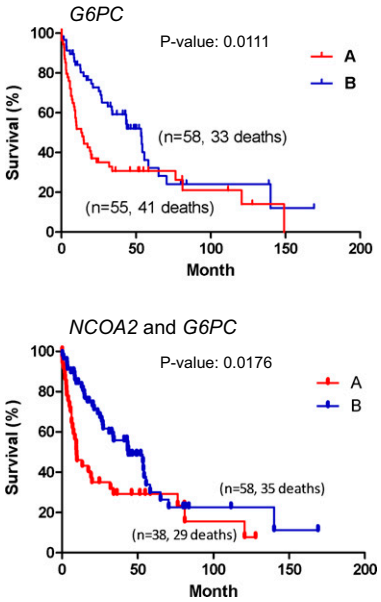

Fig. S10. Kaplan–Meier survival analysis of hepatocellular carcinoma (HCC) patients based on expression of *G6PC* alone and in combination with *NCOA2*. Overall survival of individuals with HCC clustered by low expression (cluster A) and high expression (cluster B) of *G6PC* alone (*Upper*) and *NCOA2* and *G6PC* (*Lower*). $P = 0.0111$ for *G6PC* and $P = 0.0176$ for NCOA2/*G6PC* based on a log-rank test.

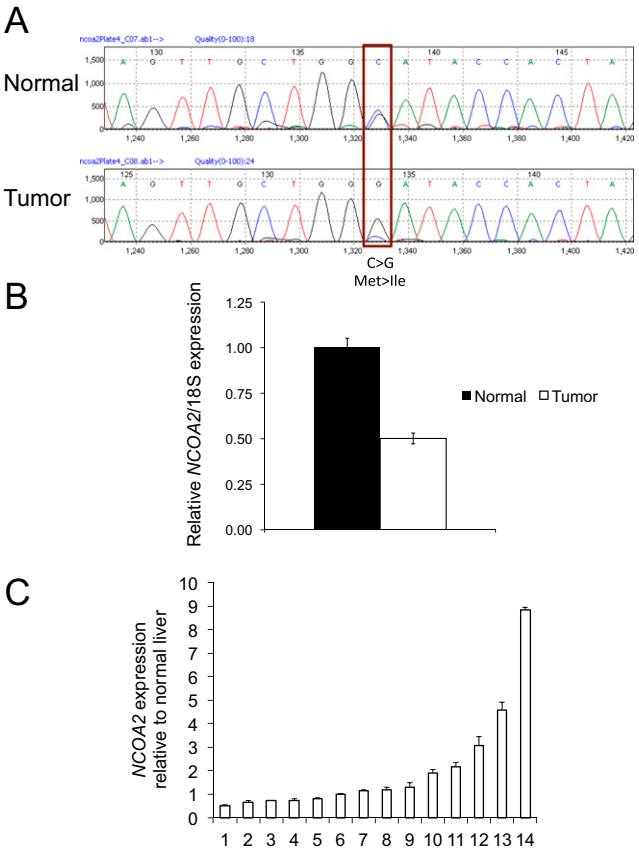

Fig. S11. Sequencing and expression of *NCOA2* in human HCC samples. (*A*) Chromatograph depicting variant that undergoes loss of heterozygosity in a human HCC sample. Loss of heterozygosity converts a previously heterozygous nonsynonymous substitution [Met > Ile at amino acid 1282, which occurs in a glutamine-rich region just upstream of the activation domain 2 (AD2) of the NCOA2 protein] to homozygosity. Upon further analysis of the HapMap and the 1000 Genomes databases, we discovered that this variant is a known SNP found at a frequency >0.017. (*B*) mRNA expression of *NCOA2* in a human HCC sample normalized to paired normal liver from the same patient. Bar graphs represent mean expression levels relative to 18SrRNA. Error bars represent SDs from three independent measurements. (*C*) *NCOA2* expression in individual HCC tumors relative to expression in paired normal liver samples.

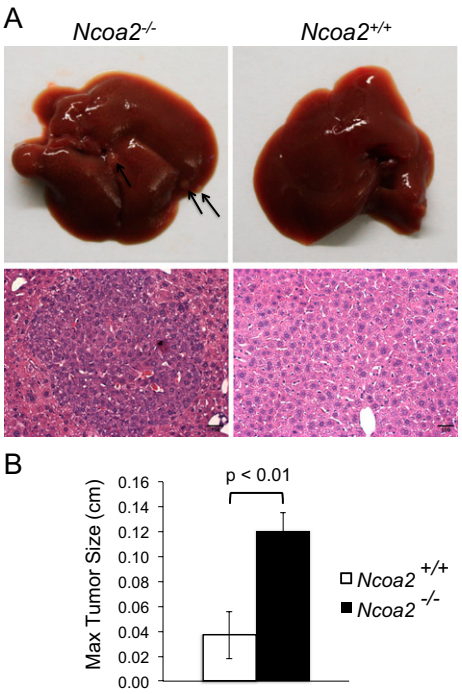

Fig. S12. Analysis of diethylnitrosamine-treated *Ncoa2*$^{+/+}$ and *Ncoa2*$^{-/-}$ mice. (*A*) (*Upper*) Representative images of diethylnitrosamine-treated livers from *Ncoa2*$^{+/+}$ and *Ncoa2*$^{-/-}$ animals. (*Lower*) Representative histology of a liver tumor nodule (*Left*) and normal liver (*Right*) from *Ncoa2*$^{-/-}$ and *Ncoa2*$^{+/+}$ animals using H&E staining. (*B*) Quantification of maximum tumor size 6 mo after diethylnitrosamine treatment in *Ncoa2*$^{+/+}$ and *Ncoa2*$^{-/-}$ knockout mice. Maximum tumor size, $P = 0.009$.

Table S1. Animals analyzed in *SB* mutagenesis screen

Mouse ID	Genotype	Sex	Tumors	No. of tumors dissected	Age at dissection (d)	Lymphoma
SB427	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	90	N
SB440	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	1	110	N
SB444	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	4	101	N
SB445	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	106	N
SB447	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	2	88	N
SB449	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	3	109	Y
SB453	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	1	109	Y
SB455	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	4	109	N
SB466	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	112	N
SB471	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	2	112	Y
SB489	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	6	111	Y
SB493	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	2	111	N
SB498	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	1	86	N
SB499	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	2	100	N
SB500	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	82	Y
SB501	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	86	N
SB503	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	2	100	Y
SB504	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	82	Y
SB506	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	1	104	N
SB518	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	111	N
SB521	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	1	104	N
SB533	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	9	106	N
SB534	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	1	106	N
SB537	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	105	N
SB540	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	6	99	Y
SB542	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	4	99	N
SB548	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	2	98	N
SB550	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	98	Y
SB576	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	100	N
SB577	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	2	100	N
SB578	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	1	100	N
SB579	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	4	101	N
SB580	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	1	103	Y
SB581	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	Y	1	103	N
SB584	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	103	N
SB586	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	104	N
SB588	MYC/+; LAP/+; T2Onc/+; Rosa/+	M	N	0	104	Y
SB465	MYC; LAP; T2Onc	M	N	0	100	N
SB485	MYC; LAP; T2Onc	M	N	0	99	N
SB487	MYC; LAP; T2Onc	M	N	0	104	N
SB490	MYC; LAP; T2Onc	M	N	0	99	N
SB491	MYC; LAP; T2Onc	M	N	0	99	N
SB508	MYC; LAP; T2Onc	M	N	0	104	N
SB509	MYC; LAP; T2Onc	M	Y	Multiple	104	N
SB510	MYC; LAP; T2Onc	M	N	0	104	N
SB516	MYC; LAP; T2Onc	M	N	0	103	N
SB527	MYC; LAP; T2Onc	M	N	0	104	N
SB536	MYC; LAP; T2Onc	M	Y	1	104	N
SB546	MYC; LAP; T2Onc	M	N	0	98	N
SB547	MYC; LAP; T2Onc	M	N	0	98	N
SB552	MYC; LAP; T2Onc	M	N	0	97	N
SB590	MYC; LAP; T2Onc	M	Y	4	98	N
SB591	MYC; LAP; T2Onc	M	Y	3	98	N
SB418	MYC; LAP; Rosa/+	M	Y	2	108	N
SB423	MYC; LAP; Rosa/+	M	N	0	110	N
SB431	MYC; LAP; Rosa/+	M	N	0	110	N
SB432	MYC; LAP; Rosa/+	M	N	0	110	N
SB433	MYC; LAP; Rosa/+	M	N	0	110	N
SB434	MYC; LAP; Rosa/+	M	N	0	110	N
SB450	MYC; LAP; Rosa/+	M	Y	Multiple	111	N
SB483	MYC; LAP; Rosa/+	M	N	0	92	N
SB520	MYC; LAP; Rosa/+	M	N	0	112	N
SB523	MYC; LAP; Rosa/+	M	Y	Multiple	111	N

Table S1. Cont.

Mouse ID	Genotype	Sex	Tumors	No. of tumors dissected	Age at dissection (d)	Lymphoma
SB524	MYC; LAP; Rosa/+	M	Y	2	111	N
SB530	MYC; LAP; Rosa/+	M	N	0	109	N
SB531	MYC; LAP; Rosa/+	M	Y	2	109	N
SB553	MYC; LAP; Rosa/+	M	Y	3	98	N
SB568	MYC; LAP; Rosa/+	M	N	0	101	N
SB573	MYC; LAP; Rosa/+	M	N	0	100	N
SB587	MYC; LAP; Rosa/+	M	N	0	98	N
SB593	MYC; LAP; Rosa/+	M	N	0	100	N

Table S2. Age of animals at the time of dissection

Cohort	No. of mice with tumors	Average age at day of dissection (d)
Quadruple-transgenic animals (63 tumors used for sequencing)	24	101.4
Triple-transgenic animals	10	103.6

Table S3. Common insertion sites in lymphomas

Chromosome	Gene name	Gene description	Range (bp)	No. of tumors	No. of independent insertion positions
10	*Myb*	Myb proto-oncogene protein (c-myb)	12,326	2	4
1	*Ncoa2*	Nuclear receptor coactivator 2 (SRC-2)	41,297	3	3
10	*5930403N24Rik*	Putative uncharacterized protein	79,301	1	3
10	*Specc1l*	Cytospin-A (SPECC1-like protein)	58,520	2	1
13	*Msh3*	DNA mismatch repair protein (Repair-3 protein)	96,001	1	3
16	*Erg*	Transcriptional regulator ERG	438	3	3
17	*Sos1*	Son of sevenless homolog 1	47,075	1	3
19	*Ppp1r3c*	Protein phosphatase 1 regulatory subunit 3C	68,923	2	2
2	*Ssfa2*	Sperm-specific antigen 2 homolog (K_i-Ras induced)	5	1	3
4	*Nbn*	Nibrin (Nijmegen breakage syndrome protein 1 homolog)	64,817	2	2
6	*Nfe2l3*	Nuclear factor erythroid 2-related factor 3	51,084	1	2

Table S4. Analysis of animals 6 mo after DEN treatment

Genotype	Body weight (g)	Liver weight (g)	No. of tumors	Maximum tumor size (cm)	Average tumor size (cm)	Spleen weight (g)	Liver weight (% body weight)
SRC-2$^{-/-}$	28.9	1.308	3	0.06	0.047	0.043	0.045
SRC-2$^{-/-}$	33.7	1.280	4	0.14	0.085	0.076	0.038
SRC-2$^{-/-}$	25.5	1.152	11	0.14	0.040	0.085	0.045
SRC-2$^{-/-}$	28.43	1.648	3	0.13	0.087	0.189	0.058
SRC-2$^{-/-}$	33.5	0.933	6	0.13	0.107	0.064	0.028
Avg.	30.006	1.264	5.4	0.12	0.073	0.092	0.043
SE	1.579	0.117	1.503	0.0152	0.013	0.025	0.005
SRC-2$^{+/+}$	34.5	1.664	0	0	0	0.146	0.048
SRC-2$^{+/+}$	35.6	1.690	0	0	0	0.167	0.047
SRC-2$^{+/+}$	30.6	1.296	2	0.08	0.055	0.107	0.042
SRC-2$^{+/+}$	32.5	1.473	1	0.02	0.020	0.097	0.045
SRC-2$^{+/+}$	35.91	1.768	4	0.085	0.061	0.175	0.049
Average	33.822	1.578	1.4	0.037	0.027	0.138	0.047
SE	1.003	0.086	0.748	0.019	0.013	0.016	0.001
Student's *t* test	0.076	0.062	0.044	0.009	0.037	0.155	0.491

Table S5. Analysis of CIS gene alterations in human tumors

Gene	Chromosome	Mutations	Nonsense substitution	Missense substitution	Deletion	Amplifications	Loss of heterozygosity
NCOA2	8	11 (2 breast, 6 lung, 1 ovary, 1 pancreas, 1 skin)	0	11	0	2	117 (1 liver)
ZFX	X	2 (2 ovary)	0	2	1	0	561 (6 liver)
SFI1	22	6 (1 CNS, 1 liver, 1 ovary, 3 UAD tract)	1	4	0	0	242 (3 liver)
SND1	7	4 (1 CNS, 2 ovary, 1 UAD tract)	0	1	0	7	136 (2 liver)
RC3H1	1	7 (2 ovary, 1 skin, 4 UAD tract)	1	4	0	4	118 (2 liver)
CTNNA3	10	10 (1 CNS, 1 lung, 2 ovary, 1 pancreas, 5 skin)	3	5	3	1	258 (2 liver)
ZFHX3	16	12 (5 ovary, 5 skin, 2 UAD tract)	2	7	1	2	205 (5 liver)
DTNB	2	5 (2 large intestine, 2 ovary, 1 UAD tract)	0	3	0	1	118 (1 liver)
ZNF189	9	5 (1 CNS, 1 kidney, 2 ovary, 1 skin)	0	4	0	1	237 (2 liver)
DNAJC6	1	3 (1 large intestine, 1 lung, 1 ovary)	1	2	0	0	138 (3 liver)
ZNF608	5	7 (1 lung, 3 ovary, 3 skin)	0	7	0	1	229 (1 liver)
BCL9	1	7 (1 breast, 2 large intestine, 2 ovary, 1 skin, 1 UAD tract)	1	3	0	5	77 (0 liver)
NFE2	12	2 (2 UAD tract)	1	1	0	1	118 (0 liver)
GHR	5	8 (1 kidney, 4 lung, 1 ovary, 1 skin, 1 UAD tract)	1	5	0	13	95 (1 liver)

Analysis of mutations in human tumors was performed using the COSMIC database (v57 release). UAD tract, upper aerodigestive tract.