Supplemental Information

SOX2 is an Amplified Lineage Survival Oncogene in Lung and Esophageal Squamous Cell Carcinomas

Adam J. Bass, Hideo Watanabe, Craig H. Mermel, Soyoung Yu, Sven Perner, Roel G. Verhaak, So Young Kim, Leslie Wardwell, Pablo Tamayo, Irit Gat-Viks, Alex H. Ramos, Michele S. Woo, Barbara A. Weir, Gad Getz, Rameen Beroukhim, Michael O'Kelly, Amit Dutt, Orit Rozenblatt-Rosen, Piotr Dziunycz, Justin Komisarof, Lucian R. Chirieac, Christopher J. LaFargue, Veit Scheble, Theresia Wilbertz, Changqing Ma, Shilpa Rao, Hiroshi Nakagawa, Douglas B. Stairs, Lin Lin, Thomas J. Giordano, Patrick Wagner, John D. Minna, Adi F. Gazdar, Chang Qi Zhu, Marcia S. Brose, Ivan Cecconello, Ulysses Ribeiro Jr., Suely K. Marie, Olav Dahl, Ramesh A. Shivdasani, Ming-Sound Tsao, Mark A. Rubin, Kwok K. Wong, Aviv Regev, William C. Hahn, David G. Beer, Anil K. Rustgi, Matthew Meyerson

Supplemental Note: Added details for RNAi experiments

After the RNAi screen described in the text of the manuscript was performed, we further evaluated the suppression of *DCUN1D1* in our screen, an effect that ranked 7th in essentiality of the 13 genes tested. We evaluated the knockdown efficacy of the 10 shRNA constructs targeting *DCUN1D1* and found that eight of these 10 shRNAs suppress DCUN1D1 protein expression (Supplemental Figure 3). These eight *DCUN1D1*-directed shRNAs show consistently less inhibition of proliferation in 3q26.33-amplified SCC cells than do the five (of six constructs screened) effective shRNA constructs targeting *SOX2* (Supplemental Figure 3). While these findings do not formally exclude the possibility of a role for *DCUN1D1* in SCC pathogenesis, they do not support the identification of *DCUN1D1* as a critical functional target of amplification in SCC.

With the later RNAi experiments using the stably infected and selected cells with shRNA constructs targeting *SOX2* or control GFP we observed both reduced proliferation and reduction in apparent anchorage-independent growth. We noted that the reduction in proliferation may contribute to this decreased anchorage-independent growth. To try to separate these phenotypes we performed parallel testing of NCI-H520 cells with *SOX2* RNAi with soft-agar and focus formation assays. The results were consistent with an effect of *SOX2* knockdown on anchorage-independent growth exceeding the effect on proliferation (Supplemental Figure 3). In addition, we also sought to evaluate if *SOX2* may be essential for cell lines with *SOX2* expression but not higher-level amplification as seen in the four 3q26.33-amplified cell lines used in the prior experiments. We found *SOX2*-targeting shRNA suppressed anchorage-independent growth in the Sox2-expressing esophageal SCC cell line TE11 that harbors 3q gain (Supplemental Figure 3), suggesting that *SOX2* is also essential for some tumors with lower-level genomic alteration at 3q26.33.

Supplemental Table 1:
GISTIC-determined peaks of amplification and deletion in lung and esophageal SCC

Lung SCC Amplification GISTIC Peaks

Cytoband	q value	Boundaries	SNP Boundaries	Genes in Peak	Candidates
3q26.33	1.53E-17	182472326-184234104	rs7653021-rs4859266	4	*SOX2*
8p12	8.19E-09	38252951-39417992	rs1488935-rs6415692	10	*FGFR1,WHSC1L1*
7p11.2	9.21E-05	54502271-55933841	rs10254916-rs2331096	6	*EGFR*

Lung SCC Deletion GISTIC Peaks

Cytoband	q value	Boundaries	SNP Boundaries	Genes in Peak	Candidates
9p21.3	2.74E-11	21944954-22057275	rs10757261-rs10965224	2	*CDKN2A/B*

Esophageal SCC Amplification GISTIC Peaks

Cytoband	q value	Boundaries	SNP Boundaries	Genes in Peak	Candidates
11q13.2	2.00E-43	68807653-69942337	rs3924679-rs12808511	10	*CCND1*
7p11.2	3.11E-06	54604765-55363608	rs4425711-rs6976824	2	*EGFR*
3q26.33	7.39E-05	182706356-183933466	rs4855031-rs2700847	1	*SOX2*
8q24.21	0.003	128351094-128699323	rs1011387-rs4562277	2	*POU5F1B, MYC*
8p12	0.003	38229373-38755603	rs2898674-rs55669404	6	*FGFR1, WHSC1L1, PPADC1B*

Esophageal SCC Deletion GISTIC Peaks

Cytoband	q value	Boundaries	SNP Boundaries	Genes in Peak	Candidates
9p21.3	1.48E-52	21944954-22031997	rs10757261-rs17694493	2	*CDKN2A/B*
Xp22.33	2.18E-10	2618641-2687133	rs1700955-rs66657603	2	
2q22.1	3.81E-06	141373426-141738261	rs12691581-rs10928100	1	*LRP1B*
18q22.1	6.16E-05	61759478-69251643	rs2587428-rs1943817	10	
3p14.2	0.0001	60352583-60385066	rs736934-rs2594252	1	*FHIT*
2q22.1	0.0004	142020039-142191813	rs7574682-rs10199508	1	*LRP1B*

Supplemental Table 2: Samples and Reagents

Samples Used for SNP Array Studies

Sample	Grade	Stage	Site	Type	SOX2 CN
1330	mod	IA	Lung SCC	Primary	2.32
1353	poor	IIB	Lung SCC	Primary	2.15
1669	mod	IIB	Lung SCC	Primary	2.04
1673	mod	IB	Lung SCC	Primary	1.99
1707	poor	1B	Lung SCC	Primary	3.49
1816	mod	IIA	Lung SCC	Primary	2.03
2032	poor	IB	Lung SCC	Primary	4.69
2132	poor	IIB	Lung SCC	Primary	3.46
2225	poor	IIIA	Lung SCC	Primary	2.22
2345	mod	IIB	Lung SCC	Primary	2.45
2523	mod	IIB	Lung SCC	Primary	2.05
2536	poor	IB	Lung SCC	Primary	2.93
2631	mod	IIB	Lung SCC	Primary	4.38
2727	mod/poor	IIB	Lung SCC	Primary	3.1
2808	mod/poor	IIB	Lung SCC	Primary	2.38
3076	poor	IA	Lung SCC	Primary	2.47
3121	mod/well	IIIB	Lung SCC	Primary	6.73
3253	poor	IB	Lung SCC	Primary	6.78
3289	mod	IB	Lung SCC	Primary	1.98
3312	mod	1A	Lung SCC	Primary	2.19
12T	mod	IB	Lung SCC	Primary	2.05
13T	mod/poor	IB	Lung SCC	Primary	3.71
15T	mod	IIB	Lung SCC	Primary	5.28
1735P	poor	III	Esophageal SCC	Primary	1.97
1737P	mod	IVa	Esophageal SCC	Primary	2.32
1739P	poor	III	Esophageal SCC	Primary	2.62
1747P	poor	III	Esophageal SCC	Primary	1.94
1749P	mod	IVa	Esophageal SCC	Primary	2.53
1759P	well	III	Esophageal SCC	Primary	2.65
1761P	well	III	Esophageal SCC	Primary	2.07
1777P	mod	III	Esophageal SCC	Primary	2.4
17T	mod	IIA	Lung SCC	Primary	2.81
1813P	mod	II	Esophageal SCC	Primary	2.65
1817P	mod	II	Esophageal SCC	Primary	2.35
1819P	poor	II	Esophageal SCC	Primary	2.63
19T	mod	IB	Lung SCC	Primary	2.24
21T	mod	IB	Lung SCC	Primary	2.33
23T	mod	IIB	Lung SCC	Primary	3.1
24T	mod	IIB	Lung SCC	Primary	2.09
2T	well	IB	Lung SCC	Primary	2.76
30T	mod	IB	Lung SCC	Primary	2.14
32T	mod	IIB	Lung SCC	Primary	2.93
33T	poor	IIB	Lung SCC	Primary	2.01
38T	mod	IB	Lung SCC	Primary	3.14
3T	mod	IB	Lung SCC	Primary	3.05
40T	mod	IIB	Lung SCC	Primary	3.99
43T	mod	IIB	Lung SCC	Primary	4.06

47T	mod	IB	Lung SCC	Primary	2
51T	poor	IIA	Lung SCC	Primary	4.38
52T	mod	IB	Lung SCC	Primary	3.69
54T	mod	IB	Lung SCC	Primary	2.23
57T	mod	IIB	Lung SCC	Primary	2.34
59T	poor	IB	Lung SCC	Primary	2.81
60T	poor	IB	Lung SCC	Primary	4.84
69T	mod	IIB	Lung SCC	Primary	3.43
71T	mod	IB	Lung SCC	Primary	2.98
72T	poor	IIA	Lung SCC	Primary	2.36
8T	poor	IIA	Lung SCC	Primary	2.28
HCE7	N/A	N/A	Esophageal SCC	Cell Line	4.28
KW1	mod	III	Esophageal SCC	Primary	2
KW10	poor	III	Esophageal SCC	Primary	2.87
KW11	mod	I	Esophageal SCC	Primary	2.39
KW12	poor	III	Esophageal SCC	Primary	2.76
KW13	well	IV	Esophageal SCC	Primary	2.99
KW14	poor	III	Esophageal SCC	Primary	3.2
KW15	mod	III	Esophageal SCC	Primary	2.66
KW16	mod	IIA	Esophageal SCC	Primary	2.39
KW17	poor	IIB	Esophageal SCC	Primary	2.76
KW18	other	III	Esophageal SCC	Primary	3.3
KW2	mod	IV	Esophageal SCC	Primary	6.11
KW3	mod	IIB	Esophageal SCC	Primary	2.84
KW4	mod	IV	Esophageal SCC	Primary	2.29
KW5	Poor	IV	Esophageal SCC	Primary	1.91
KW6	mod	I	Esophageal SCC	Primary	2.65
KW7	mod	III	Esophageal SCC	Primary	2.69
KW8	mod	III	Esophageal SCC	Primary	2.11
KW9	mod	III	Esophageal SCC	Primary	2.33
TE1	well	II	Esophageal SCC	Cell Line	2.26
TE10	well	IV	Esophageal SCC	Cell Line	7.02
TE11	mod	IV	Esophageal SCC	Cell Line	2.54
TE15	well	IV	Esophageal SCC	Cell Line	2.35
TE4	well	III	Esophageal SCC	Cell Line	3.91
TE5	poor	IV	Esophageal SCC	Cell Line	3.42
TE6	well	IV	Esophageal SCC	Cell Line	10.59
TE8	mod	III	Esophageal SCC	Cell Line	3.26
TE9	poor	IV	Esophageal SCC	Cell Line	2.38
TT	N/A	IV	Esophageal SCC	Cell Line	3.61

shRNA Vectors Used in RNAi Screen

Gene Symbol	**TRC clone name**
ATP11B	XM_087254.4-2579s1c1
ATP11B	XM_087254.4-1435s1c1
ATP11B	XM_087254.4-2284s1c1
ATP11B	XM_087254.4-526s1c1
ATP11B	XM_087254.4-1555s1c1
ATP11B	NM_029570.3-1969s1c1
ATP11B	NM_029570.3-341s1c1
B3GNT5	NM_032047.3-1286s1c1
B3GNT5	NM_032047.3-728s1c1
B3GNT5	NM_032047.3-1511s1c1

B3GNT5	NM_032047.3-736s1c1
B3GNT5	NM_032047.3-1220s1c1
DCUN1D1	NM_020640.2-467s1c1
DCUN1D1	NM_020640.2-397s1c1
DCUN1D1	NM_020640.2-235s1c1
DCUN1D1	NM_020640.2-588s1c1
DCUN1D1	NM_020640.2-757s1c1
DCUN1D1	NM_020640.2-1521s1c1
DCUN1D1	NM_020640.2-355s1c1
DCUN1D1	NM_020640.2-2035s1c1
DCUN1D1	NM_020640.2-1285s1c1
DCUN1D1	NM_033623.2-111s1c1
DNAJC19	NM_201260.1-730s1c1
DNAJC19	NM_201260.1-799s1c1
DNAJC19	NM_201260.1-616s1c1
DNAJC19	NM_201260.1-187s1c1
DNAJC19	NM_201260.1-907s1c1
FXR1	NM_005087.1-579s1c1
FXR1	NM_005087.1-130s1c1
FXR1	NM_005087.1-1028s1c1
KLHL6	NM_130446.1-1260s1c1
KLHL6	NM_130446.1-547s1c1
KLHL6	NM_130446.1-4632s1c1
KLHL6	NM_130446.1-709s1c1
KLHL6	NM_130446.1-2535s1c1
LAMP3	NM_014398.2-1041s1c1
LAMP3	NM_014398.2-304s1c1
LAMP3	NM_014398.2-842s1c1
LAMP3	NM_014398.2-1810s1c1
LAMP3	NM_014398.2-388s1c1
MCCC1	NM_020166.2-1958s1c1
MCCC1	NM_020166.2-1957s1c1
MCCC1	NM_020166.2-859s1c1
MCF2L2	NM_015078.2-466s1c1
MCF2L2	NM_015078.2-1248s1c1
MCF2L2	NM_015078.2-2760s1c1
MCF2L2	NM_015078.2-316s1c1
MCF2L2	NM_015078.2-2121s1c1
PIK3CA	NM_006218.2-1900s1c1
PIK3CA	NM_006218.2-3471s1c1
PIK3CA	NM_006218.2-2892s1c1
SOX2	NM_003106.x-787s1c1
SOX2	NM_003106.x-2378s1c1
SOX2	NM_003106.x-1060s1c1
SOX2	NM_011443.2-1665s1c1
SOX2	NM_011443.2-1027s1c1
SOX2	NM_011443.2-850s1c1
TP63	NM_003722.3-2716s1c1
TP63	NM_003722.3-870s1c1
TP63	NM_003722.3-2058s1c1
TP63	NM_003722.3-164s1c1
TP63	NM_003722.3-1218s1c1
TTC14	NM_133462.1-1838s1c1
TTC14	NM_133462.1-2468s1c1

TTC14	NM_133462.1-393s1c1
TTC14	NM_133462.1-2482s1c1
TTC14	NM_133462.1-2278s1c1

shRNA Vectors Used for Follow-up Experiments

shRNA	**TRC Clone ID**	**target**
shSOX2a	NM_011443.2-1665s1c1	CGAGATAAACATGGCAATCAA
shSOX2b	NM_003106.x-2378s1c1	CTGCCGAGAATCCATGTATAT
shGFP		ACAACAGCCACAACGTCTATA

Primers Used for Real-Time PCR

Gene	**Primer 1**	**Primer 2**
GAPDH	CATGAGAAGTATGACAACAGCCT	AGTCCTTCCACGATACCAAAGT
KRT6A	GGCTGAGGAGCGTGAACAG	CCAGGAACCGCACCTTGT
TP63	CCTTACTTTGCTGAGGGTTTGAA	CAAGGCCCTAGTGTTACCTGAATAG
SOX2	TACAGCATGTCCTACTCGCAG	GAGGAAGAGGTAACCACAGGG

Supplemental Table 3:
Genes correlated with *SOX2* expression in lung squamous cell carcinoma

Rank	Gene Name	Score
1	*SOX2*	31.7851
2	*FOXE1*	11.3745
3	*CHST7*	8.5647
4	*ARTN*	8.3056
5	*PRKX*	8.109
6	*GPX2*	7.593
7	*CSRP2*	7.5534
8	*SUSD4*	7.4062
9	*BLMH*	7.2036
10	*ADH7*	7.1618
11	*FGFR2*	6.9599
12	*RAB40B*	6.6939
13	*KLF5*	6.6201
14	*ADAM23*	6.6189
15	*ABCC5*	6.5715
16	*RRM1*	6.5583
17	*TMPRSS4*	6.4615
18	*ATP1B3*	6.3264
19	*RMND5A*	6.2492
20	*SIAH2*	6.1775
21	*PRKY*	6.1614
22	*DHX9*	6.1583
23	*RAB6B*	6.067
24	*PSMD11*	5.9397
25	*MSH6*	5.8637
26	*CST1*	5.8374
27	*CNTN1*	5.7848
28	*GDAP1*	5.7771
29	*BMP7*	5.7262
30	*CHAC1*	5.7015
31	*PLA2G4B*	5.6961
32	*BDH1*	5.6752
33	*AKR1C3*	5.606
34	*TBL1X*	5.5977
35	*MARK1*	5.5919
36	*ABCC1*	5.5832
37	*LRP8*	5.5674
38	*TP63*	5.5629
39	*ENTPD3*	5.4769
40	*DST*	5.472
41	*AKR1C4*	5.4702
42	*GPC3*	5.4598
43	*LSG1*	5.4367
44	*C3orf58*	5.4189
45	*NEO1*	5.4185
46	*KRT6A*	5.3818
47	*CYB5R2*	5.3704
48	*FOXRED2*	5.3328
49	*UGT1A9*	5.322
50	*USP39*	5.2764

Expression profiling data from 92 stage I/II tumors were analyzed to identify the 10 tumors with highest and lowest *SOX2* expression. The top 50 genes whose expression distinguishes these two groups are listed in descending rank order.

Supplemental Figure Legends

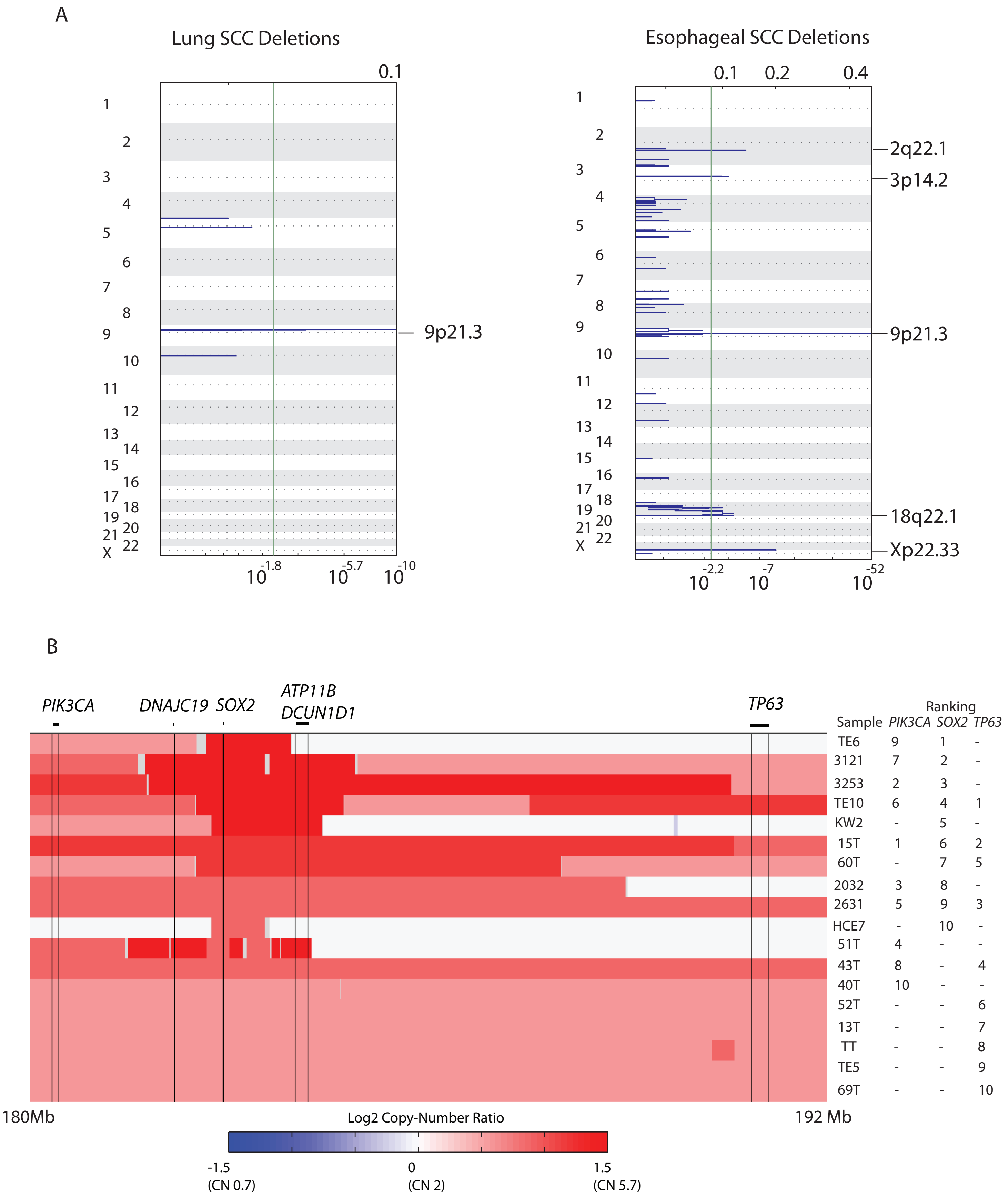
Supplemental Figure 1: Additional Data from SNP Array Analysis A) Plots of GISTIC scores for recurrent high-level deletions in 47 primary SCCs of the lung (left) and 40 esophageal SCC samples (29 primary samples and 11 cell lines; right). X-axis shows the G-score (top) and false discovery rate (q-value; bottom) for recurrent deletions with a green line demarcating FDR cut-off of 0.005. Labels on right denote the position of the peak of the significantly altered region. B) Copy number at 3q26-28 loci in samples with highest copy numbers in each of *SOX2, PIK3CA* or *TP63*. Copy number is depicted across a region of 3q spanning 180-192 Mb. The copy number is charted as log2 CN ratio of the tumor to normal (see color bar). Positions of *SOX2*, *PIK3CA*, *TP63* and *SOX2*'s immediate neighboring genes *DNAJC19* and (*ATP11B*/*DCUN1D1)* are marked. The 18 samples comprising the group of samples with the ten highest copy numbers at the *SOX2*, *TP63* or *PIK3CA* loci are shown and are depicted in decreasing order of copy number at *SOX2*.

Supplemental Figure 2: Correlations of *SOX2* Amplifications and Expression A) SNP array determined copy-number at *SOX2* and *SOX2* mRNA expression are plotted for 27 lung SCC samples. B) Comparison of *SOX2* mRNA expression in those samples meeting or not-meeting the threshold for amplification via SNP arrays (3.6 inferred copies) showing significant increase in amplified cases. C) Measurement of Sox2 protein via immunohistochemistry in tissue microarray normal esophagus, and esophageal SCC cases. Cores representing samples of normal esophagus and cases with low and high Sox2 expression are shown. The chart shows Sox2 immunohistochemistry staining in cases with *SOX2* amplification (>10 copies per FISH) compared to those with inferred diploid copy at *SOX2*.

Supplemental Figure 3: Additional Data from RNAi Experiments A) Immunoblots demonstrate knockdown efficiency of each of the 10 shRNA constructs against *DCUN1D1* compared to control (shGFP) in HCC95 lung SCC cells. B) Graphed are the effects of the eight effective *DCUN1D1* shRNA constructs and five *SOX2* constructs with effective knockdown in the RNAi screen. For each cell line, the impact on viability of each shRNA construct relative to control (shGFP and shLacZ) is plotted horizontally with shRNA constructs ranked by efficacy. Lower ratios are indicative of reduced viability. C) Comparative effects of shSOX2b on foci formation and anchorage-independent growth in NCI-H520 cells. The numbers of soft-agar colonies and foci formed for the shSOX2b relative to shGFP are plotted (+/- standard deviation) with p-values labeled with asterisks as in Figure 2 in the main text. D) Immunoblots showing expression of Sox2 and Vinculin in TE11 cells relative to NIH-H520, HCC95 and NCI-H1437. The chart shows anchorage-independent growth of TE11 expressing shRNAs targeting *SOX2* or *GFP* (+/- standard deviation) with p-values labeled with asterisks as in Figure 2.

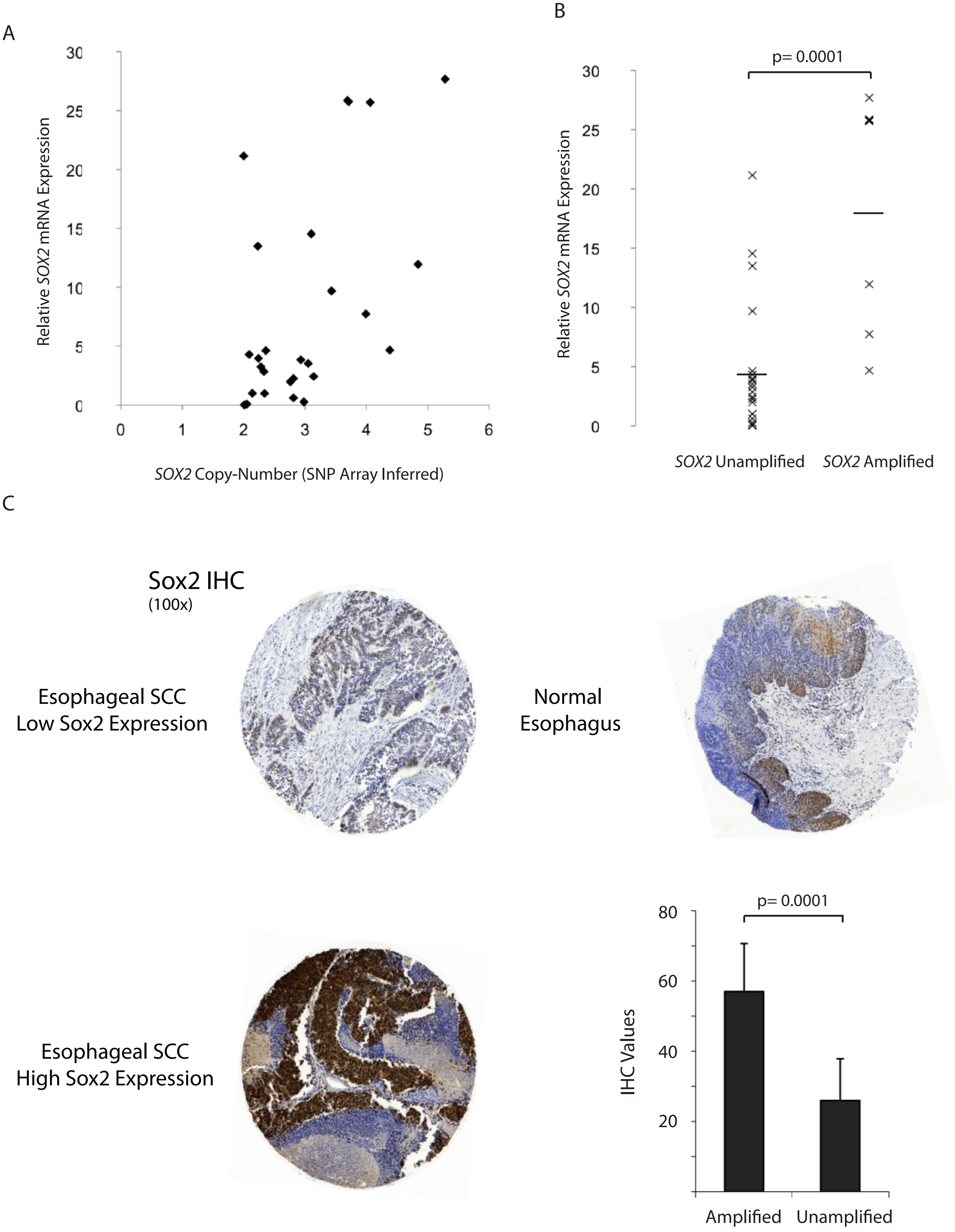
Supplemental Figure 4: Amplification and Expression of *SOX2* and *NKX2-1* Differ Between SCCs and Adenocarcinoma of the Lung. A) Segmented copy number on chromosome 3q with position of *SOX2* labeled. SCCs of the lung (47) are depicted (columns) in the left box and compared to an equal sized set of lung adenocarcinomas. For equivalent analysis, the adenocarcinoma samples pictured were randomly selected from a larger sample collection. The color scale ranges from blue (deletion) to white (neutral copy number) to red (amplification). On the right plot, G-scores for amplification along 3q are plotted for the two groups. B) Copy number data on chromosome 14q from 47 samples of lung squamous cell carcinomas and 47 randomly selected lung adenocarcinomas; G-scores for amplifications for the two groups are plotted on the right. The position of *NKX2-1* on 14q is labeled. C-D) Expression of mRNA for *SOX2* and *NKX2-1* in lung SCC (panel C; n=92) and lung adenocarcinoma (panel D; n=353). Box-plots show median expression and 25th/75th percentiles with outside lines showing limits of 95% of samples. Y-axis is log2 expression in arbitrary expression units. P-values for significance of differential expression for panels C and D are labeled.

Supplemental Figure 1

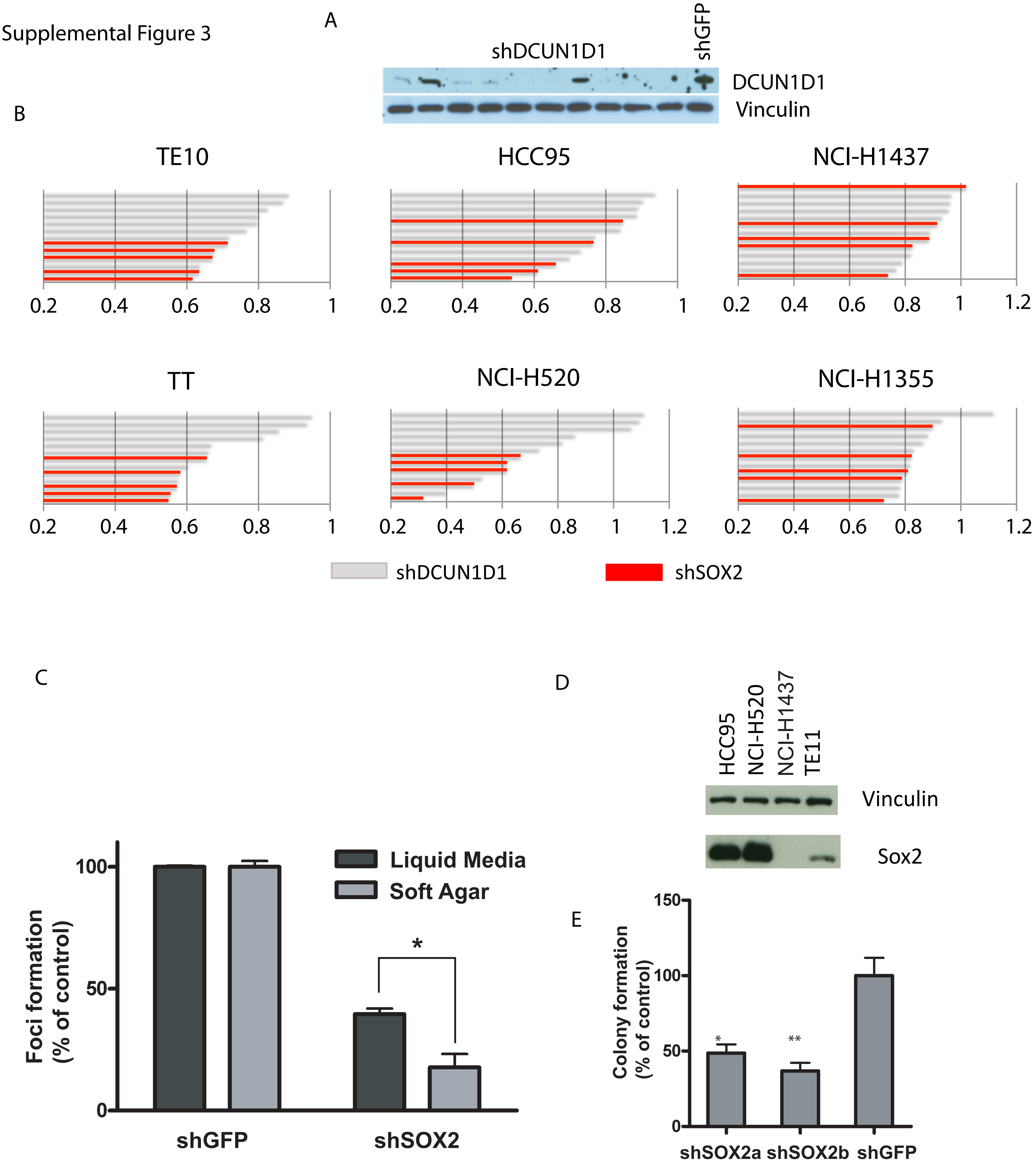


	Ranking		
Sample	*PIK3CA*	*SOX2*	*TP63*
TE6	9	1	-
3121	7	2	-
3253	2	3	-
TE10	6	4	1
KW2	-	5	-
15T	1	6	2
60T	-	7	5
2032	3	8	-
2631	5	9	3
HCE7	-	10	-
51T	4	-	-
43T	8	-	4
40T	10	-	-
52T	-	-	6
13T	-	-	7
TT	-	-	8
TE5	-	-	9
69T	-	-	10

Supplemental Figure 2



Supplemental Figure 3



Supplemental Figure 4