Supplementary Table 1

Cell Line	NTHL1 Expression	NSCLC Subtype	Origin	*TP53* Status	*EGFR* Status	*Ras* Status	Additional Information	Reference(s)
Lung								
HBEC	elevated	-	normal bronchial epithelium	wildtype	wildtype	wildtype	hTERT and CDK4 immortalized	ATCC; (1)
Beas2B	low	-	normal bronchial epithelium	wildtype	wildtype	wildtype	cells form colonies in semi-solid matrix, but did not for tumors in immunosuppressed mice	ATCC; (2)
NSCLC								
A549	elevated	carcinoma	primary tumor explant	wildtype	wildtype	mutant		ATCC; (2)
H460	elevated	large cell carcinoma	lung plural effusion	wildtype	wildtype	mutant		ATCC;(2); (3)
H1299	undetected	carcinoma	metastatic lymph node	null	wildtype	wildtype		ATCC; (2)
H1975	elevated	adeno-carcinoma	primary tumor	mutant	mutant	wildtype	non-smoker	ATCC; (2); (4); (3)
H226	low	squamous	metastatic plural effusion	wildtype	wildtype	wildtype		ATCC;(2); (5); (4)
H522	elevated	adeno-carcinoma	metastatic plural effusion	mutant	wildtype	wildtype	stage 2; non-smoker	ATCC; (2); (6); (3)
HCC827	elevated	adeno-carcinoma	primary tumor	mutant	mutant	wildtype		ATCC; (2); (3)
H1792	elevated	adeno-carcinoma	metastatic plural effusion	mutant	wildtype	mutant	stage 4; smoker	ATCC; (2)
Caul1	elevated	epidermoid carcinoma	metastatic plural effusion	null	wildtype	mutant	grade III	ATCC; (2); (3)

Supplementary Table 2

Cell Line	NTHL1 Expression	Origin	*TP53* Status	*Ras* Status	Additional Information	Reference(s)
Caco2	low	colon	mutated	wildtype	heterozygous *TP53* mutation; grade II	ATCC; (7); (8)
HT29	elevated	colon	mutated	wildtype	grade II	ATCC
HCT116	elevated	colon	wild type	mutated	Mismatch repair deficient (MLH1)	ATCC; (9); (10)
SW480	low	colon	mutated	mutated	Dukes' type B, colorectal adenocarcinoma	ATCC
HeLa	elevated	cervix	wild type	wildtype	E6 viral protein results in null p53 expression	ATCC; (11); (12)

Supplementary Table 3

Data Set	Reporting Group	Provisional Data Available	NSCLC Subtype	Information Type	Publication Year	Reference
Pan-Lung cancer	TCGA	No	Pan	CNV	2016	(13)
Lung Adenocarcinoma	Broad Institute	No	Adenocarcinoma	CNV	2012	(14)
Lung Adenocarcinoma	TCGA	Yes	Adenocarcinoma	CNV and RNA	2014	(15)
Lung Squamous Cell Carcinoma	TCGA	Yes	Squamous	CNV and RNA	2012	(16)

Supplementary Table 4

Project	Primer Direction	Primer Sequence	Original Plasmid	Plasmid Cloned Into	Reference/ Source
NTHL1-FLAG Subcloning	Forward	5'-ACACTGGCGGCCGTTACTAGTGGATCCT-3'	pCMV6-AC-NTHL1-GFP	pcDNA3.1(+)	Origene (PS100010) Invitrogen (V79020)
	Reverse	5'-ACGACTCACTATAGGGAGACCCAAGCTT-3'			
CATmut-FLAG Site-directed Mutagenesis	Forward	5'-TGTTGGGCCCCAGATGGCACACCTGG -3'	pcDNA3.1(+)	N/A	Invitrogen (V79020)
	Reverse	5'-CCCGGCAGCGCCACCAGC-3'			
qRT-PCR NTHL1	Forward	5'-CAGCATCCTGCAGACAGATGA-3'	N/A	N/A	Sigma
	Reverse	5'-GTCCACTGCAATGCCTGACAC-3'			

SUPPLEMENTARY DATA TABLE AND FIGURE LEGENDS

Supplementary Table 1: Non-small cell lung cancer (NSCLC) cell line characteristics. Information on each cell line used in this study including NSCLC subtype, origin, *TP53* gene status, *EGFR* and *RAS* gene mutational status. Information in this table was used in Figure 1B and Supplementary Figure 1A. Information was obtained from the American Type Culture Collection (ATCC), the COSMIC database (2), and referenced publications.

Supplementary Table 2: **Colon and HeLa cell line characteristics**. Information on each cell line including tissue origin, *TP53* status, and *RAS* mutational status. Information was obtained from the American Type Culture Collection (ATCC) and the referenced publications.

Supplementary Table 3: Information on cBioPortal NSCLC datasets used in this study. Datasets were accessed through cBioPortal (17,18) and mined in Oncoprint for *NTHL1* copy number variation (CNV) or RNA data. Percentages of NTHL1 amplification or RNA upregulation from each study is reported as a summary in Figure 1A**.** Provisional data sets are preliminary data before validation.

Supplementary Table 4: Cloning and oligonucleotide used in experiments. Plasmids used for *NTHL1* cloning and the corresponding oligonucleotides used are displayed with their sources. Oligonucleotides for qRT-PCR of *NTHL1* in control and non-small cell lung cancer are shown.

Supplementary Figure 1: NTHL1 protein levels vary between non-small cell lung cancer cell lines (NSCLC). A) Immunoblot of a panel of NSCLC cell lines reveals that NTHL1 is elevated in NSCLC compared to non-transformed HBEC cells. Gels for lung cancer screens were run with a total protein content of 25 µg of protein. B) Immunoblots for NTHL1 expression from the NSCLC cell line panel were quantified and demonstrate that NTHL1 expression is significantly increased. C) Immunoblots for NEIL2 expression from the NSCLC cell line panel were quantified and demonstrate that NEIL2 expression does not widely vary. NEIL2 expression is not statistically significant in comparison between the HBEC control and the NSCLC cell lines. D) Immunoblotting for NTHL1 expression in HeLa cells and colon cancer cell lines demonstrates that NTHL1 expression is elevated in multiple cancer types. E) Quantification of immunoblots from the HeLa and colon cancer panel display a significant increase in NTHL1 expression compared to the reference control cell line Caco-2. P values are as follows: $p \leq 0.05$ * ; $p \leq 0.01$ ** ; $p \leq 0.001$ ***.

Supplementary Figure 2: DNA damage is induced by NTHL1 and CATmut overexpression. A) Dot plot of DNA damage (% DNA) in comet tails for 50 individual HBEC cells from three independent experimental replicates. Hydrogen peroxide (H_2O_2) treatment was used as a positive control for DNA damage. NTHL1 and CATmut expression results in a cell population that contains an increase in DNA damage compared to empty vector (Vector) control cells. B) Functional schematic of the DR-GFP

assay used to measure homologous recombination (HR) activity in U2OS cells. The DR-GFP cassette contains two inactive GFP genes separated by a DNA linker. One is disrupted by the DNA sequence recognized by the exogenously introduced I-Sce1 restriction enzyme (SceGFP). The other is an inactive GFP that serves as the homologous recombination template for repair (iGFP). Upon expression of the I-Sce1 enzyme, a DSB is generated at the I-Sce1 cleavage site. If HR is active, the DSB will be repaired using the iGFP cassette to generate a repaired SceGFP gene that results in fluorescently active GFP protein (GFP+). C) Flow cytometry on U2OS cells for GFP signal (FL-1) after an I-Sce1 induced DSB. GFP positive cells represent cells that are able to repair the I-Sce1 break via HR. NTHL1 and CATmut overexpressing cells display a decrease in the percentage of cells that repair the DSB by HR compared to empty vector (Vector) control cells. D) Immunoblot for NTHL1-Flag and I-Sce1-HA in cells that were used in the DR-GFP assay. E) Immunoblot of NTHL1-Flag (72 hours following expression) in cells used for the DsR-7F4 end joining assay. I-Sce1 is IFP fluorescently tagged and cells were normalized for I-Sce1 expression upon analysis. F) HBEC cells expressing GFP alone, or either NTHL1-GFP or CATmut-GFP were cell cycle sorted using GFP intensity and propidium iodine staining. Chi-square analysis of the cell cycle phase distribution demonstrates no statistical significance ($p=0.27$) among the samples.

Supplementary Figure 3: Genomic instability is induced by NTHL1 and CATmut overexpression. A) HBEC cells expressing NTHL1-GFP were FACS sorted for GFP signal for experiments in Figure 4A. Cells were sorted in three different populations based on GFP intensity termed low, intermediate (Int), and High. B) In a separate unrelated experiment, HBEC cells overexpressing NTHL1-GFP were scored for micronucleus formation experiments in Figure 4D. Genomic instability is specific to cells overexpressing NTHL1-GFP and not seen from GFP expression alone. C) Ectopic expression of red fluorescent protein (dsRED) was employed as a control for general protein overexpression in HBEC. D) Expression of the G12D mutant form of K-Ras (a well-known oncogene variant in NSCLC) was employed as a positive control for micronucleus assays. E) Representative examples of a binucleated cell without a micronucleus, typical binucleated cell with one micronucleus, and binucleated cells with multiple micronuclei, as scored for micronucleus assays with NTHL1 and CATmut overexpression.

Supplementary Figure 4: NTHL1 expression levels in soft agar clones and parental HBEC clones. A) Quantification of immunoblot of NTHL1 protein showing fold change of protein levels in soft agar clones compared to the parental HBEC line. B) Immunoblot of NTHL1 expression in the HBEC line, cells transiently overexpressing NTHL1, and ten different clonal derivatives of non-transfected HBECs. Immunoblots were run and developed at the same time, but were transferred onto different membranes in order to include all ten non-transfected HBEC clones.

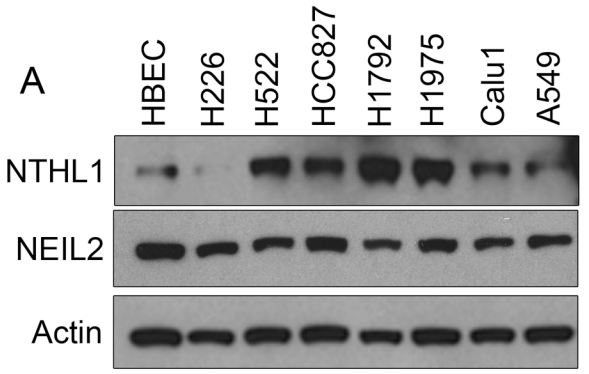
Supplementary References

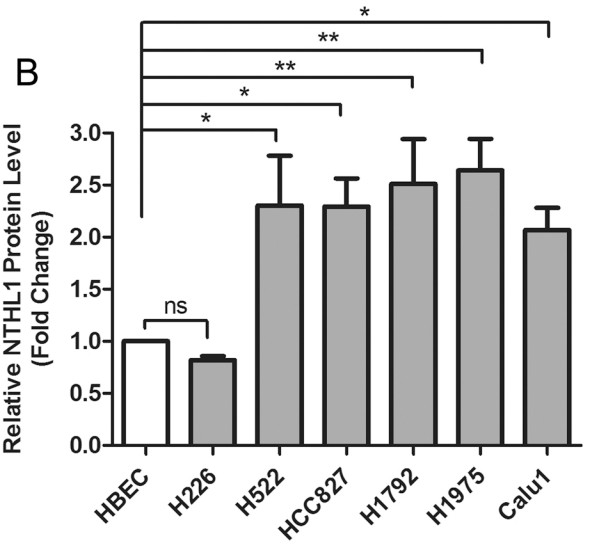
1. Sato, M., Larsen, J.E., Lee, W., Sun, H., Shames, D.S., Dalvi, M.P., Ramirez, R.D., Tang, H., DiMaio, J.M., Gao, B. *et al.* (2013) Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations. *Mol Cancer Res*, **11**, 638-650.
2. Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S. *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*, **43**, D805-811.
3. Giovannetti, E., Zucali, P.A., Peters, G.J., Cortesi, F., D'Incecco, A., Smit, E.F., Falcone, A., Burgers, J.A., Santoro, A., Danesi, R. *et al.* (2010) Association of polymorphisms in AKT1 and EGFR with clinical outcome and toxicity in non-small cell lung cancer patients treated with gefitinib. *Mol Cancer Ther*, **9**, 581-593.
4. Franklin, W.A., Haney, J., Sugita, M., Bemis, L., Jimeno, A. and Messersmith, W.A. (2010) KRAS mutation: comparison of testing methods and tissue sampling techniques in colon cancer. *J Mol Diagn*, **12**, 43-50.
5. Zhang, H., Zhan, C., Ke, J., Xue, Z., Zhang, A., Xu, K., Shen, Z., Yu, L. and Chen, L. (2016) EGFR kinase domain mutation positive lung cancers are sensitive to intrapleural perfusion with hyperthermic chemotherapy (IPHC) complete treatment. *Oncotarget*, **7**, 3367-3378.
6. Yang, X., Zhao, M., Xia, M., Liu, Y., Yan, J., Ji, H. and Wang, G. (2012) Selective requirement for Mediator MED23 in Ras-active lung cancer. *Proc Natl Acad Sci U S A*, **109**, E2813-2822.
7. Brink, M., de Goeij, A.F., Weijenberg, M.P., Roemen, G.M., Lentjes, M.H., Pachen, M.M., Smits, K.M., de Bruine, A.P., Goldbohm, R.A. and van den Brandt, P.A. (2003) K-ras oncogene mutations in sporadic colorectal cancer in The Netherlands Cohort Study. *Carcinogenesis*, **24**, 703-710.
8. Kralj, M., Husnjak, K., Korbler, T. and Pavelic, J. (2003) Endogenous p21WAF1/CIP1 status predicts the response of human tumor cells to wild-type p53 and p21WAF1/CIP1 overexpression. *Cancer Gene Ther*, **10**, 457-467.
9. Liu, Y. and Bodmer, W.F. (2006) Analysis of P53 mutations and their expression in 56 colorectal cancer cell lines. *Proc Natl Acad Sci U S A*, **103**, 976-981.
10. de las Alas, M.M., Aebi, S., Fink, D., Howell, S.B. and Los, G. (1997) Loss of DNA mismatch repair: effects on the rate of mutation to drug resistance. *J Natl Cancer Inst*, **89**, 1537-1541.
11. Leroy, B., Girard, L., Hollestelle, A., Minna, J.D., Gazdar, A.F. and Soussi, T. (2014) Analysis of TP53 mutation status in human cancer cell lines: a reassessment. *Hum Mutat*, **35**, 756-765.
12. Romano, D., Maccario, H., Doherty, C., Quinn, N.P., Kolch, W. and Matallanas, D. (2013) The differential effects of wild-type and mutated K-Ras on MST2 signaling are determined by K-Ras activation kinetics. *Mol Cell Biol*, **33**, 1859-1868.
13. Niu, B., Scott, A.D., Sengupta, S., Bailey, M.H., Batra, P., Ning, J., Wyczalkowski, M.A., Liang, W.W., Zhang, Q., McLellan, M.D. *et al.* (2016) Protein-structure-

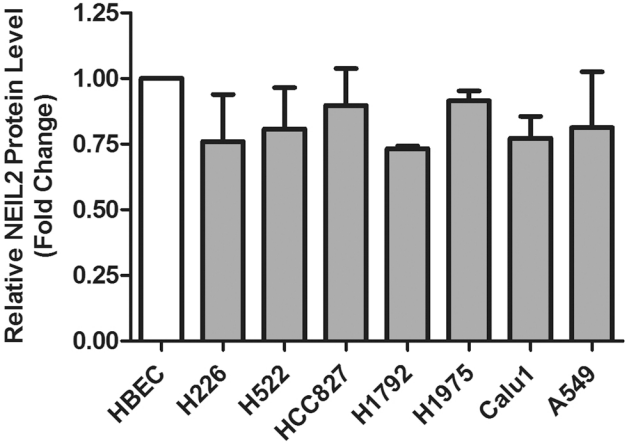
guided discovery of functional mutations across 19 cancer types. *Nat Genet*, **48**, 827-837.

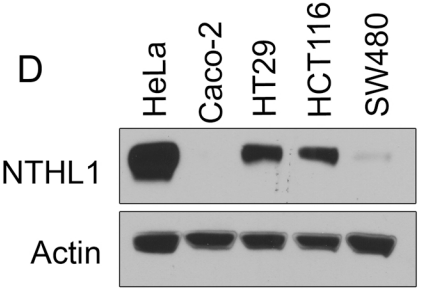
14. Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A. *et al.* (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*, **150**, 1107-1120.
15. Cancer Genome Atlas Research, N. (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, **511**, 543-550.
16. Cancer Genome Atlas Research, N. (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, **489**, 519-525.
17. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, **2**, 401-404.
18. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, **6**, pl1.

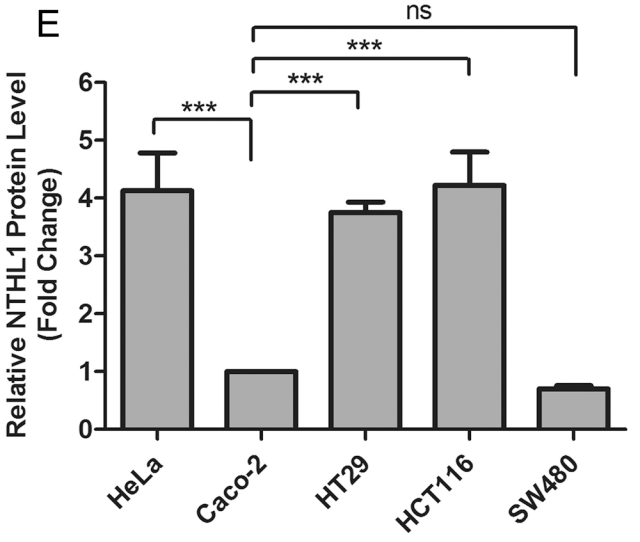
Supplementary Figure 1



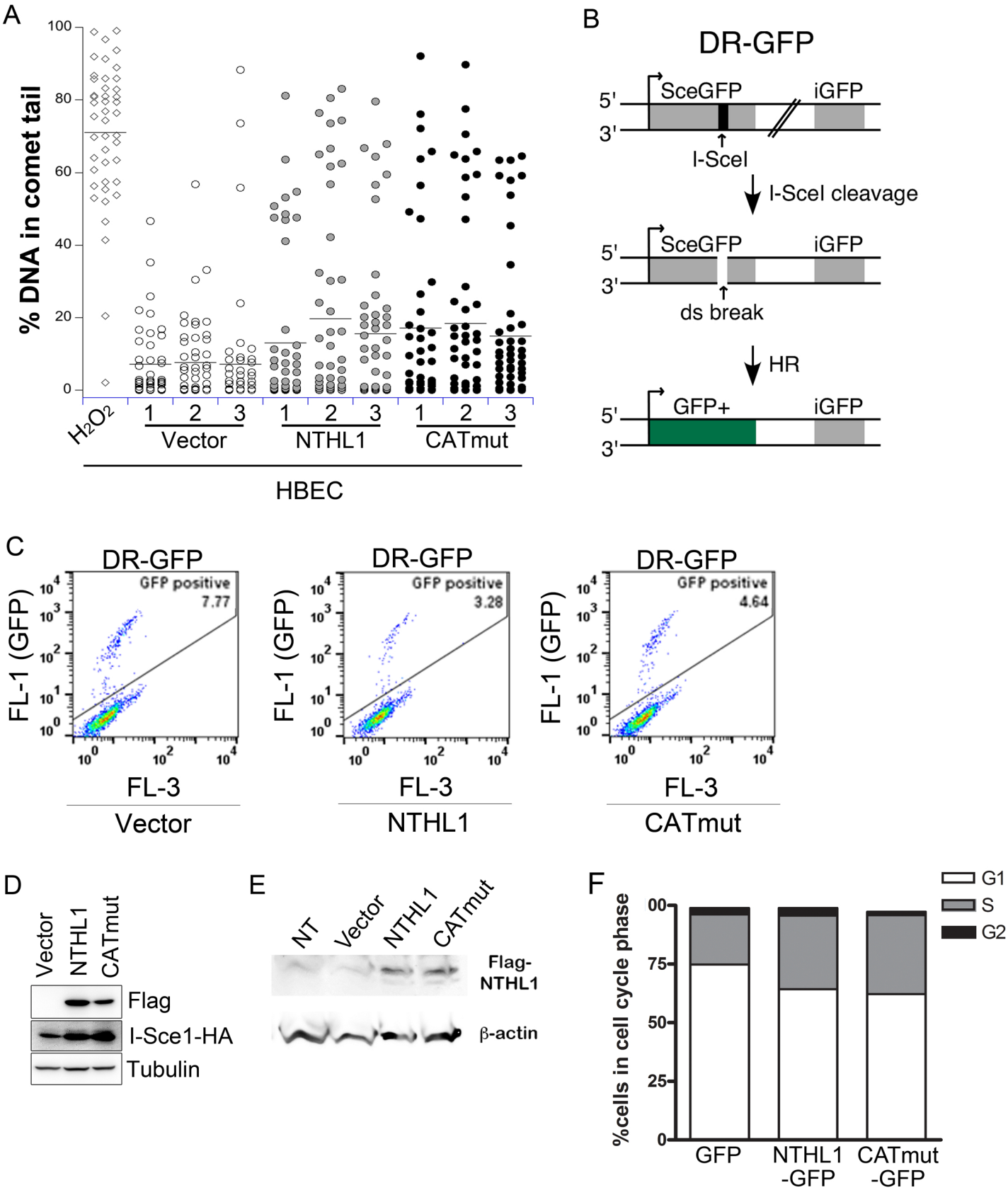




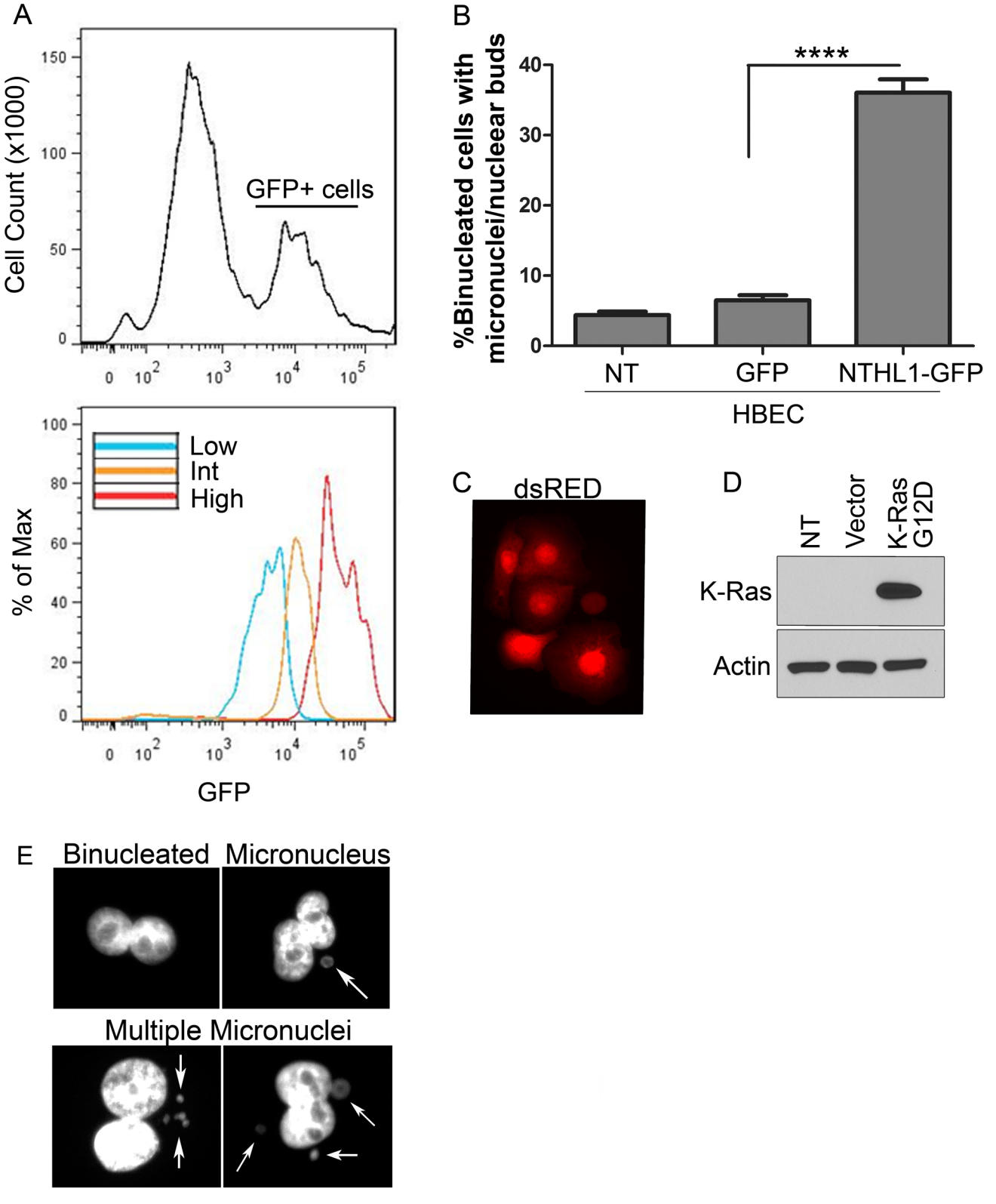




Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4