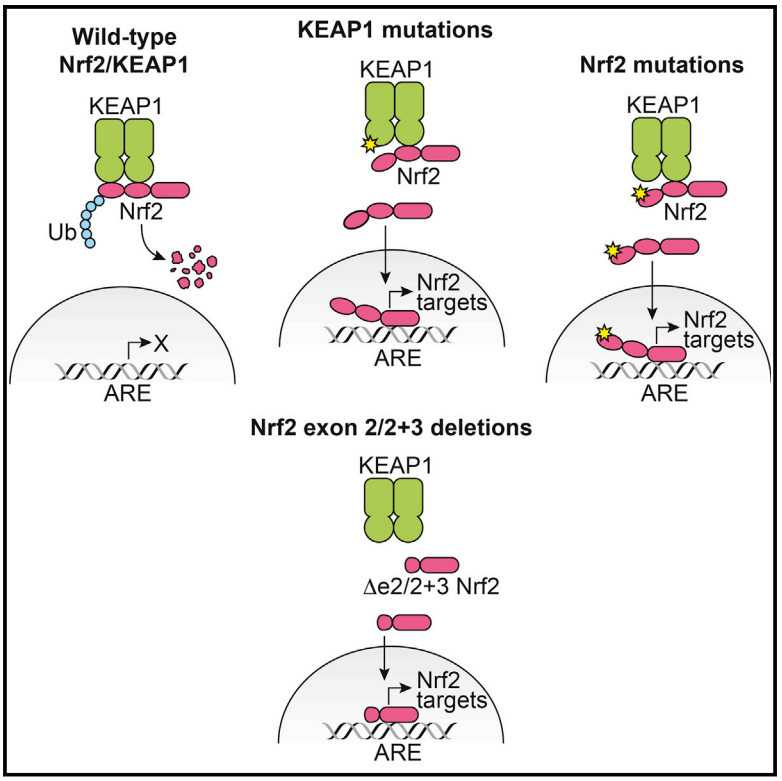
Article

Cell Reports

Recurrent Loss of *NFE2L2* Exon 2 Is a Mechanism for Nrf2 Pathway Activation in Human Cancers

Graphical Abstract



Highlights

- Recurrent removal of *NFE2L2* exon 2 occurs in lung cancers and head and neck cancers
- Exon removal results in Nrf2 stabilization and downstream pathway activation
- Exon deletions explain Nrf2 activation in patients without mutations in Nrf2 or KEAP1

Authors

Leonard D. Goldstein, James Lee, Florian Gnad, ..., Somasekar Seshagiri, Robert Gentleman, David Stokoe

Correspondence

stokoe.david@gene.com

In Brief

Goldstein et al. report recurrent removal of exon 2 from *NFE2L2* transcripts in lung cancers and head and neck cancers. Exon 2 loss prevents Nrf2 degradation and results in protein stabilization and transcription of Nrf2 target genes. Exon deletions explain Nrf2 activation in patients without detectable mutations in the pathway.

Goldstein et al., 2016, Cell Reports *16*, 2605–2617
September 6, 2016
http://dx.doi.org/10.1016/j.celrep.2016.08.010

Cell Reports
Article

OPEN ACCESS
CellPress

Recurrent Loss of *NFE2L2* Exon 2 Is a Mechanism for Nrf2 Pathway Activation in Human Cancers

Leonard D. Goldstein,[1,3,7] James Lee,[2,7] Florian Gnad,[3] Christiaan Klijn,[3] Annalisa Schaub,[2,8] Jens Reeder,[3] Anneleen Daemen,[3] Corey E. Bakalarski,[3,4] Thomas Holcomb,[5] David S. Shames,[5] Ryan J. Hartmaier,[6] Juliann Chmielecki,[6] Somasekar Seshagiri,[1] Robert Gentleman,[3,9] and David Stokoe[2,10,*]

[1]Department of Molecular Biology
[2]Department of Discovery Oncology
[3]Department of Bioinformatics and Computational Biology
[4]Department of Protein Chemistry
[5]Department of Oncology Biomarker Development
Genentech Inc., South San Francisco, CA 94080, USA
[6]Foundation Medicine Inc., Cambridge, MA 02141, USA
[7]Co-first author
[8]Present address: Helmholtz Center Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany
[9]Present address: 23AndMe Inc., Mountain View, CA 94041, USA
[10]Lead Contact
*Correspondence: stokoe.david@gene.com
http://dx.doi.org/10.1016/j.celrep.2016.08.010

SUMMARY

The Nrf2 pathway is frequently activated in human cancers through mutations in Nrf2 or its negative regulator KEAP1. Using a cell-line-derived gene signature for Nrf2 pathway activation, we found that some tumors show high Nrf2 activity in the absence of known mutations in the pathway. An analysis of splice variants in oncogenes revealed that such tumors express abnormal transcript variants from the *NFE2L2* gene (encoding Nrf2) that lack exon 2, or exons 2 and 3, and encode Nrf2 protein isoforms missing the KEAP1 interaction domain. The Nrf2 alterations result in the loss of interaction with KEAP1, Nrf2 stabilization, induction of a Nrf2 transcriptional response, and Nrf2 pathway dependence. In all analyzed cases, transcript variants were the result of heterozygous genomic microdeletions. Thus, we identify an alternative mechanism for Nrf2 pathway activation in human tumors and elucidate its functional consequences.

INTRODUCTION

Lung cancer is the primary cause of cancer-related death for men and women in the United States. The majority of lung cancers are non-small-cell lung cancers (NSCLCs), and most of these belong to either adenomatous or squamous subtypes. Recent studies have identified the genomic alterations that underlie these tumors, showing distinct patterns of mutational processes (Cancer Genome Atlas Research Network, 2012, 2014; Imielinski et al., 2012). While both subtypes show a high frequency of TP53 and NF1 mutations, KRAS and EGFR mutations are more common in adenomatous tumors, whereas PIK3CA mutations are more prevalent in squamous tumors. Adenomatous and squamous lung tumors additionally show recurrent mutations in KEAP1, an adaptor protein that recruits the Cul3 ubiquitin ligase to the transcription factor Nrf2 (encoded by the *NFE2L2* gene). In addition, Nrf2 is frequently mutated in squamous tumors (Shibata et al., 2008). Both of these alterations cause loss of Nrf2 ubiquitination and stabilization of the protein, resulting in high-level transcription of Nrf2 target genes (Leinonen et al., 2014).

KEAP1/Nrf2 represents one of the most important cellular pathways for sensing and responding to increases in oxidative stress. Under normal conditions, Nrf2 is constitutively targeted for ubiquitination by KEAP1, mediated through N-terminal DLG (low-affinity) and ETGE (high-affinity) interaction motifs that bind a KEAP1 dimer (McMahon et al., 2006; Tong et al., 2006a). When oxidative stress or electrophile levels increase, key cysteine residues in KEAP1 become oxidized, decreasing the ability of KEAP1 to cause Nrf2 ubiquitination (Tong et al., 2006b). Nrf2 target genes are situated downstream of defined antioxidant response elements (AREs), which restore cellular homeostasis by increasing cellular reducing power through increases in glutathione levels, NADPH levels, and upregulation of drug transporters (Hirotsu et al., 2012). Tumor-associated mutations in *NFE2L2* and *KEAP1* thereby hijack a highly regulated cellular process that provides a selective advantage for the affected cells in a manner that is not completely understood.

In this study, we identify tumor-specific alterations that have been missed by conventional sequencing and copy number approaches. Using publicly available RNA sequencing (RNA-seq) data from The Cancer Genome Atlas (TCGA) project, we identify recurrent transcript variants in key oncogenes that are not detected in normal tissue samples. One of the top hits from this analysis is *NFE2L2*, with alterations seen in squamous NSCLC and head and neck squamous carcinoma (HNSC). We explore

the mechanisms underlying these alterations and their consequences on Nrf2 activity and function.

RESULTS

Identification of NSCLC Cell Lines with Mutations in *KEAP1* or *NFE2L2*

We documented the status of *KEAP1* and *NFE2L2* in a panel of 113 NSCLC cell lines profiled by RNA-seq, exome sequencing (exome-seq), or SNP arrays to identify mutations, copy-number changes, and loss of heterozygosity (LOH) of these genes (Figure S1A; Table S1). *KEAP1* mutations were found in 29/113 cell lines (26%), and *NFE2L2* mutations were detected in 4/113 cell lines (4%). With the exception of NCI-H661, all *KEAP1* mutant cell lines showed homozygous expression of the mutant allele, which was associated with copy-neutral LOH. In contrast, the *NFE2L2* mutations were heterozygous and not associated with LOH. A further two cell lines (HCC1534 and NCI-H1437) showed no detectable *KEAP1* mRNA, due to bi-allelic loss of the *KEAP1* gene. The Nrf2 mutations were in the previously identified hotspots in the KEAP1 interface regions (Figure S1B) (Shibata et al., 2008) and comprised point mutations and an in-frame three-amino-acid deletion. The mutations in KEAP1 were spread throughout the primary sequence (Figure S1C), with few obvious hotspots. However, when mapped onto the KEAP1/Nrf2 peptide crystal structure (Fukutomi et al., 2014), the mutations clustered in the loops extending from the KEAP1 core beta propeller close to the interaction site with Nrf2 (Figure S1D).

Identification of a Mutant KEAP1 Gene Signature

To determine the transcriptional consequences of KEAP1 mutations in NSCLC cell lines, we asked which genes were significantly differentially expressed (adjusted $p < 0.01$, absolute mean $\log_2$ fold change > 1) in the KEAP1 mutant compared to wild-type cell lines. Overall, 27 genes were significantly upregulated in the KEAP1 mutant cell lines (Figure 1A; Figure S1E; Table S2), many of which have previously been identified as Nrf2 target genes from chromatin immunoprecipitation sequencing (ChIP-seq), RNA-seq, or microarray studies (Chorley et al., 2012; Hirotsu et al., 2012; MacLeod et al., 2009; Malhotra et al., 2010). Only one gene was identified as significantly downregulated using these cutoffs (*HSPB1*). The protein targets of these genes were also increased in KEAP1 mutant cell lines compared to wild-type (Figure S1F).

Hierarchical clustering of 230 TCGA lung adenocarcinomas based on the expression of these 27 genes resulted in the division of two major groups (Figure 1B). One group was mainly characterized by high expression of the 27 signature genes and contained 43 tumors, out of which 32 (74%) were KEAP1 mutant. The other group, characterized by low expression, contained 187 tumors, out of which 179 (96%) were KEAP1 wild-type. When using the same gene set to cluster TCGA lung squamous cell carcinomas, Nrf2 and KEAP1 mutant tumors were distinguished from the Nrf2/KEAP1 wild-type tumors (Figure 1C), suggesting that Nrf2 mediates most of the transcriptional consequences of KEAP1 loss/mutation. Interestingly, there were several squamous NSCLC tumors that showed high expression of the 27 signature genes without any known mutations in either KEAP1 or Nrf2.

Identification of Abnormal *NFE2L2* Transcript Variants in Tumor Samples

We asked whether Nrf2 pathway activation in these tumors might be driven by transcript alterations not recognized by whole-exome-seq. To address this question, we performed de novo prediction of splice variants from RNA-seq data and identified transcript variants in known oncogenes that are recurrently observed in TCGA cancer samples but are rarely seen in normal samples from the genotype-tissue expression (GTEx) project (see Experimental Procedures). We included 19 cancer types with at least 100 patients (6,359 samples in total). In 54 analyzed oncogenes, we identified nine recurrent candidate cancer-specific splice variants (>1% and $n \geq 3$ for a given cancer type) (Table S3). Using the same detection criteria as in cancer samples, none of these variants could be detected in normal controls (2,958 samples in total). Grouping together related variants with shared splice sites yielded five independent alterations in four oncogenes (Figure 2A). These alterations included several well-documented oncogenic transcript variants, including *EGFRvIII* in brain cancers (Cho et al., 2011), *MET* exon 14 skipping in lung adenocarcinoma (Kong-Beltran et al., 2006), and *CTNNB1* exon 3 deletions in colorectal cancers (Iwao et al., 1998). Interestingly, we also observed previously uncharacterized splice variants in *NFE2L2*, which occurred in patients with squamous NSCLC (3.3%; 16/481) and at lower prevalence in patients with HNSC (1.5%; 6/403) (Figure 2A). A more detailed analysis of *NFE2L2* splice variants in lung squamous cell carcinoma revealed two splice variants co-occurring in the same patients, corresponding to a skip of *NFE2L2* exon 2 in mRNAs transcribed from either one of two alternative promoters (2.1%; 10/481) (Figure 2B). Two additional splice variants co-occurred in a distinct set of patients (1.2%; 6/481), corresponding to a skip of *NFE2L2* exons 2 and 3 (exon 2+3) in mRNAs with either one of the two alternative transcript starts (Figure 2B). All patients expressing *NFE2L2* splice variants lacking exon 2 or exon 2+3 also showed expression of normal *NFE2L2* transcripts as evidenced by split reads supporting inclusion of exon 2. Both exons 2 and 3 are part of the *NFE2L2* coding sequence. For transcripts transcribed from the distal promoter, skip of exon 2 or exon 2+3 was predicted to result in an in-frame deletion (p.D16_Q104del or p.D16_E134del for NP_006155) (Figure S2). Transcripts for the proximal promoter have a start codon in exon 2. In this case, the skip of exon 2 or exon 2+3 was predicted to cause translation from downstream internal start codons, resulting in an N-terminal truncation (p.M1_C103del or p.M1_G163del for NP_001138884). The recurrence of *NFE2L2* transcripts lacking exon 2 and preservation of coding potential suggest that these splice variants may present gain-of-function events conferring a selective advantage. This is supported by the finding that exon 2 encodes the Neh2 domain, containing the DLG and ETGE motifs, which allow for interaction with KEAP1 (Itoh et al., 1999) and which are mutated in 15% of squamous lung cancers (Cancer Genome Atlas Research Network, 2012).

NFE2L2 Exon 2 Deletion Is Associated with Nrf2 Target Gene Expression in Tumors

To assess whether the observed *NFE2L2* transcript variants can account for Nrf2 pathway activation in patients without known

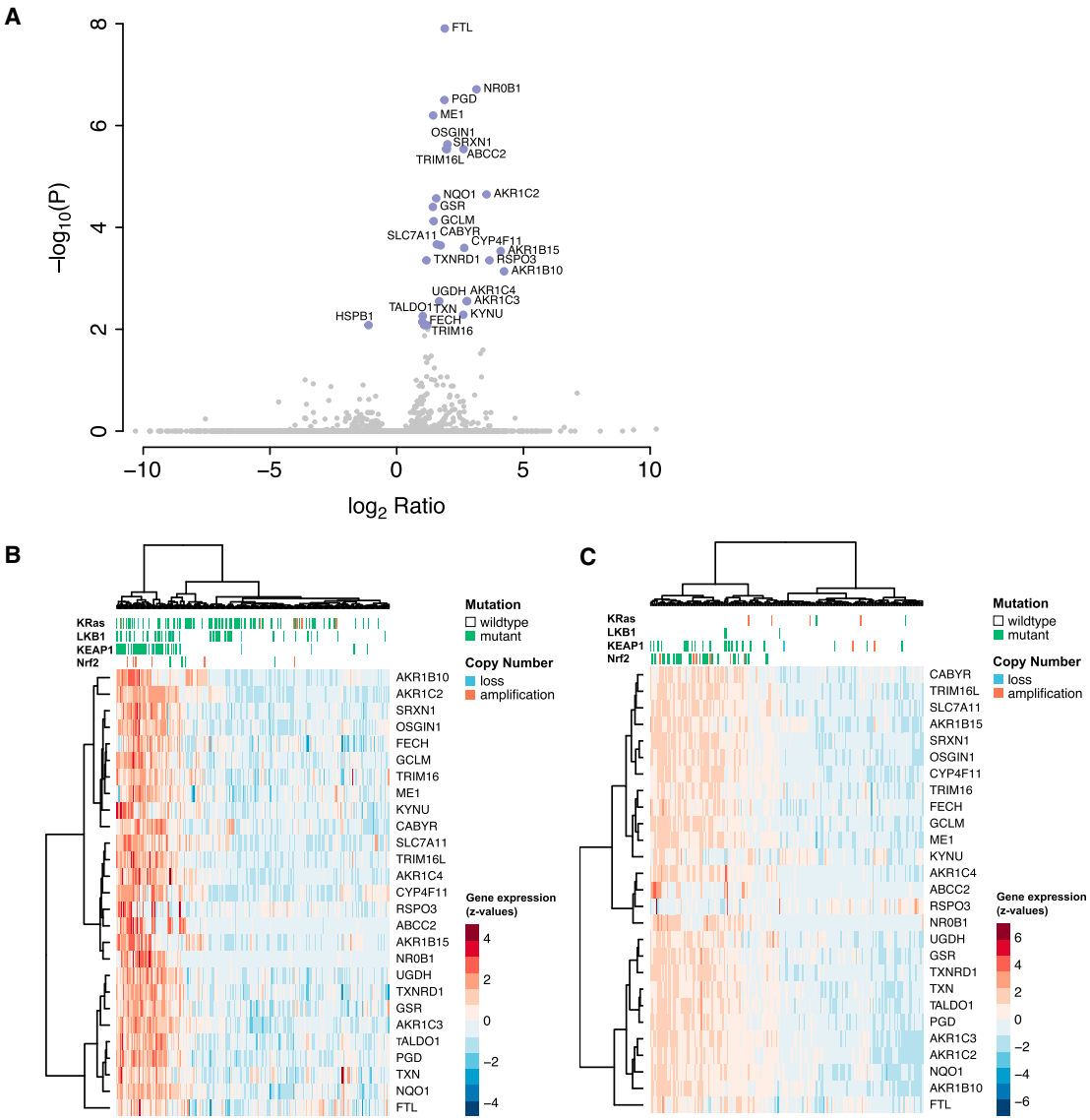


Figure 1. A Cell-Line-Derived Mutant KEAP1 Gene Signature Classifies Human Lung Tumors

(A) Volcano plot of $\log_2$ fold changes and adjusted *p* values for differential expression analysis between mutant (n = 25) and wild-type (n = 74) *KEAP1* NSCLC cell lines. Significantly differentially expressed genes (absolute $\log_2$ fold change > 1; adjusted p < 0.01) are indicated.

(B) Using unsupervised ward clustering, the KEAP1 gene signature classifies 32 (80%) of the 40 *KEAP1* mutant lung adenocarcinomas from TCGA. The mutation status for selected genes is indicated indicated for each sample.

(C) The KEAP1 gene signature classifies 19 (86%) of 22 *KEAP1* mutant and 27 (100%) of 27 *NFE2L2* mutant lung squamous cell carcinomas from TCGA.

See also Figure S1 and Tables S1 and S2.

mutations in *KEAP1* or *NFE2L2*, we analyzed the co-occurrence of *NFE2L2* transcript variants and Nrf2 pathway mutations. In the TCGA collection, 178 of the squamous lung tumors with RNA-seq data were also profiled by exome-seq. In this subset, ten tumors (6%) displaying the exon 2 or exon 2+3 deletion were mutually exclusive with 48 tumors (27%) showing mutations in either *NFE2L2* or *KEAP1* (Figure 2C). Moreover, all exon-2-deleted tumors showed high expression of the 27 candidate Nrf2 target genes (Figure 2D). Similar observations were made for 275 TCGA head and neck cancers, where *NFE2L2* exon deletions in five tumors (2%) were mutually exclusive with mutations in *NFE2L2* or *KEAP1* in 26 tumors (9%) (Figures 2E and 2F). These results suggest that the deletion of exon 2 represents an alternative mechanism for activation of Nrf2 in a subset of squamous NSCLC and head and neck cancers.

Validation of *NFE2L2* Transcript Variants in Cell Lines

To identify cell line models for further study, we analyzed RNA-seq data from a large panel of human cancer cell lines (described in Klijn et al., 2015; unpublished data) and searched for evidence

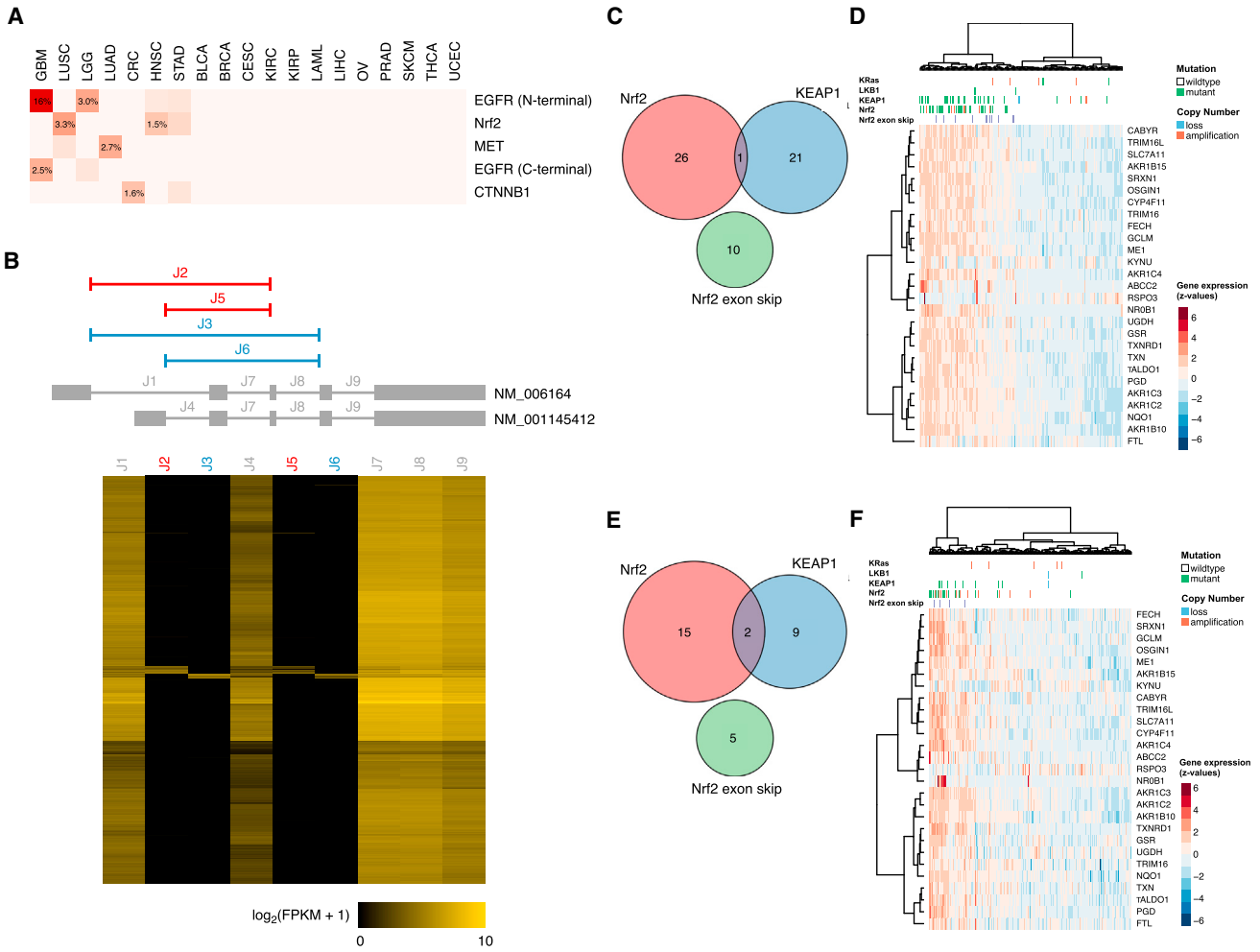


Figure 2. *NFE2L2* Transcript Variants Lacking Exon 2 Are Observed in Lung and in Head and Neck Squamous Cell Carcinoma

(A) Heatmap of frequency of recurrent transcript alterations in 19 cancer types. GBM, glioblastoma; LUSC, lung squamous cell carcinoma; LGG, lower grade glioma; LUAD, lung adenocarcinoma; CRC, colorectal cancer; HNSC, head and neck squamous cell carcinoma; STAD, stomach adenocarcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast cancer; CESC, cervical cancer; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LIHC, liver hepatocellular carcinoma; OV, ovarian carcinoma; PRAD, prostate adenocarcinoma; SKCM, skin cutaneous melanoma; THCA, thyroid carcinoma; UCEC, endometrial carcinoma.

(B) *NFE2L2* exons and splice junctions predicted from RNA-seq data. Two annotated transcripts are shown in gray. Identified junctions corresponding to the skip of exon 2 (J2, J5) or exon 2+3 (J3, J6) are shown in red and blue, respectively. Heatmap illustrating read evidence for junctions (columns) across 481 TCGA lung squamous cell carcinoma patients (rows) on an FPKM scale after $\log_2(x + 1)$ transformation.

(C) Venn diagram illustrating mutual exclusive occurrence of *NFE2L2* transcript alteration and mutation in *KEAP1* or *NFE2L2*.

(D) Clustering of squamous NSCLC based on 27 candidate Nrf2 target genes. Mutation status and *NFE2L2* transcript alteration are indicated for each sample.

(E and F) Venn diagram and clustering for head and neck cancers, otherwise as in (C) and (D).

See also Figure S2 and Table S3.

of the identified splice variants. Out of 611 cell lines, we identified one multiple myeloma cell line, KMS-27, and one hepatocellular carcinoma cell line, JHH-6, both showing evidence for a heterozygous skip of *NFE2L2* exon 2 by junction reads (Figure 3A). We validated the *NFE2L2* exon 2 skipping by RT-PCR in JHH-6 and KMS-27 mRNA. Using a series of forward and reverse primers derived from exon 1 and from exons 3 and 4, we confirmed the exon 2 deletion (Δe2 *NFE2L2*) in mRNA isolated from JHH-6 and KMS-27 cells, respectively (Figure S3A). Sequencing of the PCR products confirmed the expected deletion of exon 2 (Figure S3B). Based on RNA-seq data, we did not detect point mutations in the coding sequence of *NFE2L2* or *KEAP1* in JHH-6 or KMS-27 (Klijn et al., 2015; data not shown).

Given that Nrf2/KEAP1 alterations are common in hepatocellular carcinoma (10%) but infrequent in multiple myeloma (0%) (http://www.cbioportal.org), we focused our attention on JHH-6 cells. We asked whether the exon-2-deleted form of Nrf2 protein (Δe2 Nrf2) was expressed in this line. We performed western blotting of whole-cell lysates from JHH-6 cells, as well as the KEAP1 mutant HUH-1 line, and from HuCCT1 as a

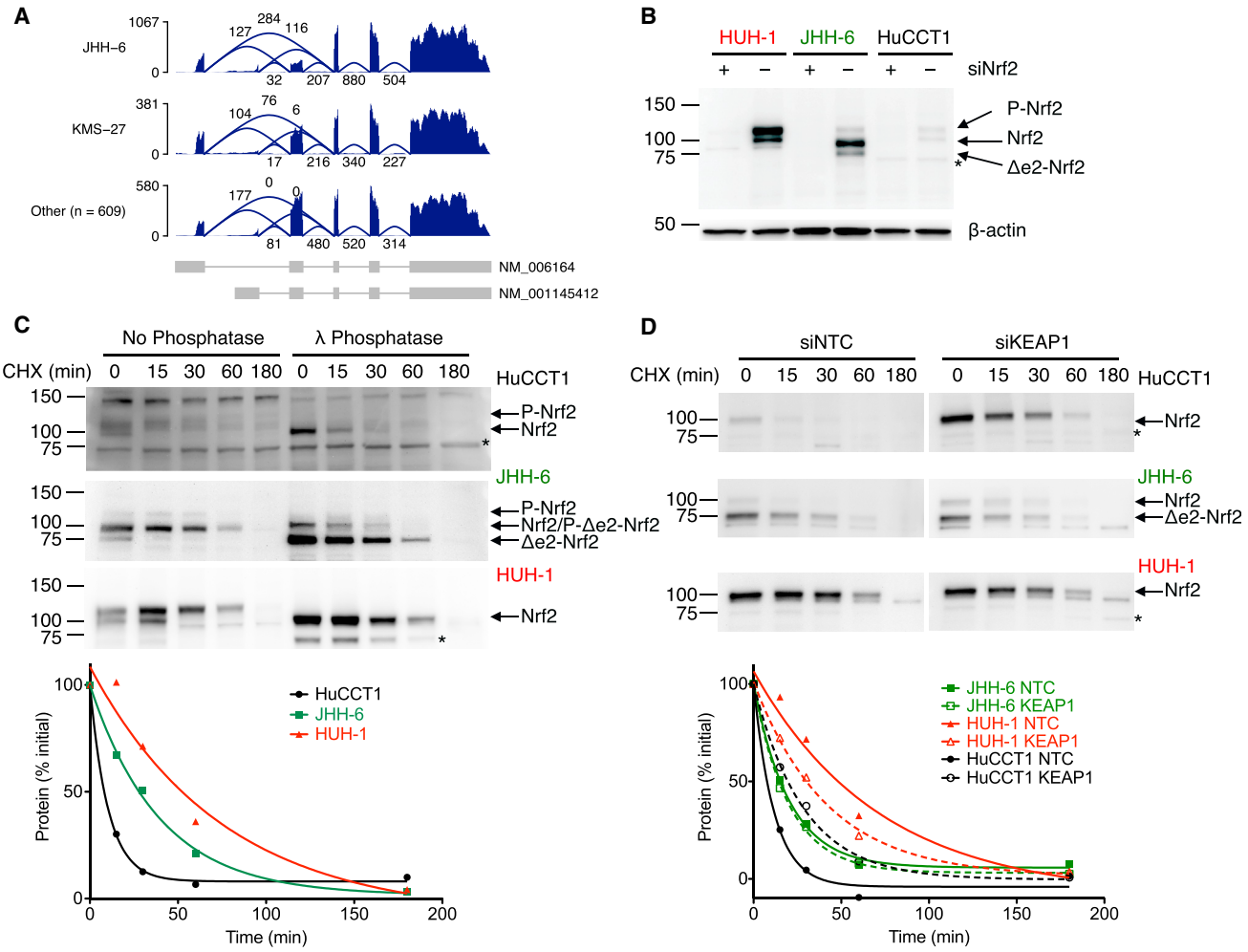


Figure 3. Validation of Altered *NFE2L2* mRNA and Nrf2 Protein in KMS-27 and JHH-6 Cell Lines

(A) RNA-seq data from KMS-27 and JHH-6 cells show presence of junction reads skipping exon 2. Read counts for each sample were scaled by a size factor based on total read count for the sample divided by the median per-sample count. The bottom panel shows average scaled counts across 609 samples.

(B) Protein lysates from the indicated cell lines were separated by SDS-PAGE, and Nrf2 and actin were visualized by western blotting. The bands representing wild-type Nrf2, phosphorylated Nrf2 (P-Nrf2), and exon-2-deleted Nrf2 (Δe2-Nrf2) are indicated. Asterisk represents a likely non-specific band, as it is not depleted by Nrf2 siRNA (siNrf2).

(C) HuCCT1, JHH-6, or HUH-1 cells were grown in six-well dishes and then treated with 100 μg/ml cycloheximide (CHX) for the indicated times. The lysates were incubated with either buffer or 400 U lambda phosphatase for 30 min before separation by SDS-PAGE and western blotting with Nrf2 antibodies. Band intensities of dephosphorylated full-length Nrf2 (for HuCCT1 and HUH-1 cells) or Δe2-Nrf2 (for JHH-6 cells) were then quantified and fitted to a one-phase decay curve.

(D) Cells were treated as in (C), except that they were transfected with either siNTC (50 nM) or siKEAP1 (50 nM) for 72 hr prior to harvest. Lysates were all treated with 400 U lambda phosphatase for 30 min prior to separation by SDS-PAGE and western blotting with Nrf2 antibodies. Full-length Nrf2 (for HuCCT1 and HUH-1 cells) and Δe2-Nrf2 (for JHH-6 cells) were quantified and fitted to a one-phase decay curve.

See also Figure S3.

representative wild-type *KEAP1* liver cancer cell line. The levels of Nrf2 in JHH-6 cells were comparable to those seen in HUH-1 cells, which were much higher than in the wild-type KEAP1 HuCCT1 cells (Figure 3B). Moreover, a species with smaller molecular weight, consistent with a deletion of exon 2, was detectable in JHH-6 and was reduced upon *NFE2L2* small interfering RNA (siRNA) transfection, confirming that it, indeed, represents a form of Nrf2. We suspected that a phosphorylated form of Δe2 Nrf2 might co-migrate with unphosphorylated wild-type Nrf2 in the 4%–12% gels used. Indeed, dephosphorylation of JHH-6 lysates confirmed that Δe2 Nrf2 was significantly more abundant than the wild-type form (Figure 3C, middle panel). Similarly, KMS-27 cells expressed Δe2 Nrf2, which was the major species apparent upon dephosphorylation (Figure S3C).

We also analyzed the stability of Nrf2 in the three liver cancer cell lines, using cycloheximide to halt total protein synthesis. We used dephosphorylated lysates to allow more accurate quantification of total Nrf2. The experiments showed increased stability of Δe2 Nrf2 in JHH-6 cells, comparable to Nrf2 in HUH-1 cells, which were both more stable than Nrf2 in HuCCT1

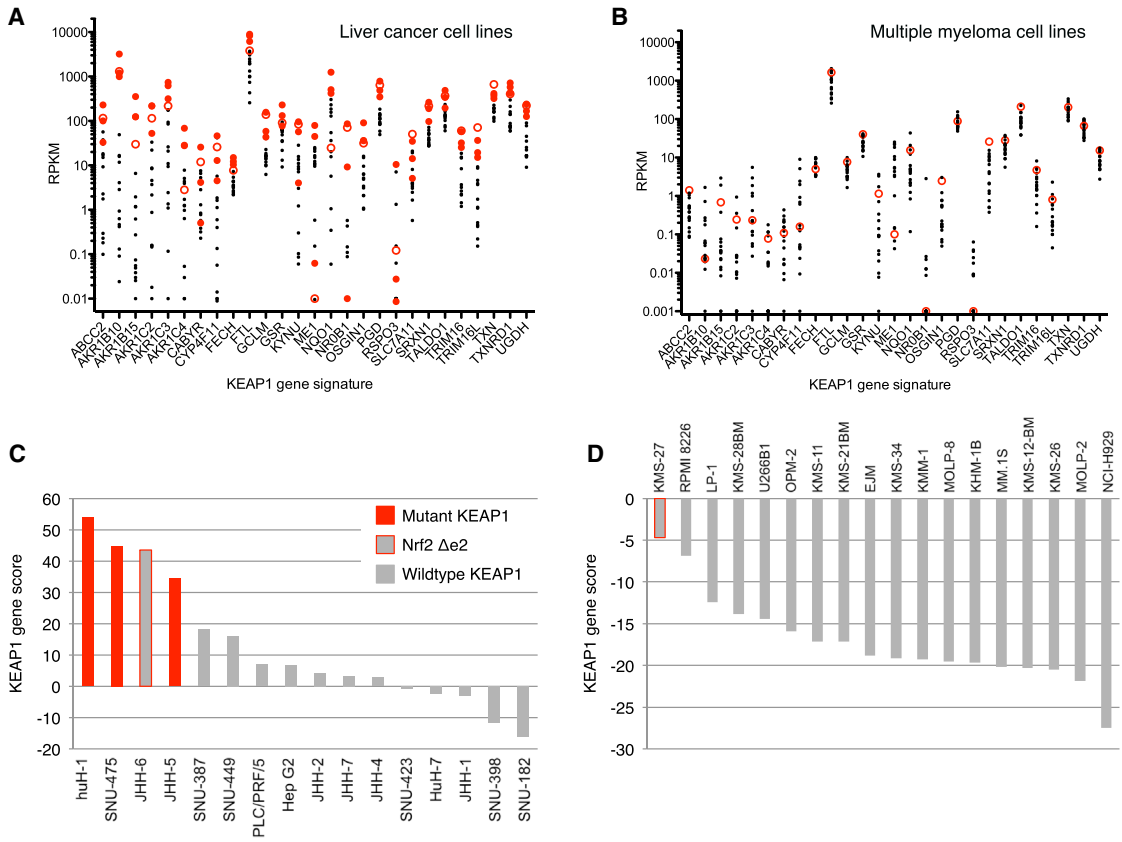


Figure 4. Nrf2 Is Activated in Cell Lines Expressing Exon-2-Deleted *NFE2L2*

(A and B) Expression of 27 Nrf2/KEAP1 target genes in 16 hepatocellular carcinoma cell lines (A) and 18 multiple myeloma cell lines (B) using RNA-seq data (described in Klijn et al., 2015; unpublished data). KMS-27 and JHH-6 are indicated as red open circles, and mutant *KEAP1* liver cancer cell lines are indicated as solid red circles.

(C and D) Bar graphs show the KEAP1 gene score (sum of *Z* scores for the 27 Nrf2/KEAP1 target genes determined over the full dataset) in the 16 hepatocellular carcinoma cell lines (C) and 18 multiple myeloma cell lines (D). *KEAP1* and *NFE2L2* alterations are indicated as filled and outlined boxes, respectively.

cells (Figure 3C). Δe2 Nrf2 in JHH-6 cells also showed prominent nuclear localization (Figure S3D). Next, we asked whether the deletion of exon 2 in JHH-6 cells made Nrf2 refractory to regulation by KEAP1. Knockdown of *KEAP1* in HuCCT1 cells resulted in increased steady-state levels of Nrf2 due to increased stability (Figure 3D). However, knockdown of *KEAP1* in JHH-6 cells did not affect the levels or stability of Δe2 Nrf2 (although the wild-type form of Nrf2 in these cells showed increased stability). As expected, knockdown of *KEAP1* did not increase the stability of wild-type Nrf2 in the KEAP1 mutant HUH-1 cell line (Figure 3D). *KEAP1* mRNA levels were equally decreased by *KEAP1* siRNA in all three cell lines (Figure S3E).

NFE2L2 Exon 2 Deletion Is Associated with Nrf2 Target Gene Expression in Cell Lines

Using the Nrf2/KEAP1 gene signature on RNA-seq data from 16 hepatocellular carcinoma cell lines, we found that JHH-6 cells show high expression of Nrf2 target genes, similar to cell lines with *KEAP1* mutations (Figure 4A). Similarly, analysis of 18 multiple myeloma cell lines demonstrated that KMS-27 cells show high expression of these genes (Figure 4B). Expression of the 27 candidate target genes can be summarized by a "KEAP1 gene score," calculated as the sum of *Z* scores for individual target genes, with *Z* scores computed across the 611 cell lines examined. This results in a single score per cell line that reflects the extent of overexpression of signature genes. The KEAP1 gene score confirms that JHH-6 cells show a similar score as *KEAP1* mutant liver cancer cell lines (Figure 4C) and that KMS-27 cells show the highest score among multiple myeloma cell lines (Figure 4D), despite cell lines from this cancer showing a low overall KEAP1 gene score.

Deletion of Exon 2 Results in Dependency on Nrf2

Next, we addressed whether JHH-6 cells expressing Δe2 *NFE2L2* were more dependent on the expression of Nrf2 protein, compared to wild-type *NFE2L2*-expressing HuCCT1 cells. Knockdown of *NFE2L2* in JHH-6 cells caused a marked decrease in cell viability, similar to that seen in the mutant *KEAP1* hepatocellular carcinoma cell line HUH-1. In contrast, *NFE2L2* knockdown had a more modest effect on the viability of HuCCT1 cells (Figure 5A). This was not due to defective *NFE2L2* knockdown in HuCCT1 cells, as knockdown was

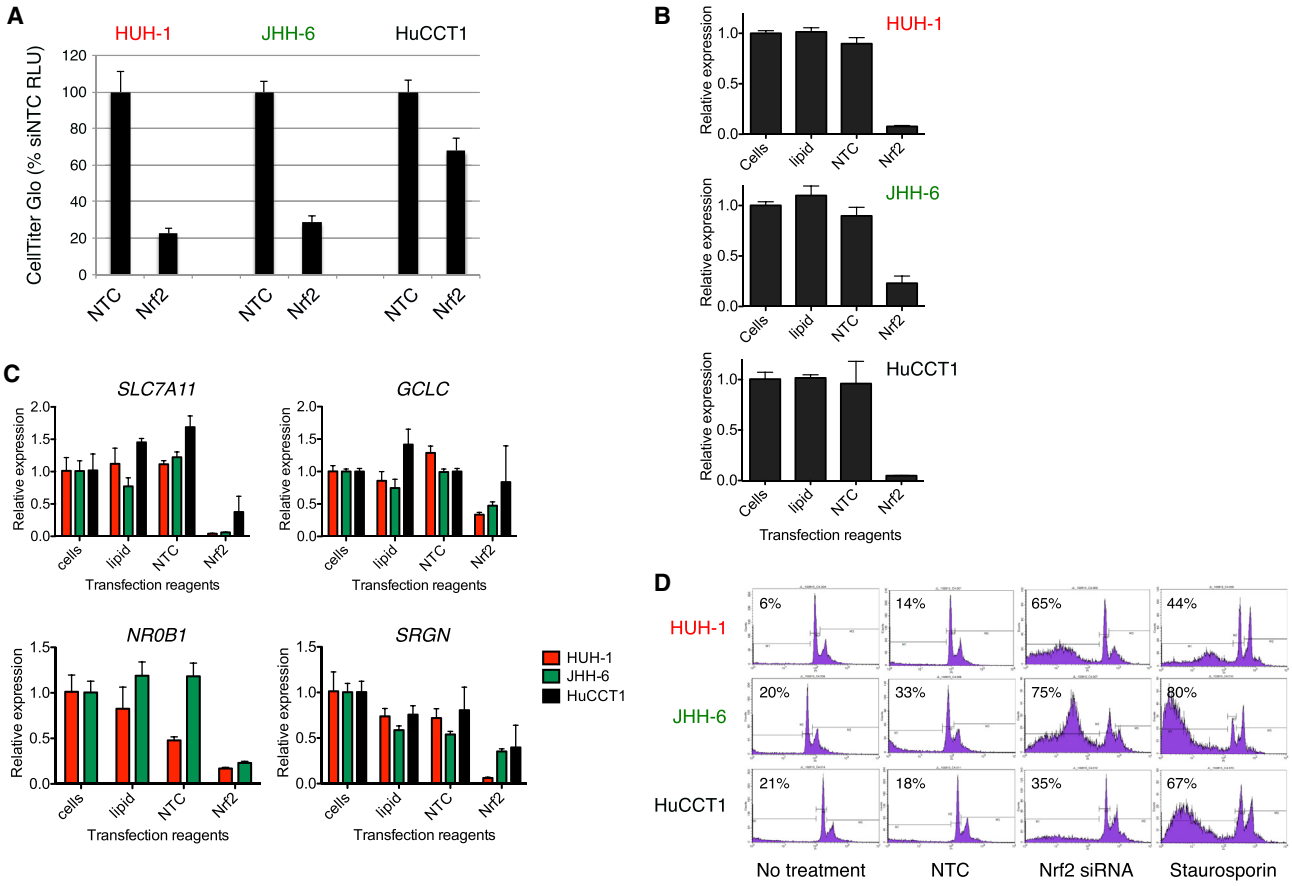


Figure 5. Requirement for Exon-2-Deleted Nrf2 for Cell Viability

(A) HUH-1, JHH-6, and HuCCT1 cells were seeded into 96-well plates containing either non-targeted siRNAs (NTC) or siRNAs targeting Nrf2. Viability was measured 4 days later using CellTiter-Glo. RLU, relative luminescence units.

(B) HUH-1, JHH-6, and HuCCT1 cells were grown in six-well dishes (Cells), treated with lipid agents only (lipid), and transfected with siRNAs against non-targeting control (NTC) or *NFE2L2* sequences (Nrf2). 48 hr later, total RNA was isolated, and expression of *NFE2L2* was measured using Taqman probes.

(C) Cells were treated as in (B), but Taqman probes against the indicated Nrf2 target genes were used. *NR0B1* expression was below detection levels in HuCCT1 cells.

(D) The indicated cell lines were reverse transfected with non-targeting control siRNA or *NFE2L2* siRNA and were analyzed 96 hr later for propidium iodide (PI) staining. The percent staining in the sub-G1 compartment is indicated. Staurosporine treatment (24 hr) was used as a positive control.

Error bars in (A)–(C) indicate mean ± SD (n = 3).

equally efficient in all three cell lines (Figure 5B). Knockdown of *NFE2L2* also resulted in the decreased expression of four Nrf2 target genes, although the effect was less pronounced in the wild-type *KEAP1* HuCCT1 cell line (Figure 5C). Decreased viability was likely due, at least in part, to apoptosis, as measured by an increase in fragmented DNA (Figure 5D).

Deletion of Exon 2 or Exon 2+3 Abolishes KEAP1 Interaction and Maintains Target Gene Expression

To address how loss of *NFE2L2* exon 2 affects the ability of Nrf2 to be regulated by KEAP1, we used transient expression in 293 cells. KEAP1 decreased the expression of full-length Nrf2 but had lesser effects on the expression of Nrf2 lacking exon 2 or exon 2+3 (Figure 6A, upper panel). The inhibitory effect of KEAP1 on full-length Nrf2 expression was mostly abolished by the proteasome inhibitor MG132, as expected. Full-length Nrf2 and KEAP1 interacted with each other, whereas deletion of exon 2 or exon 2+3 completely abolished the ability of KEAP1 to bind Nrf2 (Figure 6A, lower two panels). As a result of this, altered Nrf2 remained stable following KEAP1 expression, in contrast to wild-type Nrf2 (Figure 6B), although the altered forms of Nrf2 appeared to have slightly decreased intrinsic stability. Altered Nrf2 isoforms were transcriptionally active, as judged by their ability to increase Nrf2 target gene expression (Figure 6C). Most genes were similarly increased by Δe2 Nrf2 or Δe2+3 Nrf2, compared to full-length Nrf2, and were resistant to the effects of KEAP1 overexpression. Interestingly, Δe2+3 Nrf2 was defective for increasing *GPX2* expression, suggesting that there might be subtle differences in the transcriptional activation of this form of Nrf2. Consistent with this observation, we also noticed that 21 of the 27 target genes shown in Figure 1, as well as *GPX2*, showed lower

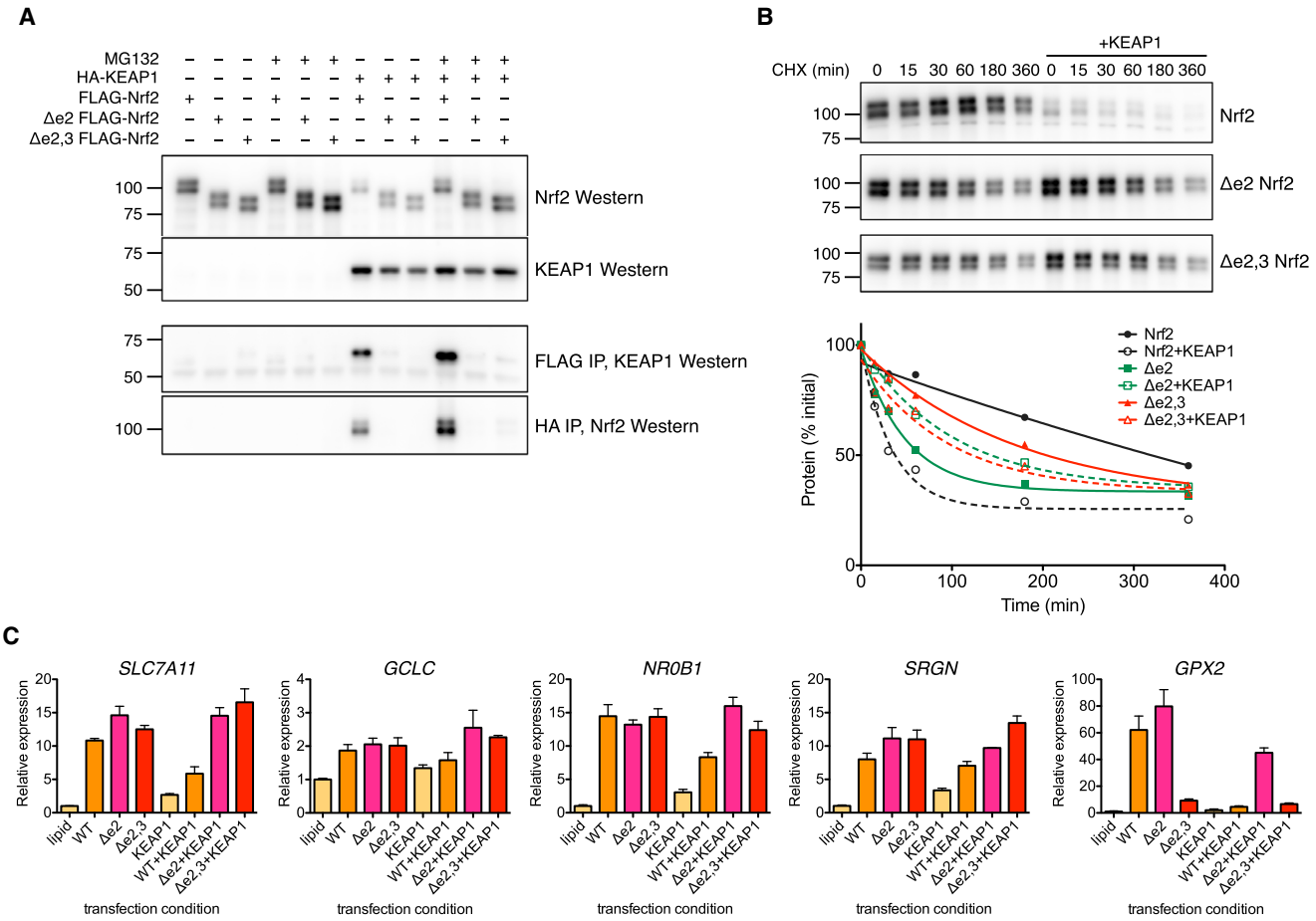


Figure 6. Effect of *NFE2L2* Exon 2 Deletion on Interaction with KEAP1

(A) 293 cells were transfected with plasmids expressing FLAG-Nrf2, Δe2 FLAG-Nrf2, Δe2,3 FLAG-Nrf2, or HA-KEAP1. 48 hr after transfection, cells were lysed, and either whole-cell lysates (top gels) or anti-FLAG or HA immunoprecipitates (IPs) were analyzed by western blotting using the indicated antibodies.

(B) 293 cells were transfected with the same plasmids as described in (A) and then treated with 100 μg/ml cycloheximide (CHX) for the indicated times. Cells were lysed and separated by SDS-PAGE and western blotted using Nrf2 and β-actin antibodies.

(C) Cells were treated as in (A) but harvested for total RNA, which was used to analyze the expression of the indicated genes using Taqman RT-PCR. Error bars indicate mean ± SD (n = 3).

See also Figure S4.

median expression in exon 2+3-deleted squamous lung tumors, compared to exon 2-deleted tumors ($p < 0.01$, two-sided binomial test; Figure S4).

NFE2L2 Transcript Variants Are Caused by Genomic Alterations

Finally, we investigated the mechanism underlying the observed *NFE2L2* transcript alterations. Analysis of exome-seq data for KMS-27 and JHH-6 showed a relative decrease in reads mapping to exon 2, suggesting that the observed transcript variants could be the result of genomic alterations (Figure 7A). Indeed, whole-genome sequencing of JHH-6 and KMS-27 showed that these cell lines harbor microdeletions surrounding *NFE2L2* exon 2, spanning 4,685 and 2,981 nt, respectively (Figure 7B; Figure S5A). To investigate the causal mechanism in patients, we analyzed comprehensive hybrid capture-based genomic profiling data (Foundation Medicine) from a large cohort of clinical samples including different types of cancers (n = 58,707). In this dataset, we identified 11 out of 1,218 squamous NSCLC tumors that showed a decrease in copy number for exon 2 or exon 2+3 compared to nearby control regions (Experimental Procedures; Figure 7C). The focal nature of the deletions can be appreciated by investigating log ratios from defined genomic regions targeted for sequencing (Figure S5B). We also identified seven squamous NSCLC tumors with discordant read pairs that were consistent with structural variants encompassing several kilobases of DNA and affecting exon 2 or exon 2+3 (Figure 7D). In total, 16 out of 1,218 squamous NSCLC patients showed evidence for genomic alterations affecting *NFE2L2* exon 2 or exon 2+3, and the identified events were mutually exclusive with point mutations or indels in *NFE2L2* and *KEAP1*, which were observed in 171 patients and are known to activate this pathway. Although we found a total of 45 cases with evidence for genomic alterations in the complete cohort, squamous

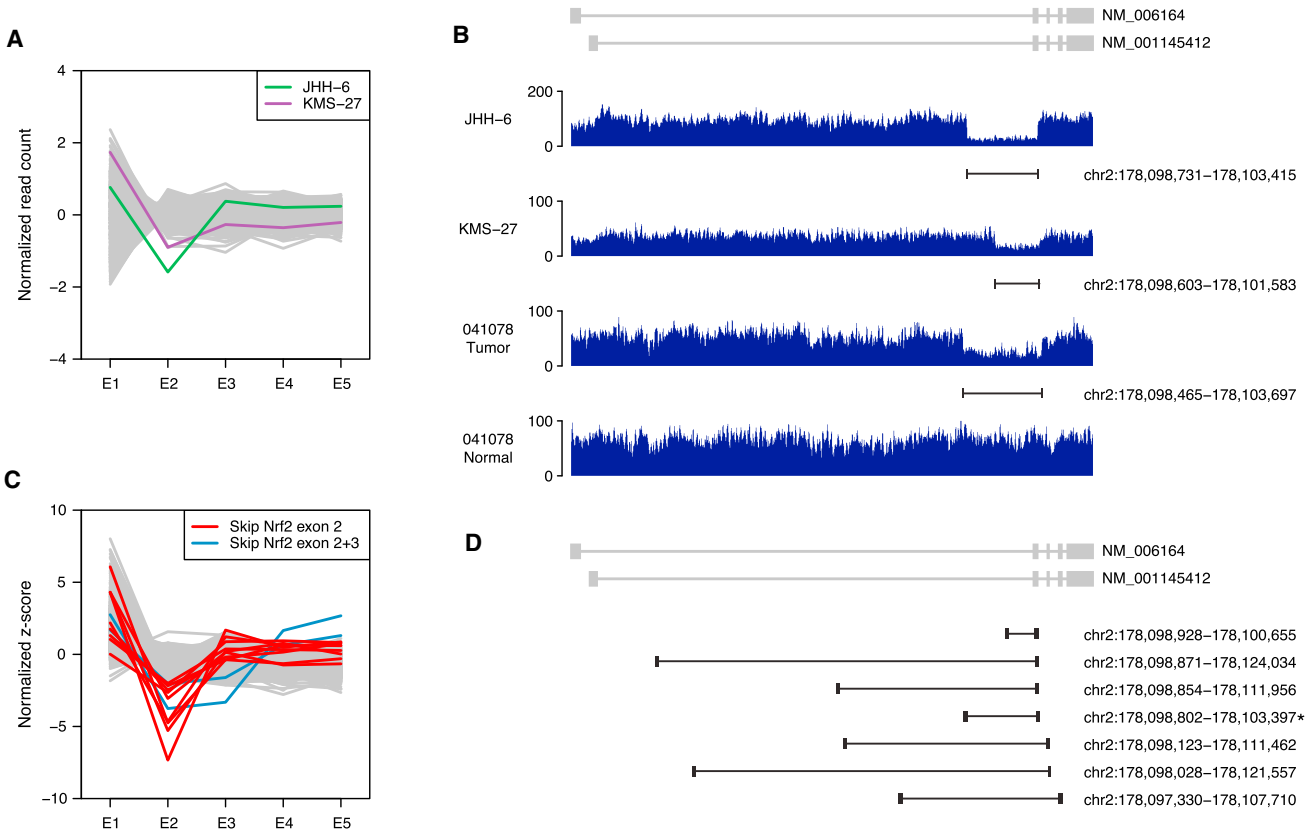


Figure 7. Genomic Microdeletions Underlie the Loss of *NFE2L2* Exon 2

(A) Normalized exome-seq read counts for *NFE2L2* exons across a panel of cancer cell lines (unpublished data). JHH-6 and KMS-27 are highlighted in green and purple, respectively. Per-exon read counts were mean centered across exons and samples after $\log_2(x + 1)$ transformation. E, exon.

(B) Heterozygous genomic deletions of *NFE2L2* exon 2 identified by whole-genome sequencing of cell lines JHH-6 and KMS-27 and a lung tumor and adjacent normal tissue.

(C) Normalized *Z* scores for exon read coverage across 1,218 squamous NSCLC tumors. Tumors with identified exon 2 deletion or exon 2+3 deletion are highlighted in red and blue, respectively. *Z* scores were mean centered across exons for each sample.

(D) Genomic location of structural variants affecting *NFE2L2* exon 2 or exon 2+3 identified based on discordant read pairs. Asterisk indicates an inversion.

See also Figures S5 and S6.

NSCLC was the only cancer type that showed a statistically significant enrichment (adjusted $p = 3 \times 10^{-14}$, two-sided Fisher's exact test). We note that our ability to detect structural variants in this dataset is limited. Identification of heterozygous deletions is challenging due to variability in read coverage, while breakpoints for structural variants are only identified if they are located in close proximity to a genomic region targeted for sequencing. Therefore, we analyzed an additional cohort of 45 squamous NSCLC tumors for which both RNA and DNA were available. RT-PCR analysis identified a single patient with loss of exon 2, which was strongly enriched in the tumor compared to adjacent normal tissue (Figure S6A). RNA-seq analysis confirmed that the transcript variant was expressed in the identified tumor but was absent in adjacent normal tissue (Figure S6B). Expression of Nrf2 target genes was elevated to a similar extent as in TCGA tumors with known mutations in the pathway, whereas the adjacent normal tissue showed low target gene expression (Figure S6C). Finally, whole-genome sequencing confirmed that the transcript variant was the result of a somatic genomic microdeletion of 5,233 nucleotides surrounding exon 2 (Figure 7B; Figure S5A). These data suggest that genomic microdeletions are a clinically relevant mechanism for Nrf2 pathway activation.

DISCUSSION

Genomic analysis of a large panel of lung cancer cell lines (Klijn et al., 2015; data not shown) allowed us to identify genes that are differentially expressed in *KEAP1* mutant compared to wild-type cell lines. All but one of the genes identified showed increased expression, consistent with the evidence that the KEAP1 target Nrf2 is a positive regulator of gene transcription. Fifteen of the 27 genes had been previously identified as Nrf2 target genes from ChIP-seq and RNA-seq experiments in cells from different tissues and species (Chorley et al., 2012; Hirotsu et al., 2012; Malhotra et al., 2010), and a further 8 were identified as showing similar expression to "classic" Nrf2 target genes (DeNicola et al., 2015) or upregulated upon *KEAP1* knockdown in keratinocytes

(MacLeod et al., 2009), implying that the set of genes regulated by Nrf2 is highly conserved across different tissues and conditions. This has practical value in the use of a single gene signature to identify tumors with Nrf2 activation in both NSCLC and HNSC (Figure S6D). Interestingly, the Nrf2/KEAP1 signature is only activated in tumors. Matched normal samples for lung and for head and neck tumors showed only low Nrf2 target gene activity (Figure S6E), suggesting that inhibition of the pathway might have selective benefit in tumors showing pathway deregulation compared to normal tissues.

Intragenic genomic deletions that result in activation of proto-oncogenes have previously been reported for a number of genes, including *EGFR* and *CTNNB1*. Such variants are not routinely assayed, which is due, in part, to the limitations of current genomic technologies. Alterations affecting individual exons and involving small copy-number changes are difficult to detect by exome-seq. Therefore, intragenic deletions have remained relatively unexplored, and new variants are still being discovered. Recent studies of small-cell lung cancer and adult T cell leukemia/lymphoma identified recurrent microdeletions in *TP73*, *IKZF2*, and *CARD11* using whole-genome sequencing (George et al., 2015; Kataoka et al., 2015). Here, we used publicly available RNA-seq data generated as part of the TCGA project to identify recurrent transcript alterations in known oncogenes. Due to differences between patient cohorts, it is difficult to accurately assess the general prevalence of *NFE2L2* exon deletions. When analyzing TCGA lung squamous cell cancers with available RNA-seq data (n = 481), we identified 3% (16/481) of patients with the deletion of *NFE2L2* exon 2 or exon 2+3. When analyzing the subset of patients with available exome-seq data (n = 178), for which somatic mutation calling can be performed, the proportion of patients with *NFE2L2* exon deletions was 6% (10/178). Thus, accounting for *NFE2L2* exon deletions increased the percentage of patients with putative Nrf2 pathway activation from 27% (48/178) to 33% (58/178) in lung squamous carcinoma and from 9% (26/275) to 11% (31/275) in head and neck squamous carcinoma, compared to assessing mutations in *NFE2L2* or *KEAP1* by exome-seq alone (Figures 2C and 2E). Analysis of real-world clinical samples from patients that underwent genomic profiling suggested a prevalence of *NFE2L2* exon deletions in 1%–2% of lung squamous cell carcinoma. However, the latter analysis lacks sensitivity, since only unambiguous deletions were considered. In the future, more comprehensive screening of known cancer genes may be achievable through the sequencing of complete gene loci, including introns, or by combining data from exome-seq and RNA-seq experiments.

We confirmed deletions of *NFE2L2* exon 2 at the genomic, transcript, and protein levels using cell-line models JHH-6 and KMS-27. Cell-line studies further demonstrated that deletion of *NFE2L2* exon 2 or exon 2+3 results in the loss of interaction with KEAP1, Nrf2 stabilization, induction of a Nrf2 transcriptional response, and Nrf2 pathway dependence. However, we acknowledge that these results were obtained in the specific cellular and genetic context of the cell lines used in this study. Interestingly, even in cell lines with mutations in *KEAP1* or *NFE2L2*, Nrf2 protein levels are still unstable. Nrf2 stability is likely regulated by additional factors; for example, β-TrCP (Chowdhry et al., 2013) and/or SIAH2 (Baba et al., 2013). Differing levels of these or other E3 ligase adaptors could explain residual Nrf2 instability, as well as differences in Nrf2 stability between JHH-6 and HUH-1 cells.

Deletion of exon 2 provides an elegant mechanism to increase Nrf2 activity by removing the interaction site with KEAP1 while keeping the remainder of the gene functionally intact. Indeed, biochemical analyses confirmed an almost complete loss of KEAP1 binding and stabilization of Nrf2 when exon 2 is deleted (Figures 6A and 6B). When considering *NFE2L2* point mutations found in tumors, mutations surrounding the ETGE high-affinity binding site result in complete loss of KEAP1 interaction, whereas mutations in the lower affinity DLG motif vary in their ability to disrupt the Nrf2/KEAP1 complex (Fukutomi et al., 2014; Shibata et al., 2008). However, even point mutations that do not disrupt the complex change the nature of the interaction so as to prevent KEAP1-mediated ubiquitination of Nrf2 (Shibata et al., 2008). While the interaction with KEAP1 is similarly abolished in the case of deletion of exon 2+3, exon 3 contains the Neh4 domain that has been previously implicated in transcriptional activation by Nrf2 through binding to CBP (CREB [cAMP response element-binding protein]-binding protein) (Katoh et al., 2001). Neh4 (exon 3) and Neh5 (exon 4) were shown to act synergistically in recruiting CBP. The Neh4 and Neh5 domains of Nrf2 have also been shown to be important for binding the Mediator complex (Sekine et al., 2015) and the SWI/SNF component BRG1 (Zhang et al., 2006), demonstrating the importance of these domains in Nrf2 transcriptional activation. Consistent with this, we observe decreased ability of Δe2+3 Nrf2 to induce some Nrf2 target genes compared to Δe2 Nrf2 or point mutations in *NFE2L2* (Figure 6C; Figure S4). However, recurrence of Δe2+3 Nrf2 suggests that it is selected in the evolution of lung tumors. It will be interesting to define more precisely the transcriptional profile of this Nrf2 isoform to help inform the specific subset of genes that are required for the tumorigenic properties of Nrf2 in lung cancer. It might also be interesting to test any emerging CBP inhibitors for their ability to inhibit GPX2 and other Nrf2 target genes in KEAP1/Nrf2 mutant tumors.

Deletions found in human tumors that remove the interaction domain with E3 ligases have also been seen in other genes. For example, 7 out of 222 colorectal tumors showed small genomic deletions surrounding exon 3 of β-catenin (Iwao et al., 1998) that remove the interaction site for its E3 ligase β-TrCP (Hart et al., 1999). Similarly, the majority of TMPRSS-ERG and TMPRSS-ETV1 fusion proteins found in prostate cancer encode truncated versions of transcription factors that render them resistant to ubiquitination and degradation mediated by SPOP or COP1, respectively (An et al., 2015; Vitari et al., 2011). In addition, mutations resulting in *MET* exon 14 skipping remove amino acid residue Y1003, which is required for Cbl recruitment and subsequent ubiquitination and downregulation. Therefore, small intragenic deletions represent effective mechanisms for nascent oncogenes to escape normal degradation during tumor initiation and evolution. This suggests that increased use of both whole-genome sequencing and integrative analysis with RNA-seq will help identify oncogenes activated by this and other mechanisms.

EXPERIMENTAL PROCEDURES

TCGA and GTEx Data

TCGA RNA-seq data were obtained from the Cancer Genomics Hub at the University of California, Santa Cruz. Raw FASTQ files were downloaded in batches as they became available. For samples without available FASTQ files, BAM files were converted to FASTQ. The latest TCGA downloads were based on a manifest created on July 14, 2015. Mutation and copy-number data were retrieved from cBioPortal (http://www.cbioportal.org), using the R software package CGDS-R (Cerami et al., 2012; Gao et al., 2013). Raw FASTQ files for GTEx RNA-seq data were downloaded from dbGaP (http://www.ncbi.nlm.nih.gov/gap) using accession code phs000424.v4.p1.

RNA-Seq Analysis

RNA-seq data were aligned to the human reference genome (GRCh37/hg19) using GSNAP version 2013-10-10 (Wu and Nacu, 2010) (parameters: -M 2 -n 10 -B 2 -i 1 -N 1 -w 200000 -E 1 --pairmax-rna=200000 --clip-overlap). Gene expression levels for RefSeq genes were quantified based on counts and RPKM (reads per kilobase of target and million reads sequenced) values.

Derivation of a Mutant KEAP1 Gene Expression Signature

RNA-seq data for 99 NSCLC cell lines (Table S1) were retrieved from the European Genome-phenome Archive: EGAS00001000610 (Klijn et al., 2015). Using the DESeq R package (Anders and Huber, 2010) we measured differential gene expression between *KEAP1* mutant (point mutation or genomic deletion) and wild-type cell lines, reported as fold change and associated adjusted *p* values. Signature genes were selected based on an adjusted $p < 0.01$ and mean $\log_2$ fold change >1. For ward clustering of samples and genes (using Euclidean distance) variance-stabilized count data were used. The NMF (Non-negative Matrix Factorization) R package was used to create associated heatmaps. The KEAP1 gene score was computed as the sum of *Z* scores for 27 signature genes based on read counts after voom transformation (Law et al., 2014).

Splice Variant Analysis

Analysis of splice variants was performed using the R/Bioconductor package SGSeq (version 1.3.14) (Goldstein et al., 2016). We predicted exons and splice junctions from BAM files for 7,384 TCGA samples (parameters: alpha = 2, psi = 0, beta = 0.2, and gamma = 0.2) at 54 genomic loci of known oncogenes (Vogelstein et al., 2013). Predictions were merged across samples and assembled into a splice graph. Splice events consisting of two or more alternative splice variants were identified from the graph. Evidence for splice variants was quantified in terms of FPKM (fragments per kilobase and million fragments mapped) and estimates of relative usage. Relative usage estimates with a denominator less than 20 were set to NA. Splice variants detected in TCGA samples were also quantified in 2,958 GTEx samples from normal human tissues (GTEx Consortium, 2015).

Identification of Cancer-Specific Splice Variants

We considered internal splice variants (not involving alternative transcript starts or ends) and required the start and end of each splice variant to either overlap or extend exons that belong to annotated RefSeq transcripts downloaded from the UCSC (University of California, Santa Cruz) Genome Browser website (Pruitt et al., 2005; Rosenbloom et al., 2015). Retained introns were excluded. We considered 19 TCGA cancer types that included samples from at least 100 patients (6,359 cancer samples in total) and selected splice variants with (1) an FPKM >2 and estimated relative usage >0.2 in at least one cancer sample, (2) an FPKM <1 in > 99.9% of GTEx samples, and (3) an FPKM of 0 in >97.5% of GTEx samples. FPKM-based criteria were required to be satisfied at both the start and the end of the splice variant. Variants satisfying the FPKM-based criteria for which relative usage could not be estimated were included after manual inspection. Filtering for recurrence (> 1%, $n \geq 3$ for a given cancer type) resulted in nine variants, each corresponding to a single exon-exon junction. In subsequent analyses, these variants were considered present in samples with an FPKM >2 (Table S3).

Analysis of the FoundationCORE Comprehensive Genomic Profiling Database

Samples were processed and sequenced similarly as previously described (Frampton et al., 2013). We screened the November 2015 data freeze of FoundationCORE (n = 58,707) for *NFE2L2* exon 2 or exon 2+3 deletions using two approaches. First, rearrangement calls based on discordant read pairs and/or split reads were examined for the loss of *NFE2L2* exon 2 or exon 2+3. Since intronic regions of *NFE2L2* are not captured, this approach is limited to the identification of deletions in which breakpoints occur near intron-exon boundaries or within exons. The second approach uses copy-number log ratios for individual bait regions. Copy-number log ratios were determined with an in-house algorithm educated to the specific tumor cellularity of each sample. A *Z* score was calculated comparing the log ratio for each exon in *NFE2L2* to control polymorphism capture regions immediately adjacent to *NFE2L2* (n = 15; evenly spaced every ~1 MB from ~3 MB upstream and ~12 MB downstream of *NFE2L2*). Exon 2 deletions with and without a concurrent exon 3 deletion were identified (for simplicity, referred to as "exons of interest," or EOIs). EOI deletions were called if (1) *Z* scores were less than −2 for EOIs and not for non-EOIs in *NFE2L2* and (2) there was a log-ratio drop of 0.2 from non-EOIs in *NFE2L2*. To investigate whether specific cancer types were associated with genomic alterations in *NFE2L2* exon 2 or exon 2+3 (detected by copy-number log ratio or discordant read pairs), we considered 85 cancer types with at least 100 patients. For each cancer type, we used a two-sided Fisher's exact test to assess whether cases were enriched or depleted compared to the rest of the dataset. The *p* values were adjusted for multiple testing using the Benjamini-Hochberg method.

Cell Culture

KMS-27 (RPMI-1640), JHH-6 (Williams Media E), HuCCT1 (RPMI-1640), and HUH-1 (DMEM) cells were from JCRB (Japanese Collection of Research Bioresources Cell Bank), and 293 (Eagle's minimal essential medium; EMEM) cells were from ATCC; they were cultured in the indicated media in the presence of 2 mM glutamine and 10% fetal bovine serum (FBS).

Western Blotting

Cell lysates were prepared with RIPA Buffer (Sigma) supplemented with complete EDTA-free protease inhibitor (Roche) and phoSTOP (Roche), Phosphatase Inhibitor Cocktail 2 (Sigma), and Phosphatase Inhibitor Cocktail 3 (Sigma). Lysates were run on Novex Tris-Glycine 4%–12% gradient gels (ThermoFisher) and then transferred onto iBlot nitrocellulose (Invitrogen). Blots were pre-incubated in 5% skim milk powder (Merck) in Tris-buffered saline with Tween 20 (TBST; 10 mM Tris [pH 8], 150 mM NaCl, 0.1% Tween 20), followed by 5% BSA (Sigma) in TBST containing antibodies. Secondary antibodies used were ECL Anti-Rabbit HRP (horseradish peroxidase) and ECL Anti-Mouse HRP (both from GE Healthcare). Blots were developed with the ChemiGlow Chemiluminescence Substrate Kit (Protein Simple) and visualized with a FluorChem HD2 imager (Protein Simple). Antibodies used in this study are against KEAP1 (Cell Signaling, G1D10), Nrf2 (Abcam, ab62352), HSP90 (Cell Signaling, 4877), HDAC2 (Cell Signaling, 5113), β-actin (Sigma, A2228), hemagglutinin (HA; Roche, 11815016001), FLAG (Sigma, F2426). Lambda phosphatase was from New England Biolabs (P0753L), and phosphatase inhibitors were omitted from the lysis buffer in these experiments.

Cell Viability and DNA Fragmentation Analysis

siRNAs were reverse transfected into cells with Dharmafect 2 reagent (ThermoFisher) and Opti-MEM (GIBCO). Four days post-transfection, cells were measured for viability using CellTiter-Glo reagent (Promega), and luminescence was detected on an EnVision Multilabel Reader (PerkinElmer). Four days post-transfection, cells were measured for apoptosis using propidium iodide (LifeTechnologies) staining and flow cytometry following a published protocol (Riccardi and Nicoletti, 2006). Staurosporine (1 μM; Enzo Life Sciences) was added to positive control cells 24 hr pre-staining. *NFE2L2* siRNAs were: 5′-TGACAGAAGTTGACAATTA-3′ and 5′-GTAAGAAG CCAGATGTTAA-3′ and were used along with siGENOME Non-Targeting siRNA Pool #2 (GE Healthcare) as control siRNA. Stained cells were analyzed with a Becton Dickinson FACSCaliber instrument. siRNAs targeting *KEAP1* were from Dharmacon (L012453-00).

Taqman Analysis

Total cellular RNA was extracted with an RNeasy Kit (QIAGEN). RNA was converted to cDNA using a cDNA Reverse Transcription Kit (Applied Biosystems), and cDNAs were amplified with Taqman Gene Expression primer-probe sets (ThermoFisher), using Taqman Gene Expression Master Mix reagents (Applied Biosystems). Taqman amplification/detection was performed on the QuantStudio7 Flex Real-Time PCR System. The Hs00975961_g1 primer-probe set was used to detect *NFE2L2* (ThermoFisher). The Nrf2 target gene Taqman primer-probe sets used were as follows: *SLC7A11* (Hs00921938_m1), *SGRN* (Hs00921938_m1), *NR0B1* (Hs03043658_m1), *GCLC* (Hs00155249_m1), *GPX2* (Hs01591589_m1), and *KEAP1* (Hs00202227_m1), all from ThermoFisher.

293 Transfections

Target DNAs in pRK5 were transfected into cells using Lipofectamine 2000 (ThermoFisher) and Opti-MEM (GIBCO) as recommended by manufacturer's protocol.

ACCESSION NUMBERS

The accession number for the RNA-seq and whole-genome sequencing data reported in this paper is European Genome-phenome Archive (EGA; http://www.ebi.ac.uk/ega/): EGAS00001001740.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.08.010.

AUTHOR CONTRIBUTIONS

L.D.G. performed splice variant and related analysis. A.D. and J.R. coordinated the downloading and processing of TCGA data. C.K. coordinated the downloading and processing of GTEx data and performed the gene signature analysis. J.L. and A.S. performed RNA and protein validation of Nrf2 exon 2 deletion in KMS-27 and JHH-6 cells and transfection of 293 cells with Nrf2 constructs. F.G. generated the KEAP1/Nrf2 gene signature. C.E.B. performed protein comparisons in KEAP1 mutant and wild-type NSCLC cell lines. T.H. and D.S.S. contributed the squamous lung cancer cohort allowing whole-genome sequencing of a primary tumor sample with exon 2 deletion. R.J.H. and J.C. developed methods and analyzed genomic deletions in FoundationCORE. S.S., R.G., and D.S. analyzed data and supervised the project. All authors wrote the manuscript.

ACKNOWLEDGMENTS

The results shown in this study are, in part, based upon data generated by the TCGA Research Network (http://cancergenome.nih.gov/) and the GTEx Project. We thank the Genentech sequencing and the gCell distribution cores as well as Allison Bruce for help with the Graphical Abstract. We thank Jeff Settleman and Shiva Malek for useful discussions and insights. L.D.G., J.L., F.G., C.K., J.R., A.D., C.E.B., T.H., D.S.S., S.S., R.G., and D.S. are or have been employees of Genentech, and some hold shares in Roche. R.J.H. and J.C. are employees of and equity holders in Foundation Medicine. R.G. is an employee of 23AndMe Inc.

Received: December 5, 2015
Revised: June 3, 2016
Accepted: July 31, 2016
Published: August 25, 2016

REFERENCES

An, J., Ren, S., Murphy, S.J., Dalangood, S., Chang, C., Pang, X., Cui, Y., Wang, L., Pan, Y., Zhang, X., et al. (2015). Truncated ERG oncoproteins from TMPRSS2-ERG fusions are resistant to SPOP-mediated proteasome degradation. Mol. Cell *59*, 904–916.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Baba, K., Morimoto, H., and Imaoka, S. (2013). Seven in absentia homolog 2 (Siah2) protein is a regulator of NF-E2-related factor 2 (Nrf2). J. Biol. Chem. *288*, 18393–18405.

Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. Nature *489*, 519–525.

Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543–550.

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.

Cho, J., Pastorino, S., Zeng, Q., Xu, X., Johnson, W., Vandenberg, S., Verhaak, R., Cherniack, A.D., Watanabe, H., Dutt, A., et al. (2011). Glioblastoma-derived epidermal growth factor receptor carboxyl-terminal deletion mutants are transforming and are sensitive to EGFR-directed therapies. Cancer Res. *71*, 7587–7596.

Chorley, B.N., Campbell, M.R., Wang, X., Karaca, M., Sambandan, D., Bangura, F., Xue, P., Pi, J., Kleeberger, S.R., and Bell, D.A. (2012). Identification of novel NRF2-regulated genes by ChIP-seq: influence on retinoid X receptor alpha. Nucleic Acids Res. *40*, 7416–7429.

Chowdhry, S., Zhang, Y., McMahon, M., Sutherland, C., Cuadrado, A., and Hayes, J.D. (2013). Nrf2 is controlled by two distinct β-TrCP recognition motifs in its Neh6 domain, one of which can be modulated by GSK-3 activity. Oncogene *32*, 3765–3781.

DeNicola, G.M., Chen, P.H., Mullarky, E., Sudderth, J.A., Hu, Z., Wu, D., Tang, H., Xie, Y., Asara, J.M., Huffman, K.E., et al. (2015). NRF2 regulates serine biosynthesis in non-small cell lung cancer. Nat. Genet. *47*, 1475–1481.

Frampton, G.M., Fichtenholtz, A., Otto, G.A., Wang, K., Downing, S.R., He, J., Schnall-Levin, M., White, J., Sanford, E.M., An, P., et al. (2013). Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat. Biotechnol. *31*, 1023–1031.

Fukutomi, T., Takagi, K., Mizushima, T., Ohuchi, N., and Yamamoto, M. (2014). Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1. Mol. Cell. Biol. *34*, 832–846.

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.

George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., et al. (2015). Comprehensive genomic profiles of small cell lung cancer. Nature *524*, 47–53.

Goldstein, L.D., Cao, Y., Pau, G., Lawrence, M., Wu, T.D., Seshagiri, S., and Gentleman, R. (2016). Prediction and Quantification of Splice Events from RNA-Seq Data. PLoS ONE *11*, e0156132.

GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.

Hart, M., Concordet, J.P., Lassot, I., Albert, I., del los Santos, R., Durand, H., Perret, C., Rubinfeld, B., Margottin, F., Benarous, R., and Polakis, P. (1999). The F-box protein beta-TrCP associates with phosphorylated beta-catenin and regulates its activity in the cell. Curr. Biol. *9*, 207–210.

Hirotsu, Y., Katsuoka, F., Funayama, R., Nagashima, T., Nishida, Y., Nakayama, K., Engel, J.D., and Yamamoto, M. (2012). Nrf2-MafG heterodimers contribute globally to antioxidant and metabolic networks. Nucleic Acids Res. *40*, 10228–10239.

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.

Itoh, K., Wakabayashi, N., Katoh, Y., Ishii, T., Igarashi, K., Engel, J.D., and Yamamoto, M. (1999). Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. Genes Dev. *13*, 76–86.

Iwao, K., Nakamori, S., Kameyama, M., Imaoka, S., Kinoshita, M., Fukui, T., Ishiguro, S., Nakamura, Y., and Miyoshi, Y. (1998). Activation of the beta-catenin gene by interstitial deletions involving exon 3 in primary colorectal carcinomas without adenomatous polyposis coli mutations. Cancer Res. *58*, 1021–1026.

Kataoka, K., Nagata, Y., Kitanaka, A., Shiraishi, Y., Shimamura, T., Yasunaga, J., Totoki, Y., Chiba, K., Sato-Otsubo, A., Nagae, G., et al. (2015). Integrated molecular analysis of adult T cell leukemia/lymphoma. Nat. Genet. *47*, 1304–1315.

Katoh, Y., Itoh, K., Yoshida, E., Miyagishi, M., Fukamizu, A., and Yamamoto, M. (2001). Two domains of Nrf2 cooperatively bind CBP, a CREB binding protein, and synergistically activate transcription. Genes Cells *6*, 857–868.

Klijn, C., Durinck, S., Stawiski, E.W., Haverty, P.M., Jiang, Z., Liu, H., Degenhardt, J., Mayba, O., Gnad, F., Liu, J., et al. (2015). A comprehensive transcriptional portrait of human cancer cell lines. Nat. Biotechnol. *33*, 306–312.

Kong-Beltran, M., Seshagiri, S., Zha, J., Zhu, W., Bhawe, K., Mendoza, N., Holcomb, T., Pujara, K., Stinson, J., Fu, L., et al. (2006). Somatic mutations lead to an oncogenic deletion of met in lung cancer. Cancer Res. *66*, 283–289.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Leinonen, H.M., Kansanen, E., Pölönen, P., Heinäniemi, M., and Levonen, A.L. (2014). Role of the Keap1-Nrf2 pathway in cancer. Adv. Cancer Res. *122*, 281–320.

MacLeod, A.K., McMahon, M., Plummer, S.M., Higgins, L.G., Penning, T.M., Igarashi, K., and Hayes, J.D. (2009). Characterization of the cancer chemopreventive NRF2-dependent gene battery in human keratinocytes: demonstration that the KEAP1-NRF2 pathway, and not the BACH1-NRF2 pathway, controls cytoprotection against electrophiles as well as redox-cycling compounds. Carcinogenesis *30*, 1571–1580.

Malhotra, D., Portales-Casamar, E., Singh, A., Srivastava, S., Arenillas, D., Happel, C., Shyr, C., Wakabayashi, N., Kensler, T.W., Wasserman, W.W., and Biswal, S. (2010). Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through ChIP-seq profiling and network analysis. Nucleic Acids Res. *38*, 5718–5734.

McMahon, M., Thomas, N., Itoh, K., Yamamoto, M., and Hayes, J.D. (2006). Dimerization of substrate adaptors can facilitate cullin-mediated ubiquitylation of proteins by a "tethering" mechanism: a two-site interaction model for the Nrf2-Keap1 complex. J. Biol. Chem. *281*, 24756–24768.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *33*, D501–D504.

Riccardi, C., and Nicoletti, I. (2006). Analysis of apoptosis by propidium iodide staining and flow cytometry. Nat. Protoc. *1*, 1458–1461.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. *43*, D670–D681.

Sekine, H., Okazaki, K., Ota, N., Shima, H., Katoh, Y., Suzuki, N., Igarashi, K., Ito, M., Motohashi, H., and Yamamoto, M. (2015). The mediator subunit MED16 transduces NRF2-activating signals into antioxidant gene expression. Mol. Cell. Biol. *36*, 407–420.

Shibata, T., Ohta, T., Tong, K.I., Kokubu, A., Odogawa, R., Tsuta, K., Asamura, H., Yamamoto, M., and Hirohashi, S. (2008). Cancer related mutations in NRF2 impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. Proc. Natl. Acad. Sci. USA *105*, 13568–13573.

Tong, K.I., Katoh, Y., Kusunoki, H., Itoh, K., Tanaka, T., and Yamamoto, M. (2006a). Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model. Mol. Cell. Biol. *26*, 2887–2900.

Tong, K.I., Kobayashi, A., Katsuoka, F., and Yamamoto, M. (2006b). Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism. Biol. Chem. *387*, 1311–1320.

Vitari, A.C., Leong, K.G., Newton, K., Yee, C., O'Rourke, K., Liu, J., Phu, L., Vij, R., Ferrando, R., Couto, S.S., et al. (2011). COP1 is a tumour suppressor that causes degradation of ETS transcription factors. Nature *474*, 403–406.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873–881.

Zhang, J., Ohta, T., Maruyama, A., Hosoya, T., Nishikawa, K., Maher, J.M., Shibahara, S., Itoh, K., and Yamamoto, M. (2006). BRG1 interacts with Nrf2 to selectively mediate HO-1 induction in response to oxidative stress. Mol. Cell. Biol. *26*, 7942–7952.