A core NRF2 gene set defined through comprehensive transcriptomic analysis predicts selective drug resistance and poor multi-cancer prognosis

George Luo[1], Harshita Kumar[2], Kristin Aldridge[3], Stevie Rieger[4], Ethan Jiang[5], Ernest R. Chan[6], Ahmed Soliman[7], Haider Mahdi[8,9,10], and John J. Letterio[11,12,13]

1 Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, Ohio
2 Department of Engineering, Case Western Reserve University, Cleveland, Ohio
3 Triterpenoid Therapeutics Inc., Tunbridge Vermont
4 Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio
5 Booth School of Business, University of Chicago, Chicago, Illinois
6 Institute for Computational Biology, Case Western Reserve University, Cleveland, Ohio
7 Department of Pediatrics, SUNY Downstate Hospital, Brooklyn, New York
8 Magee Women's Research Institute, University of Pittsburgh Medical Center, Pittsburgh, PA
9 Magee Women's Hospital, University of Pittsburgh Medical Center, Pittsburgh, PA,
10 Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA,
11 The Angie Fowler Adolescent and Young Adult Cancer Institute, University Hospitals Rainbow Babies & Children's Hospital, Cleveland, Ohio
12 The Case Comprehensive Cancer Center, Cleveland, Ohio
13 Department of Pediatrics, Case Western Reserve University, Cleveland, Ohio

Corresponding Authors:
George Luo, Case Western Reserve University School of Medicine, 2103 Cornell Rd., Wolstein Research Bldg. Rm 3501, Cleveland, Ohio 44106; gxl263@case.edu

John J. Letterio, M.D., The Angie Fowler Adolescent and Young Adult Cancer Institute, University Hospitals Rainbow Babies & Children's Hospital, 11100 Euclid Avenue, Cleveland, Ohio 44106; jjl26@case.edu

Running Title: 'Novel NRF2 score predicts selective drug resistance and cancer prognosis'

Key Words: NRF2, Biomarker, Drug Resistance, Cancer Prognosis, Transcriptomics,

ABSTRACT

The NRF2-KEAP1 pathway plays an important role in the cellular response to oxidative stress but may also contribute to metabolic changes and drug resistance in cancer. We investigated the activation of NRF2 in human cancers and fibroblast cells through KEAP1 inhibition and cancer associated KEAP1/NRF2 mutations. We define a core set of 14 upregulated NRF2 target genes from seven RNA-Sequencing databases that we generated and analyzed, which we validated this gene set through analyses of published databases and gene sets. An NRF2 activity score based on expression of these core target genes correlates with resistance to drugs such as PX-12 and necrosulfonamide but not to paclitaxel or bardoxolone methyl. We validated these findings and also found NRF2 activation led to radioresistance in cancer cell lines. Finally, our NRF2 score is prognostic for cancer survival and validated in additional independent cohorts for novel cancers types not associated with NRF2-KEAP1 mutations. These analyses define a core NRF2 gene set that is robust, versatile, and useful as a NRF2 biomarker and for predicting drug resistance and cancer prognosis.

INTRODUCTION

Nuclear factor-erythroid 2-related factor 2 (*NFE2L2*, NRF2) is a transcription factor that plays a key role in the regulation of cellular defense mechanisms against oxidative stress and other forms of cellular damage [1]. NRF2 is responsible for the activation of a wide range of antioxidant and detoxifying enzymes, which help to protect cells from damage caused by reactive oxygen species (ROS) and other toxic molecules [1]. Kelch-like ECH-associated protein 1 (KEAP1) is an adaptor protein that acts as a negative regulator of NRF2 by targeting it for ubiquitination and degradation [1, 2]. In normal cells, KEAP1 acts to prevent NRF2 from becoming overactive and inducing excessive oxidative stress [1, 2]. However, in cancer cells and other disease states, the NRF2-KEAP1 pathway can become dysregulated, leading to the activation of NRF2 and the increased expression of its NRF2 target genes [2, 3]. The consequences of constitutive NRF2 activation in cancer include increased chemoresistance and a poor overall prognosis [2-5].

The NRF2-KEAP1 pathway is therefore a crucial component of cellular defense mechanisms, and its dysregulation can play a major role in the development of diseases such as cancer, neurodegeneration, and inflammation [5-8]. In recent years, the NRF2-KEAP1 pathway has become a target for the development of new drugs for the treatment of these and other diseases [9, 10]. Two examples of therapeutics that directly target the NRF2-KEAP1 pathway include dimethyl fumarate for multiple sclerosis and omaveloxone for Frederich's ataxia [11-13]. Thus, a deeper understanding of the NRF2-KEAP1 pathway and its role in cellular biology is of significant interest to both basic science and

clinical researchers. However, despite the interest and attention of this pathway, there are few reliable biomarkers that can be used to measure NRF2 activity for the wide range of context in which it is being researched [14, 15].

For biomarkers of NRF2 activity, researchers have used various genes to evaluate NRF2 activity but there is little consensus despite decades of research on NRF2 signaling and target genes [12-16]. Yagishita *et al.* and Kopacz *et al.* both report the common usage of *NQO1* as a proxy for NRF2 signaling in research and clinical trials but secondary options vary from study to study, including *HMOX1*, *AKR1C1*, *GSTM1*, and *GSTP1* [12-17]. While *NQO1* may be a great first option, it would be wise to have other confirmatory options to use as well as know analytically what are the most representative NRF2 target gene that be used under any context. Kopacz *et al.* details caveats of sole use of *NQO1* or *HMOX1* as NRF2 biomarkers [17]. After all, many genes may be regulated by various factors so by having multiple genes, the result is less likely to be affected by other signals and noise [18].

Here we first address this issue by defining a core set of NRF2 target genes that faithfully represents NRF2 activation in a wide array of contexts through analyses of a variety of cancer cell types and normal cells, as well as both pharmacological and genetic activation of NRF2. Generating these databases allowed us to find the most universally upregulated target genes of NRF2 while also enabling us to compare differences between the activation of NRF2 in these various contexts. We define a list of 14 core upregulated target genes of NRF2 which are present in every database that we generated, which we validate with published NRF2 datasets and compare with prior publications. This analysis allowed us to use this gene set as a biomarker and representation of NRF2 activation.

Using these core target genes, we generate a NRF2 change score which is useful for identifying both which and to what degree specific genetic modifications and drugs activate NRF2. A similar but distinct NRF2 activity score can predict which drugs are affected by NRF2 activation and which cell lines are most resistant to these drugs. We validate the modeling of selective drug resistance in our own cell lines. Finally, we investigated whether this NRF2 activity score is associated with prognosis in a variety of cancers. We trained the NRF2 activity score in 22 The Cancer Genome Atlas (TCGA) cancers [19], and found that it passed training in twelve cancers. We were then able to test validation in eight of these twelve cancers with independent cohorts and demonstrated that four cancers have significantly poorer prognosis associated with NRF2 activity, as defined by our NRF2 activity score. Overall, we identify a robust and reliable NRF2 signature that can be used in a wide range of contexts including predicting NRF2-associated drug resistance and as a biomarker of a poor cancer prognosis.

RESULTS

Generation of Seven Transcriptomic Databases of NRF2 Activation

In order to determine the core target genes of NRF2, we investigated gene expression changes from both genetic and pharmacological targeting of the NRF2/KEAP1 pathway (Fig. 1A). We treated myeloma (RPMI-8226), ovarian (OVCAR-8), glioma (SF8286), and primary dermal fibroblast cells with an NRF2 activator [20], the synthetic oleanane triterpenoid CDDO-2P-Im (1-[2-Cyano-3,12-dioxooleana-1,9(11)-dien-28-oyl]-4(-pyridin-2-yl)-1H-imidazole), for six hours before extracting RNA for RNA-sequencing. We then confirmed activation of NRF2 through upregulation of *NQO1*, a well-

established target of NRF2 [1, 2], in each of these cells (Fig. 1B-E). We also used CRISPR-Cas9 to knock out *KEAP1* in ARH-77, a B-cell lymphoblastic cell line, and verified the *KEAP1* deletion at the DNA, RNA, and protein levels (Fig. 1F-H). Furthermore, NQO1 was upregulated at baseline compared to the wild-type (WT) ARH-77 cells (Fig. 1I and J).

We also leveraged the available public databases of the Cancer Cell Line Encyclopedia (CCLE) [21] and TCGA tumor samples [19] to explore activation of NRF2. Approximately 20% of non-small lung cancer cells (NSCLC) have mutations in *KEAP1* or *NFE2L2* [5], which was also true in these datasets. After separating TCGA lung adenocarcinoma (LUAD) into WT and KEAP1/NRF2 mutated samples, we investigated the top differentially expressed genes (DEGs) of which *NQO1* was among the top genes identified (Fig. 1K). We repeated the analysis for NSCLC cell lines found in the CCLE on DepMap [22] and verified *NQO1* was upregulated in the KEAP1/NRF2 mutated group (Fig. 1L). Furthermore, we confirmed ROS pathway activation, a key NRF2 related pathway, through gene set enrichment analysis (GSEA) for our RNA-Sequencing generated datasets (Supplemental Fig. S1). These datasets all demonstrate NRF2 activation in different cellular and molecular contexts which enables us to develop a more accurate and comprehensive perspective of NRF2 activation.

A Core NRF2 Transcriptional Signature is Present after NRF2 Activation

After we created different databases of NRF2 activation, we investigated the most frequently upregulated targets of NRF2 activation, a list of 15 candidate core NRF2 genes (Fig. 2A, Supplemental Table S1). Additionally, many other genes were upregulated in most scenarios of NRF2 activation (Fig. 2B, Supplemental Table S2-10). These candidate

core NRF2 genes such as *NQO1*, *GCLC*, *GCLM*, and *SRXN1* have been well-documented as NRF2 target genes in the literature and CHIP-Seq datasets (Supplemental Fig. S2). Interestingly, we noticed most of these core genes, such as *ABHD4* and *SRXN1,* showed similar upregulation across all cell types and methods of NRF2 activation (Fig. 2C-D, and Supplemental Fig. S3). However, *AKR1C3* and other members of the *AKR1C* family displayed considerably higher activation in genetic models of NRF2 activation compared to short term drug-induced activation (Fig. 2E, Supplemental Fig. S4, and Supplemental Table S1-3). Another interesting observation is that higher concentrations of drug does not always equal higher activation of NRF2 target genes. This appears to be somewhat cell line dependent and gene specific as other core NRF2 genes show elevated upregulation at high doses (Supplemental Fig. S3J-K). We also used an *in silico* model of motif binding, LASAGNA-Search 2.0, to identify NRF2-ARE (Antioxidant Response Element) sequences near the promoters of these core and nearly universal NRF2 genes (Fig. 2F and Supplemental Fig. S5A). This analysis combined with previously published NRF2 CHIP-Seq datasets suggests that NRF2 directly binds to promoters of these core genes and directly upregulated transcription, supporting their positions within an initial list of candidate core NRF2 target genes (Supplemental Fig. S2).

Defining Conditionally Upregulated and Novels Target Genes of NRF2 Activation

Besides finding the most frequently upregulated genes, we examined if certain genes were specifically activated in one condition or another (Fig. 3A). We found certain genes such as *HMOX1* and *AKIRIN2* which are primarily upregulated in drug induced NRF2 activation but not present in either the TCGA LUAD or NSCLC cell lines (Fig. 3B

and C), consistent with the literature [23, 24]. On the other hand, genes related to enzymes in the pentose phosphate pathway (*G6PD*, *TKT*, *TALDO1*, and *PGD*) were consistently upregulated in the genetically activated NRF2 models but infrequently in the drug-induced models (Fig. 3D-G). However, this may be due to the relatively short duration of drug exposure. The pentose phosphate pathway is of particular importance as it is thought to help protect cells from oxidative stress and apoptosis [25].

We also uncovered two putative novel NRF2 target genes that are upregulated in most of our models, *B4GALNT1* and *POPDC3* (Fig. 3H-I). We found a putative NRF2/ARE motif near the promoter of B4GALNT1 though not for POPDC3 (Supplemental Fig. S6E). Finally, we also found novel downregulated genes (*PLAT, CDC42EP1, EVA1A)* in our models of NRF2 activation but these genes were not always downregulated or expressed (Supplemental Fig. S6B-D). NRF2 has been reported to interfere with transcriptional regulation of proinflammatory cytokines such as IL-6 and IL-1B [26], but we did not investigate immune cells in these experiments. Nevertheless, we observed decreased expression of immune related genes such as *CXCL6* and *CXCL1* in TCGA LUAD and KEAP1 knockout ARH-77 cells supporting the anti-inflammatory effects of NRF2 activation (Supplemental Fig. S7A-D). Because synthetic triterpenoids such as CDDO-2P-Im also inhibit NF-κB [27], it is difficult to comment on the effects of drug-induced NRF2 activation on inflammation, although suppression of inflammatory markers is observed.

Validation of the Core NRF2 Gene Set in Other Published Datasets

Although we believe our core NRF2 gene set can faithfully represent NRF2 activity, we needed to validate its utility in other contexts of NRF2 activation. For this, we first

tested other TCGA cancers which had considerable NRF2/KEAP1 mutations. We found lung squamous cell carcinoma (LUSC), head and neck squamous cell carcinoma (HNSC) and liver hepatocellular carcinoma (LIHC) to have significant numbers of mutations in NRF2/KEAP1. We found that both LUSC and HNSC validated all our core and nearly universal NRF2 gene set (Fig. 4A-B and Supplemental Fig. S8A-B). However, LIHC only validated 13/15 core NRF2 genes and 10/16 of the nearly universal NRF2 gene set (Fig. 4C and Supplemental Fig. S8C), though LIHC also had a lower number of NRF2/KEAP1 mutations.

We selected two published studies [28, 29] to investigate the validity our NRF2 gene set in other datasets and publications. First, in three different prostate cell lines treated with sulforaphane by Beaver *et al*., we found upregulation of 14/15 of the candidate core NRF2 gene and 10/16 of nearly universal NRF2 genes (Fig. 4D and Supplemental Fig. S8D) . *CBR3* was the only NRF2 gene in our initial core set that did not pass multiple validation sets so we chose to omit *CBR3* from the final core NRF2 biomarker gene set. Overall, these two different validations strongly demonstrate the validity and versatility of our 14 core NRF2 gene set in human tissues.

In the rat model of NRF2 gene deletion by Taguki *et al*., two different germline NRF2 KO mutations were generated and used to investigate NRF2 dependent gene expression changes [29]. Mice were divided into the following groups and treated with vehicle or NRF2 activator, CDDO-Im: WT + vehicle, WT + CDDO-IM, NRF2 KO + vehicle, NRF2 KO + CDDO-Im. We found that the majority of core NRF2 target genes were upregulated by CDDO-Im treatment but unchanged in the setting of NRF2 gene deletion (Fig. 4E and Supplemental Fig. S8E). However, other genes (*OSGIN1, SLC7A11, ME1,*

FTH1, FTL) showed a similar trend of activation following CDDO-Im treatment but did not reach significance in every case. Finally, we observed that *AKR1C3* did not display any trend suggesting NRF2 dependency which is similar to literature reports that AKR1C genes are not homologous between human and mice [30]. We also performed analysis with the nearly universal NRF2 gene set, finding most exhibited a trend suggesting NRF2 dependency (Supplemental Fig. S8F and G). Therefore, we believe our core NRF2 gene can also work for rodent models, with a few adjustments for genes known to be nonhomologous.

Finally, we compared our core NRF2 gene set with previously published gene sets from six studies [23, 31-35]. Overall, every single core NRF2 gene was a part of previous NRF2 gene sets with 8 genes (*GCLM, SRXN1, GCLC, GSR, NQO1, SLC7A11, AKR1C3, ME1)* that were a part of 4 of more out of the 6 NRF2 gene sets examined (Fig 4F). Despite our core gene set having the fewest number of genes of these gene set, the majority of overlapping genes from at least 4 of 6 previous NRF2 gene sets were present in our core gene set (Supplemental Table S11). Our nearly universal NRF2 gene set contain a mix of genes that were present in multiple previous gene sets, only one previous gene set, or not found in any of these previous gene sets (Supplemental Fig. S8H). Thus, our core NRF2 gene set contains most of the common features of previous NRF2 gene sets while being fewer but still applicable to the many settings examined.

Next, we introduce the NRF2 Change Score (N2CS), which is a useful measure of drug or genetic effects on NRF2 activity. Using gene expression changes of the 14 core NRF2 genes, we can approximate the degree to which NRF2 is changed with scores above 1 suggesting high activation with a score of 0 as the baseline (See Methods for

calculations). We demonstrate the utility of the N2CS in the various conditions in which we examined the response to CDDO-2P-Im in our study, in the KEAP1 knockout, in the response to sulforaphane from validation studies, as well as in the response to other anti-cancer agents from published datasets (Fig. 4G and H). Interestingly, we also noticed that higher concentrations of CDDO-2P-Im do not further activate NRF2 which is also the case with other NRF2 activators [36]. We found that both bortezomib and doxorubicin induce NRF2 activation (Fig. 4H), consistent with literature [37, 38], but paclitaxel does not significantly affect NRF2 activity. Thus, the N2CS can be used to screen or check RNA-Sequencing data for modulators of NRF2 activity.

NRF2 Activity Score Correlates with Resistance to Radiation and Certain Drugs

Using the final core NRF2 target genes, we can also generate NRF2 activity scores (N2AS) to represent the degree of NRF2 activation for individual cell lines and tumor samples. The N2AS is useful for comparing cells or samples of the same group for differences in NRF2 activation. In the following experiments, we demonstrate the utility of the N2AS.

The N2AS can be used to find whether specific non-conservative mutations in NRF2/KEAP1 affect NRF2 activity. We first used N2AS to rank all CCLE NSCLC cell lines in DepMap and investigated how mutation status of NRF2/KEAP1 was related to N2AS (Fig. 5A). While most mutations are associated with higher N2AS, we do observe several mutations are associated with medium or low N2AS, suggesting these mutations do not activate NRF2. Furthermore, we correlated N2AS with a functional output, sensitivity of oxidative stress-mediated drugs from the DepMap database (106). We found N2AS

significantly correlated positively with resistance to PX-12 and necrosulfonamide in NSCLC cells (Fig. 5B and C). Not surprisingly, PX-12 has been previously demonstrated to induce apoptosis in an oxidative stress dependent manner [39] while necrosulfonamide has been linked to reacting with cystines previously [40]. These two drugs were among the top drug correlations with N2AS and we further validated this observation in our KEAP1 knockout cell line (Fig. 5D and E, Supplemental Table S12). We also tested correlation with individual NRF2 core genes and our gene set had a higher correlation than any single gene and as high correlation as other published NRF2 gene sets (Supplemental Fig. S9A-B). Although radiation is not in the database, we also tested if NRF2 activation would lead to resistance against radiation (Fig. 5F). Our findings confirmed radioresistance associated with NRF2 activation, and other studies report similar findings [41]. Interestingly, we also observe enhanced colony formation with KEAP1 knockout in ARH-77 cells (Supplemental Fig. S9C-E), a cell line with poor colony formation efficiency [42].

We also investigated drugs that were not correlated with the N2AS (Supplemental Table S12). We found both paclitaxel and bardoxolone methyl (CDDO-Me), a synthetic oleanane triterpenoid, to have little correlation with N2AS (Fig 5G and H). We then tested both drugs in our cell lines and found the resistance of KEAP1 knockout cells to paclitaxel and bardoxolone methyl was absent in comparison to resistance to both PX-12 and necrosulfonamide (Fig 5I and J). This suggests that our modeling of NRF2 activity and drug resistance is appropriate and can predict relative resistance mediated by NRF2 for drugs in the DepMap CTRP2 database.

NRF2 Activity is Associated with Poor Prognosis in a Variety of Cancers

Although there have been many studies linking mutations causing NRF2 activations to chemoresistance and poor clinical outcomes, many of these previous studies were concentrated on advanced stage patients with small sample sizes [43, 44]. Recently, a large study found KEAP1 mutations were associated with a worse prognosis, confirming the clinical importance of NRF2 activation in NSCLC [45]. However, despite this finding, mutation stratification alone is not significantly associated with prognosis difference in TCGA LUAD or LUSC datasets (Supplemental Fig. S10A and B). We believe this observation can be explained by our analyses of cell lines and truly demonstrates that not all mutations of NRF2/KEAP1 are equal and should be evaluated individually.

By using the N2AS to help categorize patients, we investigated whether our NRF2 activity score had prognostic significance in 22 different TCGA cancers (Fig. 6A). During the training, we focused on finding an appropriate N2AS cutoff for each cancer and a highly significant p value ($p < 0.01$). We found 12 cancers passed training and 10 cancers did not pass training which suggests these 10 cancers do not have significant prognosis associated with NRF2 (Supplemental Fig. S11). Of the 12 cancers that have potential prognosis significance associated with NRF2 activity, 10 had a poor prognosis associated with higher NRF2 activity score while Uterine Corpus Endometrial (UCEC) and Ovarian cancers had favorable prognosis associated with higher NRF2 activity (Fig. 6B, Supplemental Fig. S12A-D). Under further examination, PTEN mutation is a common feature in UCEC and is associated with both NRF2 activation and a favorable prognosis (Supplemental Fig. S12C) (34,35). PTEN inactivation can directly lead to higher NRF2

activation and thus may be simultaneously driving both NRF2 activity and a favorable prognosis.

To help analyze the categorization and significance of the N2AS in training and validation, we developed LOCC (Luo's Optimization Categorization Curve) which visualizes all hazard ratios (HR) and p-values of all possible cutoffs [46]. This method allows us to understand the selection of the cutoff while also understanding if there are other possible windows for categorization. Using this method on LUAD, we first locate cutoffs that are highly significant, where the yellow line, representing the –log (p value), is above the red line (representing a p value of 0.01) (Supplemental Fig. S13A). At this cutoff, the black line reveals the HR (the red line simultaneously represents an HR of 1). The x-axes of both graphs are lined up such that the sample distribution of N2AS can be visualized. The highest yellow peak is the most significant p value and is in most cases, the ideal cutoff point. After choosing a N2AS cutoff of 0.463 for LUAD, the N2AS is finalized for the training and a Kaplan-Meier Curve can be generated with appropriate additional calculations (HR – 1.63, $p < 0.01$, Fig 6C).

Using published East Asian cohort of LUAD (OncoSG) [47], we applied the previous N2AS cutoff to test if it can stratify patients into prognostic significant groups (Supplemental Fig. S13B). Indeed, we found higher NRF2 activity was associated with a worse prognosis (HR – 2.60, p – 0.01, Fig 6D). Using LOCC [46], we observed the cutoff was highly significant and that the cutoff was appropriate despite using a very different patient population as well as a drastically lower percentage of patients with NRF2/KEAP1 mutation (Supplemental Fig. S10E). These results demonstrate that N2AS could be used

as a prognostic biomarker for LUAD in various cohorts even without the presence of NRF2 mutations.

Following the validation for LUAD, we used other public databases to validate N2AS prognostic values for other cancers (Supplemental Fig. S13C-H). Through the same methods of optimizing cutoffs, we were able to successfully validate N2AS prognostic value in renal clear cell carcinoma (KIRC, Fig. 6E and F), liver hepatocellular carcinoma (LIHC, Fig. 6G and H), and acute myeloid leukemia (LAML, Fig. 6I and J). For all three cancers, we found high NRF2 activity was associated with lower overall survival (KIRC HR – 2.12, $p < 0.001$, LIHC – 2.16, $p < 0.001$, LAML – HR 3.87, p - 0.001). Previous studies in small cohorts have found high NRF2 levels is associated with poor prognosis in liver and renal cancers but have not been validated in larger cohorts using gene expression data [48, 49].

We also tested validation in head and neck squamous cell carcinoma (HNSC), pancreatic adenocarcinoma (PAAD), melanoma (SKCM), and bladder carcinoma (BLCA). Although these cancers passed training, they showed no significant poor prognosis correlated with N2AS in the validation cohort (Supplemental Fig. S14). For other TCGA cancers that passed training but did not have validation, it was because there was not a large suitable cohort to validate the training. Finally, we used other NRF2 gene sets to compare if we would obtain similar prognostic results. We found that other NRF2 gene sets could obtain similar results in LUAD but not LAML (Supplemental Fig. S15), showing that our core NRF2 gene set is better suited toward a broader range of applications than those selected gene markers of NRF2 activity found in the existing literature.

Discussion

The KEAP1/NRF2 pathway plays essential but complex roles in the maintenance of metabolic homeostasis, regulation of the cellular redox state and in the modulation of the inflammatory response [1, 6, 7]. These activities contribute to neuroprotection in the context of neurodegenerative disease [50], and to the suppression of other chronic diseases affecting multiple organs including the liver and lung [51, 52]. However, increasing evidence linking constitutive NRF2 activity to therapy resistance in cancer highlight the challenges that have influenced the clinical development of activators of this pathway [15]. In our study, we define a core NRF2 target gene set as well as several databases of NRF2 activations that will aid preclinical efforts designed to clarify the context-dependent roles of NRF2 in health and disease and to define appropriate clinical indications for therapeutics targeting the KEAP1/NRF2 complex. We rigorously applied our core NRF2 gene set through NRF2 activation validation studies, drug resistance modeling with further validation, and finally cancer prognosis with validation in multiple patient cohorts. During this process, we also utilized other published gene sets for comparison and found that while some could perform in certain situations, none were as useful for clinical prognosis. This discrepancy is likely because of the high noise from conditional or non-NRF2 target genes. While our core NRF2 gene set may not be perfect for all applications, it is a solid foundation as a biomarker that can be used in a variety of settings for the assessment of NRF2 activity. Subsets of the core NRF2 signature may be used as biomarkers of NRF2 activity in research and clinical assessment.

While cancer research has become increasingly focused on defining the unique genetic alterations that characterize distinct and often rare tumor histology, we believe it

is imperative to understand the commonality between groups as the latter may be informative regarding therapeutic resistance. Although NRF2 has been studied in many cancer contexts, particularly in NSCLC and HNSC, we observe the same principal target genes are upregulated in nearly all human cancer cells and that drug resistance modeled in NSCLC can be validated in a B cell model. Furthermore, our investigation of cancer prognosis demonstrates that in our successful training and validation, the trend is that high NRF2 activity is associated with a poor prognosis. As such, our understanding of NRF2 targets and drug resistance may not be limited to the cancers we validated but to many others which may have a larger impact on our understanding of therapeutic efficacy and resistance. Furthermore, the literature of chemoresistance mediated by NRF2 not only directly applies to cancer resistance but also to prevention of toxicities by these agents [53]. Therefore, a broad application of precision medicine and NRF2 targeting can be investigated in the future.

The literature has described many instances of chemoresistance by NRF2 [54, 55], but we demonstrate that not all chemoresistance by NRF2 activation is comparable. Furthermore, strategies designed to repress, knock down, or completely delete NRF2 gene expression may have very different effects on drug resistance when compared to activation of NRF2. Moreover, if NRF2 acts to protect both normal and transformed cells, it will be imperative to design strategies that obviate the effect of NRF2 specifically in tumor cells, as global repression of NRF2 will most certainly increase side effects and toxicities associated with chemotherapy and. radiation therapy. Interestingly, the pentose phosphate pathway activated by constitutive NRF2 activation may be a targetable option. Multiple studies have demonstrated CB-839 (Telaglenastat), a glutaminase inhibitor, can

inhibit cancer cells with high NRF2 activity better than those without high NRF2 activity [54, 56-58]. While this may represent an exploitable 'flaw in the protective armor' that NRF2 provides cancer cells, in the absence of a clear NRF2 biomarker, this may not have a large clinical benefit. Thus, it is not shocking that CB-839 recently failed a clinical trial in renal cell carcinoma, a cancer in which NRF2 activation may play a significant role in prognosis, given the lack of a clear biomarker to select patients for whom the drug would have potential clinical benefit [59, 60].

There are still many important questions to address before an NRF2 activity score can be used as an indicator for therapy. Many anti-cancer drugs can activate NRF2 through ROS generation, but this acute NRF2 activation is clearly different from the constitutive NRF2 activation observed with genetic mutations. We showed a short-term exposure to NRF2 activators will upregulate many core NRF2 targets similarly but other genes such as *AKR1C*, *G6PD*, *PGD*, *TKT*, *TALDO1* exhibit much stronger upregulation in the genetic models. Furthermore, synthetic oleaneane triterpenoids (SOTs), which activate NRF2 but also have anti-cancer properties [61], synergize with several anti-cancer agents [62, 63]. The latter likely reflects the fact that these SOTs are multi-targeted agents, modulating the activity of NFκB, STAT3, mTOR and other key regulators of cancer cell growth and viability [61]. Given the likelihood that individual NRF2 target genes may exhibit time-dependent and bi-modal regulation, it is conceivable that therapy might be tailored such that different drugs regimens that exploit low NRF2 activity with oxidative stress-mediated therapy (drugs that have resistance correlated with N2AS) may be followed by therapeutics that are less affected by N2AS.

Computational biology has been a field of great promise and interest but also presents many questions and problems as well [64]. Despite knowing the mutations of cancer cells in the CCLE and tumor samples from TCGA, we can find the general differences, yet it is hard to estimate the effect of specific unique mutations. Here we demonstrate that a NRF2 gene expression signature can help differentiate NRF2 activity in both cell lines and patient samples. However, instead of traditional methods of selecting cutoffs such as median or quartiles, we chose a training method to find the optimal cutoff and employed validation methodology which has been implemented by other studies [65, 66]. Through this novel approach, a visualization of the cutoff selection process was implemented to facilitate analysis and an enhanced understanding of the entire picture of prognosis predicted by this gene set (LOCC) [46]. Previous studies have used the receiver operating characterization (ROC) curve to help continuous variable cutoffs validation despite its better applicability to diagnosis as opposed to prognosis [66]. However, the ROC does not help us understand why biomarkers fail or succeed in validation and is suited for large differences in odds or hazard ratio [67]. With LOCC [46], there is an abundance of information which can be seen and then calculated for comparison across different genes or gene sets. When validation is unsuccessful or partly successful, the reason can be seen: e.g., no peak throughout, correctly identified peak but not enough significance or incorrect cutoff leading to low significance. We believe machine learning is the future of computational biology, but we can also supervise and understand the process which will lead to less misinterpretations and more awareness of the validations and failures.

Our N2AS is a simple unweighted average of the z-scores of the core NRF2 target genes, which is both foundational and imperfect. We tried to be as comprehensive and unbiased in our searching for core NRF2 target genes and modeling of N2AS activity to find results. Even though we did not optimize the N2AS for any application, we were still able to find and validate significant results from drug resistance to cancer prognosis, suggesting that this pathway is critical for these outcomes. With a clear direction, we can optimize the analyses for each cancer with slightly different NRF2 target genes and used weighted scoring knowing that our fundamental goal is still the quantitation of NRF2 activation, which is validated. If we optimize using machine learning without knowing whether NRF2 activation is important, we may find significant differences, yet it may not be directly related to NRF2 activity, as many genes exhibit complicated transcriptional regulation orchestrated by distinct activators and repressors acting in concert. Therefore, our simplistic N2AS is important but we understand that future research should investigate appropriate optimization with regards to each cancer's context including population differences and multivariate analysis.

Finally, we ponder why a universal pathway that directly affects chemoresistance such as NRF2/KEAP1 is prognostic in only some cancers and not others. Even though our results are novel in finding NRF2 is prognostic for large cohorts of AML, LIHC, KIRC, and possibly other cancers, there are still many questions as to why we observed this effect in these cancers specifically. For most of the cancers which were either validated or promising candidates, we found a skewed distribution in which NRF2 activity was very high in certain patient samples. This also correlated with high numbers of mutations in those cancers related to NRF2/KEAP1. Yet, for LUSC, we do not observe any prognostic

significance, similar to what has been shown in other studies with NRF2 in TCGA LUSC [68]. However, most cancers do not have a skewed distribution likely owing to fewer mutations that dramatically activate NRF2. Furthermore, while not all cancers rely on oxidative stress-dependent mechanisms, if therapy is dependent on oxidative stress or impaired by NRF2 activity, then surviving cancer cells would likely have a selection pressure favoring NRF2 activation. However, such an escape mechanism may not be found using bulk RNA-seq and even be difficult to discern with single cell RNA-seq. It may be discovered through sequential RNA-Sequencing of initial and relapsed tumor samples in therapies that significantly favor NRF2 activation, a phenomenon which has been demonstrated in breast cancer [69]. As such, there is still a need for research to be done in many cancers regarding the potential role of NRF2 in affecting outcomes.

Methods

Cell lines

SF8628 (RRID:CVCL_IT46), OVCAR-8 (RRID:CVCL_1629), RPMI-8226 (RRID:CVCL_0014), ARH-77(RRID:CVCL_1072), and primary dermal fibroblast cells were all purchased from American Type Culture Collection (ATCC). Cells were tested negative for mycoplasma with ABM Mycoplasma PCR detection kit (catalog # G238) in February 2023. Cancer cell lines were authenticated using STR profiling at the CWRU genomic facility.

Cell Treatment for RNA Analysis

Cells were treated either with a control of media plus DMSO (1%) or with CDDO-2P-Im (gift from Dr. Michael Sporn, 96.3% purity by HPLC) for 6 hours in 6 well plates before RNA extraction at concentrations indicated in Supplemental Table S13. RNA was measured by Nanodrop (Thermo Fisher Scientific) and then sent to a sequencing service, LC Sciences, for RNA-Sequencing or prepared for cDNA synthesis (#4368813, Thermo Fisher Scientific).

RNA-Sequencing

A poly(A) RNA sequencing library was prepared following Illumina's TruSeq-stranded-mRNA sample preparation protocol as previously described [63]. RNA integrity was checked with Agilent Technologies 2100 Bioanalyzer. Poly(A) tail-containing mRNAs were purified using oligo-(dT) magnetic beads with two rounds of purification. After purification, poly(A) RNA was fragmented using divalent cation buffer in elevated temperature. Quality control analysis and quantification of the sequencing library were performed using Agilent Technologies 2100 Bioanalyzer High Sensitivity DNA Chip. Paired-ended sequencing was performed on Illumina's NovaSeq 6000 sequencing system.

Cutadapt was used to remove the reads that contained contamination, low quality and undetermined bases [70]. Then sequence quality was verified with FastQC (RRID:SCR_014583). HISAT2 was used to map reads to the genome of ftp://ftp.ensembl.org/pub/release-96/fasta/homo_sapiens/dna/ [71]. The mapped reads were assembled using StringTie [72]. The transcriptomes were merged to reconstruct a

comprehensive transcriptome using perl scripts and gffcompare. After the final transcriptome was generated, StringTie and edgeR were used to estimate the expression levels of all transcripts [72, 73]. Additional RNA-Sequencing analysis was performed using Gene Set Enrichment Analysis (GSEA, RRID:SCR_003199). RNA-Sequencing has been deposited to GEO (GSE228434, GSE221526, GSE229914, GSE229704, GSE229652).

Immunoblotting

Cells were lysed with cell lysis solution (#9803, Cell Signaling, MA, USA) with protease and phosphatase inhibitor (A32961, Thermo Fisher Scientific) as previously published [63]. Cell lysates were measured using bicinchoninic acid assay to measure protein concentrations and equal amounts of proteins were mixed with Nupage LDS sample buffer (NP0007, Thermo Fisher Scientific) according to manufacturer's instructions. Samples were loaded into sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel, and then transferred to PVDF membranes (Catalog #IPFL00010 Millipore, MA, USA), and incubated with blocking buffer (927-60001, Li-cor). Membranes were incubated with primary antibodies overnight at 4ºC, washed 3 times with Tris-buffered saline with 0.05% Tween-20, and then with the corresponding secondary antibody for 1 hour at room temperature. Membranes were washed 3 times with Tris-buffered saline with 0.05% Tween-20. Finally, membranes were visualized with Li-Cor Odyssey DLx using Image Studio. Antibodies against KEAP1 (Thermo Fisher Scientific, #10503-2-AP, RRID: AB_2132625), NQO1 (Thermo Fisher Scientific, #11451-1-AP, RRID:AB_2298729), β-actin (66009-1-Ig, Proteintech, RRID:AB_2687938) and GAPDH (sc-32233, Santa Cruz Biotechnology, RRID:AB_627679) were used.

KEAP1 Knockout Generation

ARH-77 with deletion of the genes encoding KEAP1 was generated through CRISPR-Cas9 deletion. Cells were electroporated in R Buffer (Thermo Fisher Scientific, BR5) containing 150 ng/µl sgRNA and 500 ng/ul recombinant TrueCut Cas9 protein (Thermo Fisher, A36496) using Invitrogen Neon Electroporator with 2 pulses of 1200 V, 2 ms duration. Three sgRNA sequences, 5' CCGUGUAGGCGAAUUCAAUG 3', 5' ACCAACGGGCUGCGGGAGCA 3', and 5' UGGGCCAUGAACUGGGCGGC 3', were used for electroporation. Cells were then seeded at a density of one cell per well in 96 well plates and grown until colonies were established. Colonies were tested for DNA changes through PCR using KEAP1 primers, 5' AGGTGCTGGCCCAGTCCCAA 3' and 5' GCCGATGGCATTGCTGGGGT 3'. RNA was extracted from cells and tested for *NQO1* (Thermo Fisher Scientific, Hs01045993_g1) and *KEAP1* (Thermo Fisher Scientific, Hs01003506 _m1) expression with *GAPDH* normalization (Hs02758991_g1).

RNA-Seq Dataset Analysis

RNA-Sequencing data for each sample in each cell line was analyzed in Matlab (RRID:SCR_001622). Data were processed such that any genes with average expression values less than 1 fragments per kilobase of exon per million mapped fragments (FPKM) for all concentrations were removed. Mattest was used to conduct an unpaired t-test. PCA (principal component analysis) helped generate principal component coefficients to quantify the two most important principal components, which could then be visualized. The visualization helped determine the exclusion of sample 1 in control RPMI-8226, and sample 3 in the wildtype ARH-77 (Supplemental Fig. S14). Validation data from 24 hour

treatment of sulforaphane was used from GSE48812. Raw data was processed into FPKM as previously described and then Transcripts Per Millions for analysis[63]. Non-small lung cancer cells data from Depmap and cancer patient TCGA data from cBioportal, [74], were analyzed in R. Expression of cell lines from Depmap were processed in $\log_2$(Transcripts Per Million + 1) while TCGA expression used $\log_2$(normalized counts + 1). Normalized counts were calculated using RSEM (batch normalized) and downloaded from cbioportal. Cell lines and tumor samples were classified as NRF2/KEAP1 wildtype or mutant if non-conservative mutations were present in the DNA sequencing data. RNA expression was compared between wildtype and knockout samples for differential expression. Gene set enrichment analysis (Broad Institute Software) was performed on RNA-sequencing data from cell lines.

NRF2 Target Genes Identification

After processing, genes that were upregulated more than 25% on average from DMSO to low CDDO-2P-Im dose were identified as candidate NRF2 upregulated genes. A two-way student t-test using unpaired unequal variance was used to calculate a P value for each comparison. Genes that had upregulation and a p value of less than 0.1 (one way t-test of 0.05) were considered candidate NRF2 target genes for each cell line. A similar classification was used for the KEAP1 KO ARH-77 model between WT and KO samples. Data from sulforaphane validation was evaluated with the same criteria of upregulation of more than 25% on average and a t-test p value of less than 0.1.

NRF2-induced genes in LUAD TCGA samples, LUSC TCGA samples, HNSC TCGA samples, LIHC TCGA samples, from the cBioportal, and the NSCLC CCLE

samples, from the DepMap, were also found by selecting the top 500 upregulated genes in each sample, ordered by lowest p value. Bar graphs were generated for individual cell line RNA-Sequencing data while box plots using quartiles were generated for comparison of multiple cell lines or tumor samples.

These lists of candidate NRF2 target genes were compared to each other to determine which genes that were induced across cell lines. Genes were determined to be highly upregulated if their expression level increased two-fold after drug treatment, following genetic modification, or if they were part of the top 100 most significant upregulated genes by p value in the TCGA or CCLE databases. A similar analysis was repeated for downregulated genes for each cell line, in which all genes with a significant (p-value < 0.1) reduction in expression were compared with no minimum threshold for downregulation. In the TCGA and CCLE datasets, the top 250 downregulated genes were chosen as targets.

For the individual NRF2 target genes, LASAGNA-search 2.0 tool [75] was used to predict *NFE2L2*/ARE binding sites. For each gene, a distance of 0 to -5000 of the transcription start site (TSS) was selected and a cutoff of 0.0005 p value was used. Binding sites were numbered according to lowest estimated p value.

Validation of NRF2 Genes in Rat Models

Gene expression data was used from GSE77377 to investigate changes in NRF2 related genes [29]. Gene expression data were averaged for each treatment group and normalized to WT with vehicle treatment. P values for gene expression changes were calculated using student t-test.

N2CS and N2AS Calculation

To calculate the N2CS and N2AS, the 14 core NRF2 biomarker gene expressions were used. Gene expression was standardized using FPKM for each set of calculations. For the N2CS, the control is set as the baseline and all gene expression changes were expressed as Log_2 Fold Change relative to the baseline. The fold changes of the 14 core NRF2 genes are used to calculate the N2CS.

$$N2CS = \frac{\sum FC_N}{n}$$

In this calculation, FC_N is the individual fold changes for each core NRF2 gene, and n is the number of core NRF2 genes which is 14 in our case. For N2AS, a baseline expression must be established from the group of samples. The mean expression for each gene from the control is set as the baseline and the standard deviation is used to help calculate a z-score for each gene expression.

$$Z = \frac{x - \mu}{\sigma}$$

Thus, the z-score is calculated by the taking the difference of the expression value, x, and the mean, μ, and divide the difference by σ, the standard deviation of this gene expression. The final N2AS is the average of the z-scores of the 14 core NRF2 genes.

$$N2AS = \frac{\sum Z_N}{n}$$

The Z_n is the individual z-scores for each gene which is summed up and divided by the total number of genes in the gene set, *n*. When evaluating other NRF2 gene sets, instead of the 14 core genes, other genes were substituted to calculated a new NRF2 activity score.

Drug Resistance Modeling

Data from Depmap of gene expression and CTRP2 experimental data was downloaded and analyzed in R [76]. Information about gene expression and drug sensitivity (Area Under the Curve) non-small lung cancer cell lines data were collected and merged in R. Cell line expression data was paired for each cell and the N2AS score was calculated for each cell line. A Pearson correlation was performed for every gene or gene set and the drug AUC. The graphs of the area under the curve and N2AS score was plotted for all cell lines using ggplot in R. For analysis of other drugs, the datasets (GSE41929, GSE48812, GSE135842, GSE199779) were used for analysis.

Cell Viability Assay

KEAP1 knockout and wildtype ARH-77 cells were plated in 96 well plates with 2.5 x 10^4 cells per well. 1% DMSO or various concentrations of drugs were added the well. After 48 hours, CellTiter-Glo™ (Promega Corp, Madison, WI) was used to investigate cell viability according to manufacturer's instructions. Necrosulfonamide (#34702, Cayman Chemical), PX-12 (#14192, Cayman Chemical), Paclitaxel (#10461, Cayman Chemical), and bardoxolone methyl (SMB00376, Sigma Aldrich) Experiments were done in triplicates and repeated twice. IC_{50} was calculated using CompuSyn software [77].

Colony Forming Assay

KEAP1 knockout and wildtype ARH-77 cells were plated in 6 well plates with 1000 cells per well. Cells were mixed in MethoCult™ H4434 Classic media (StemCell Technologies). Cells were allowed to grow for 14 days before they were imaged using Keyence BZ-X810. Colonies of more than 10 cells were counted on 10 different views at x20 focus for each well. Experiments were done in triplicates and repeated twice.

Cancer Prognosis Modeling

TCGA Data from cBioportal was collected and analyzed in R [74]. Additional data from GEO, Biostudies [78], and International Cancer Genome Consortium (ICGC) was also collected and processed as needed (Supplemental Table S15). RNA-Seq data was processed to Transcripts per million (TPM) while microarray array data was used as is. R packages Survival and Cutpointr were used to help analyze survival data.

Information about patients' clinical outcomes and N2AS gene expression profiles were combined in R. Z scores for gene expression was used when present or calculated using TPM or microarray data as needed. Patients were subsequently ranked based on their N2AS and corresponding hazard ratios and ranked log p-values were calculated for every cutoff using Survival. These numbers were graphed onto a LOCC graph.

For training, the optimal cutoff consisting of at least 10% of the cohort was selected by Cutpointr (lowest p-value) and manually verified to be appropriate by LOCC visualization. This N2AS cutoff was selected for this cancer and used in the subsequent validation. For a cancer to pass training, the p value must be lower than $p = 0.01$ for at least one cutoff. For validation, instead of using the optimal cutoff, the previous N2AS activity score was used to separate groups. For validation, a p value of 0.05 was

considered significant. Kaplan-Meier plots were generated at each cutoff and calculations for hazard ratio and p value were performed using Survival. LOCC graphs were generated according to the previous study [46].

Data Availability

All RNA-Sequencing data is online through GEO (GSE228434, GSE221526, GSE229914, GSE229704, GSE229652). Other published datasets used in this study are listed in supplemental table S14-15. Code for analysis will be available on publication.

Author's Disclosures

The authors declare no potential conflicts of interest.

Authors' Contributions

Conceptualization, G.L, J.J.L; Methodology, G.L., J.J.L., H.K., E.J., E.R.C; Formal Analysis, G.L., H.K., E.J., K.A; Investigation, G.L., H.K., K.A., E.J., S.R, Resources, G.L., K.A., S.R, A.S., H.M., Writing-Original Draft, G.L., J.L., H.K., Writing – Review & Edits, G.L., J.L., K.A., Supervision, G.L., J.J.L; Funding Acquisition, J.J.L, H.M.

Acknowledgments

G.L is supported by NIH MSTP training grant 5T32GM007250. J.J.L. is supported by the Jane and Lee Seidman Chair in Pediatric Cancer Innovation. Graphics were made with the aid of Biorender. We would like to acknowledge Dr. Michael Sporn for his donation of CDDO-2P-Im and contribution to the field of NRF2 and synthetic triterpenoids. Dr. Sporn passed away in late 2022.

REFERENCES

[1] Q. Ma, Role of nrf2 in oxidative stress and toxicity, Annu Rev Pharmacol Toxicol 53 (2013) 401-26.
[2] K. Taguchi, M. Yamamoto, The KEAP1-NRF2 System in Cancer, Front Oncol 7 (2017) 85.
[3] A. Singh, V. Misra, R.K. Thimmulappa, H. Lee, S. Ames, M.O. Hoque, J.G. Herman, S.B. Baylin, D. Sidransky, E. Gabrielson, M.V. Brock, S. Biswal, Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer, PLoS Med 3(10) (2006) e420.
[4] T. Shibata, A. Kokubu, M. Gotoh, H. Ojima, T. Ohta, M. Yamamoto, S. Hirohashi, Genetic alteration of Keap1 confers constitutive Nrf2 activation and resistance to chemotherapy in gallbladder cancer, Gastroenterology 135(4) (2008) 1358-1368, 1368 e1-4.
[5] T. Shibata, T. Ohta, K.I. Tong, A. Kokubu, R. Odogawa, K. Tsuta, H. Asamura, M. Yamamoto, S. Hirohashi, Cancer related mutations in NRF2 impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy, Proc Natl Acad Sci U S A 105(36) (2008) 13568-73.
[6] A.T. Dinkova-Kostova, R.V. Kostov, A.G. Kazantsev, The role of Nrf2 signaling in counteracting neurodegenerative diseases, FEBS J 285(19) (2018) 3576-3590.
[7] S. Saha, B. Buttari, E. Panieri, E. Profumo, L. Saso, An Overview of Nrf2 Signaling Pathway and Its Role in Inflammation, Molecules 25(22) (2020).
[8] R. Pillai, M. Hayashi, A.M. Zavitsanou, T. Papagiannakopoulos, NRF2: KEAPing Tumors Protected, Cancer Discov 12(3) (2022) 625-643.
[9] A. Cuadrado, A.I. Rojo, G. Wells, J.D. Hayes, S.P. Cousin, W.L. Rumsey, O.C. Attucks, S. Franklin, A.L. Levonen, T.W. Kensler, A.T. Dinkova-Kostova, Therapeutic targeting of the NRF2 and KEAP1 partnership in chronic diseases, Nat Rev Drug Discov 18(4) (2019) 295-317.
[10] A. Cuadrado, G. Manda, A. Hassan, M.J. Alcaraz, C. Barbas, A. Daiber, P. Ghezzi, R. Leon, M.G. Lopez, B. Oliva, M. Pajares, A.I. Rojo, N. Robledinos-Anton, A.M. Valverde, E. Guney, H. Schmidt, Transcription Factor NRF2 as a Therapeutic Target for Chronic Diseases: A Systems Medicine Approach, Pharmacol Rev 70(2) (2018) 348-383.
[11] S. Dayalan Naidu, A.T. Dinkova-Kostova, Omaveloxolone (Skyclarys(TM)) for patients with Friedreich's ataxia, Trends Pharmacol Sci 44(6) (2023) 394-395.
[12] S. Gopal, A. Mikulskis, R. Gold, R.J. Fox, K.T. Dawson, L. Amaravadi, Evidence of activation of the Nrf2 pathway in multiple sclerosis patients treated with delayed-release dimethyl fumarate in the Phase 3 DEFINE and CONFIRM studies, Mult Scler 23(14) (2017) 1875-1883.
[13] A. Hammer, A. Waschbisch, K. Kuhbandner, A. Bayas, D.H. Lee, A. Duscha, A. Haghikia, R. Gold, R.A. Linker, The NRF2 pathway as potential biomarker for dimethyl fumarate treatment in multiple sclerosis, Ann Clin Transl Neurol 5(6) (2018) 668-676.
[14] Y. Yagishita, T.N. Gatbonton-Schwager, M.L. McCallum, T.W. Kensler, Current Landscape of NRF2 Biomarkers in Clinical Trials, Antioxidants (Basel) 9(8) (2020).
[15] A.T. Dinkova-Kostova, I.M. Copple, Advances and challenges in therapeutic targeting of NRF2, Trends Pharmacol Sci 44(3) (2023) 137-149.

[16] M.A. Riedl, A. Saxon, D. Diaz-Sanchez, Oral sulforaphane increases Phase II antioxidant enzymes in the human upper airway, Clin Immunol 130(3) (2009) 244-51.
[17] A. Kopacz, A.I. Rojo, C. Patibandla, D. Lastra-Martínez, A. Piechota-Polanczyk, D. Kloska, A. Jozkowicz, C. Sutherland, A. Cuadrado, A. Grochot-Przeczek, Overlooked and valuable facts to know in the NRF2/KEAP1 field, Free Radic Biol Med 192 (2022) 37-49.
[18] J.M. Raser, E.K. O'Shea, Noise in gene expression: origins, consequences, and control, Science 309(5743) (2005) 2010-3.
[19] J. Liu, T. Lichtenberg, K.A. Hoadley, L.M. Poisson, A.J. Lazar, A.D. Cherniack, A.J. Kovatich, C.C. Benz, D.A. Levine, A.V. Lee, L. Omberg, D.M. Wolf, C.D. Shriver, V. Thorsson, N. Cancer Genome Atlas Research, H. Hu, An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics, Cell 173(2) (2018) 400-416 e11.
[20] M. Cao, E.O. Onyango, C.R. Williams, D.B. Royce, G.W. Gribble, M.B. Sporn, K.T. Liby, Novel synthetic pyridyl analogues of CDDO-Imidazolide are useful new tools in cancer prevention, Pharmacol Res 100 (2015) 135-47.
[21] C. Cancer Cell Line Encyclopedia, C. Genomics of Drug Sensitivity in Cancer, Pharmacogenomic agreement between two cancer cell line data sets, Nature 528(7580) (2015) 84-7.
[22] A. Tsherniak, F. Vazquez, P.G. Montgomery, B.A. Weir, G. Kryukov, G.S. Cowley, S. Gill, W.F. Harrington, S. Pantel, J.M. Krill-Burger, R.M. Meyers, L. Ali, A. Goodale, Y. Lee, G. Jiang, J. Hsiao, W.F.J. Gerath, S. Howell, E. Merkel, M. Ghandi, L.A. Garraway, D.E. Root, T.R. Golub, J.S. Boehm, W.C. Hahn, Defining a Cancer Dependency Map, Cell 170(3) (2017) 564-576 e16.
[23] J.P. Rooney, B. Chorley, S. Hiemstra, S. Wink, X. Wang, D.A. Bell, B. van de Water, J.C. Corton, Mining a human transcriptome database for chemical modulators of NRF2, PLoS One 15(9) (2020) e0239367.
[24] B.N. Chorley, M.R. Campbell, X. Wang, M. Karaca, D. Sambandan, F. Bangura, P. Xue, J. Pi, S.R. Kleeberger, D.A. Bell, Identification of novel NRF2-regulated genes by ChIP-Seq: influence on retinoid X receptor alpha, Nucleic Acids Res 40(15) (2012) 7416-29.
[25] K.C. Patra, N. Hay, The pentose phosphate pathway and cancer, Trends Biochem Sci 39(8) (2014) 347-54.
[26] E.H. Kobayashi, T. Suzuki, R. Funayama, T. Nagashima, M. Hayashi, H. Sekine, N. Tanaka, T. Moriguchi, H. Motohashi, K. Nakayama, M. Yamamoto, Nrf2 suppresses macrophage inflammatory response by blocking proinflammatory cytokine transcription, Nat Commun 7 (2016) 11624.
[27] R. Ahmad, D. Raina, C. Meyer, S. Kharbanda, D. Kufe, Triterpenoid CDDO-Me blocks the NF-kappaB pathway by direct inhibition of IKKbeta on Cys-179, J Biol Chem 281(47) (2006) 35764-9.
[28] L.M. Beaver, A. Buchanan, E.I. Sokolowski, A.N. Riscoe, C.P. Wong, J.H. Chang, C.V. Lohr, D.E. Williams, R.H. Dashwood, E. Ho, Transcriptome analysis reveals a dynamic and differential transcriptional response to sulforaphane in normal and prostate cancer cells and suggests a role for Sp1 in chemoprevention, Mol Nutr Food Res 58(10) (2014) 2001-13.

[29] K. Taguchi, M. Takaku, P.A. Egner, M. Morita, T. Kaneko, T. Mashimo, T.W. Kensler, M. Yamamoto, Generation of a New Model Rat: Nrf2 Knockout Rats Are Sensitive to Aflatoxin B1 Toxicity, Toxicol Sci 152(1) (2016) 40-52.
[30] P. Velica, N.J. Davies, P.P. Rocha, H. Schrewe, J.P. Ride, C.M. Bunce, Lack of functional and expression homology between human and mouse aldo-keto reductase 1C enzymes: implications for modelling human cancers, Mol Cancer 8 (2009) 121.
[31] L.D. Goldstein, J. Lee, F. Gnad, C. Klijn, A. Schaub, J. Reeder, A. Daemen, C.E. Bakalarski, T. Holcomb, D.S. Shames, R.J. Hartmaier, J. Chmielecki, S. Seshagiri, R. Gentleman, D. Stokoe, Recurrent Loss of NFE2L2 Exon 2 Is a Mechanism for Nrf2 Pathway Activation in Human Cancers, Cell Rep 16(10) (2016) 2605-2617.
[32] D.C. Levings, X. Wang, D. Kohlhase, D.A. Bell, M. Slattery, A distinct class of antioxidant response elements is consistently activated in tumors with NRF2 mutations, Redox Biol 19 (2018) 235-249.
[33] R. Romero, V.I. Sayin, S.M. Davidson, M.R. Bauer, S.X. Singh, S.E. LeBoeuf, T.R. Karakousi, D.C. Ellis, A. Bhutkar, F.J. Sánchez-Rivera, L. Subbaraj, B. Martinez, R.T. Bronson, J.R. Prigge, E.E. Schmidt, C.J. Thomas, C. Goparaju, A. Davies, I. Dolgalev, A. Heguy, V. Allaj, J.T. Poirier, A.L. Moreira, C.M. Rudin, H.I. Pass, M.G. Vander Heiden, T. Jacks, T. Papagiannakopoulos, Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis, Nat Med 23(11) (2017) 1362-1368.
[34] C. Tonelli, I.I.C. Chio, D.A. Tuveson, Transcriptional Regulation by Nrf2, Antioxid Redox Signal 29(17) (2018) 1727-1745.
[35] A. Namani, M. Matiur Rahaman, M. Chen, X. Tang, Gene-expression signature regulated by the KEAP1-NRF2-CUL3 axis is associated with a poor prognosis in head and neck squamous cell cancer, BMC Cancer 18(1) (2018) 46.
[36] I.M. Copple, L.M. Shelton, J. Walsh, D.V. Kratschmar, A. Lister, A. Odermatt, C.E. Goldring, A.T. Dinkova-Kostova, T. Honda, B.K. Park, Chemical tuning enhances both potency toward nrf2 and in vitro therapeutic index of triterpenoids, Toxicol Sci 140(2) (2014) 462-9.
[37] M.A. Weniger, E.G. Rizzatti, P. Perez-Galan, D. Liu, Q. Wang, P.J. Munson, N. Raghavachari, T. White, M.M. Tweito, K. Dunleavy, Y. Ye, W.H. Wilson, A. Wiestner, Treatment-induced oxidative stress and cellular antioxidant capacity determine response to bortezomib in mantle cell lymphoma, Clin Cancer Res 17(15) (2011) 5101-12.
[38] M. Songbo, H. Lang, C. Xinyong, X. Bin, Z. Ping, S. Liang, Oxidative stress injury in doxorubicin-induced cardiotoxicity, Toxicol Lett 307 (2019) 41-48.
[39] B.R. You, H.R. Shin, B.R. Han, W.H. Park, PX-12 induces apoptosis in Calu-6 cells in an oxidative stress-dependent manner, Tumour Biol 36(3) (2015) 2087-95.
[40] L. Sun, H. Wang, Z. Wang, S. He, S. Chen, D. Liao, L. Wang, J. Yan, W. Liu, X. Lei, X. Wang, Mixed lineage kinase domain-like protein mediates necrosis signaling downstream of RIP3 kinase, Cell 148(1-2) (2012) 213-27.
[41] L. Guan, D.K. Nambiar, H. Cao, V. Viswanathan, S. Kwok, A.B. Hui, Y. Hou, R. Hildebrand, R. von Eyben, B.J. Holmes, J. Zhao, C.S. Kong, N. Wamsley, W. Zhang, M.B. Major, S.W. Seol, J.B. Sunwoo, D.N. Hayes, M. Diehn, Q.T. Le, NFE2L2 Mutations Enhance Radioresistance in Head and Neck Cancer by Modulating Intratumoral Myeloid Cells, Cancer Res 83(6) (2023) 861-874.

[42] B. Drewinko, W. Mars, J.J. Stragand, S.D. Henderson, J. Latreille, B. Barlogie, J.M. Trujillo, ARH-77, an established human IgG-producing myeloma cell line. II. Growth kinetics, clonogenic capacity, chalone production, xenogeneic transplantations, and response to melphalan, Cancer 54(9) (1984) 1893-903.
[43] T. Takahashi, M. Sonobe, T. Menju, E. Nakayama, N. Mino, S. Iwakiri, S. Nagai, K. Sato, R. Miyahara, K. Okubo, T. Hirata, H. Date, H. Wada, Mutations in Keap1 are a potential prognostic factor in resected non-small cell lung cancer, J Surg Oncol 101(6) (2010) 500-6.
[44] L.M. Solis, C. Behrens, W. Dong, M. Suraokar, N.C. Ozburn, C.A. Moran, A.H. Corvalan, S. Biswal, S.G. Swisher, B.N. Bekele, J.D. Minna, D.J. Stewart, Wistuba, II, Nrf2 and Keap1 abnormalities in non-small cell lung carcinoma and association with clinicopathologic features, Clin Cancer Res 16(14) (2010) 3743-53.
[45] M.M. Saleh, M. Scheffler, S. Merkelbach-Bruse, A.H. Scheel, B. Ulmer, J. Wolf, R. Buettner, Comprehensive Analysis of TP53 and KEAP1 Mutations and Their Impact on Survival in Localized- and Advanced-Stage NSCLC, J Thorac Oncol 17(1) (2022) 76-88.
[46] G. Luo, Letterio, J., LOCC: a novel visualization and scoring of cutoffs for continuous variables, bioRxiv (2023).
[47] J. Chen, H. Yang, A.S.M. Teo, L.B. Amer, F.G. Sherbaf, C.Q. Tan, J.J.S. Alvarez, B. Lu, J.Q. Lim, A. Takano, R. Nahar, Y.Y. Lee, C.Z.J. Phua, K.P. Chua, L. Suteja, P.J. Chen, M.M. Chang, T.P.T. Koh, B.H. Ong, D. Anantham, A.A.L. Hsu, A. Gogna, C.W. Too, Z.W. Aung, Y.F. Lee, L. Wang, T.K.H. Lim, A. Wilm, P.S. Choi, P.Y. Ng, C.K. Toh, W.T. Lim, S. Ma, B. Lim, J. Liu, W.L. Tam, A.J. Skanderup, J.P.S. Yeong, E.H. Tan, C.L. Creasy, D.S.W. Tan, A.M. Hillmer, W. Zhai, Genomic landscape of lung adenocarcinoma in East Asians, Nat Genet 52(2) (2020) 177-186.
[48] M. Zhang, C. Zhang, L. Zhang, Q. Yang, S. Zhou, Q. Wen, J. Wang, Nrf2 is a potential prognostic marker and promotes proliferation and invasion in human hepatocellular carcinoma, BMC Cancer 15 (2015) 531.
[49] H. Yuki, T. Kamai, S. Murakami, S. Higashi, T. Narimatsu, T. Kambara, H. Betsunoh, H. Abe, K. Arai, H. Shirataki, K.I. Yoshida, Increased Nrf2 expression by renal cell carcinoma is associated with postoperative chronic kidney disease and an unfavorable prognosis, Oncotarget 9(47) (2018) 28351-28363.
[50] A. Uruno, M. Yamamoto, The KEAP1-NRF2 System and Neurodegenerative Diseases, Antioxid Redox Signal (2023).
[51] W. Tang, Y.F. Jiang, M. Ponnusamy, M. Diallo, Role of Nrf2 in chronic liver disease, World J Gastroenterol 20(36) (2014) 13079-87.
[52] T. Rangasamy, J. Guo, W.A. Mitzner, J. Roman, A. Singh, A.D. Fryer, M. Yamamoto, T.W. Kensler, R.M. Tuder, S.N. Georas, S. Biswal, Disruption of Nrf2 enhances susceptibility to severe airway inflammation and asthma in mice, J Exp Med 202(1) (2005) 47-59.
[53] A. Sharma, M. Parikh, H. Shah, T. Gandhi, Modulation of Nrf2 by quercetin in doxorubicin-treated rats, Heliyon 6(4) (2020) e03803.
[54] X.J. Wang, Z. Sun, N.F. Villeneuve, S. Zhang, F. Zhao, Y. Li, W. Chen, X. Yi, W. Zheng, G.T. Wondrak, P.K. Wong, D.D. Zhang, Nrf2 enhances resistance of cancer cells to chemotherapeutic drugs, the dark side of Nrf2, Carcinogenesis 29(6) (2008) 1235-43.

[55] M.B. Sporn, K.T. Liby, NRF2 and cancer: the good, the bad and the importance of context, Nat Rev Cancer, England, 2012, pp. 564-71.
[56] A. Galan-Cobo, P. Sitthideatphaiboon, X. Qu, A. Poteete, M.A. Pisegna, P. Tong, P.H. Chen, L.K. Boroughs, M.L.M. Rodriguez, W. Zhang, F. Parlati, J. Wang, V. Gandhi, F. Skoulidis, R.J. DeBerardinis, J.D. Minna, J.V. Heymach, LKB1 and KEAP1/NRF2 Pathways Cooperatively Promote Metabolic Reprogramming with Enhanced Glutamine Dependence in KRAS-Mutant Lung Adenocarcinoma, Cancer Res 79(13) (2019) 3251-3267.
[57] S.E. LeBoeuf, W.L. Wu, T.R. Karakousi, B. Karadal, S.R. Jackson, S.M. Davidson, K.K. Wong, S.B. Koralov, V.I. Sayin, T. Papagiannakopoulos, Activation of Oxidative Stress Response in Cancer Generates a Druggable Dependency on Exogenous Non-essential Amino Acids, Cell Metab 31(2) (2020) 339-350 e4.
[58] H. Ding, Z. Chen, K. Wu, S.M. Huang, W.L. Wu, S.E. LeBoeuf, R.G. Pillai, J.D. Rabinowitz, T. Papagiannakopoulos, Activation of the NRF2 antioxidant program sensitizes tumors to G6PD inhibition, Sci Adv 7(47) (2021) eabk1023.
[59] N.M. Tannir, N. Agarwal, C. Porta, N.J. Lawrence, R. Motzer, B. McGregor, R.J. Lee, R.K. Jain, N. Davis, L.J. Appleman, O. Goodman, Jr., W.M. Stadler, S. Gandhi, D.M. Geynisman, R. Iacovelli, B. Mellado, J.M. Sepulveda Sanchez, R. Figlin, T. Powles, L. Akella, K. Orford, B. Escudier, Efficacy and Safety of Telaglenastat Plus Cabozantinib vs Placebo Plus Cabozantinib in Patients With Advanced Renal Cell Carcinoma: The CANTATA Randomized Clinical Trial, JAMA Oncol 8(10) (2022) 1411-1418.
[60] J.E. Brown, Glutaminase Inhibitors-Do They Have a Role in the Treatment of Metastatic Clear-Cell Renal Cell Carcinoma?, JAMA Oncol 8(10) (2022) 1418-1419.
[61] K.T. Liby, M.B. Sporn, Synthetic oleanane triterpenoids: multifunctional drugs with a broad range of applications for prevention and treatment of chronic disease, Pharmacol Rev 64(4) (2012) 972-1003.
[62] D. Chauhan, G. Li, K. Podar, T. Hideshima, R. Shringarpure, L. Catley, C. Mitsiades, N. Munshi, Y.T. Tai, N. Suh, G.W. Gribble, T. Honda, R. Schlossman, P. Richardson, M.B. Sporn, K.C. Anderson, The bortezomib/proteasome inhibitor PS-341 and triterpenoid CDDO-Im induce synergistic anti-multiple myeloma (MM) activity and overcome bortezomib resistance, Blood 103(8) (2004) 3158-66.
[63] G. Luo, K. Aldridge, T. Chen, V. Aslot, B.G. Kim, E.H. Han, N. Singh, S. Li, T.S. Xiao, M.B. Sporn, J.J. Letterio, The synthetic oleanane triterpenoid CDDO-2P-Im binds GRP78/BiP to induce unfolded protein response-mediated apoptosis in myeloma, Mol Oncol (2023).
[64] C. Computational Pan-Genomics, Computational pan-genomics: status, promises and challenges, Brief Bioinform 19(1) (2018) 118-135.
[65] J. Budczies, F. Klauschen, B.V. Sinn, B. Gyorffy, W.D. Schmitt, S. Darb-Esfahani, C. Denkert, Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization, PLoS One 7(12) (2012) e51862.
[66] M. Ogluszka, M. Orzechowska, D. Jedroszka, P. Witas, A.K. Bednarek, Evaluate Cutpoints: Adaptable continuous data distribution system for determining survival in Kaplan-Meier estimator, Comput Methods Programs Biomed 177 (2019) 133-139.

[67] M.S. Pepe, H. Janes, G. Longton, W. Leisenring, P. Newcomb, Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker, Am J Epidemiol 159(9) (2004) 882-90.
[68] Z. Liu, M. Deng, L. Wu, S. Zhang, An integrative investigation on significant mutations and their down-stream pathways in lung squamous cell carcinoma reveals CUL3/KEAP1/NRF2 relevant subtypes, Mol Med 26(1) (2020) 48.
[69] D.B. Fox, N.M.G. Garcia, B.J. McKinney, R. Lupo, L.C. Noteware, R. Newcomb, J. Liu, J.W. Locasale, M.D. Hirschey, J.V. Alvarez, NRF2 activation promotes the recurrence of dormant tumour cells through regulation of redox and nucleotide metabolism, Nat Metab 2(4) (2020) 318-334.
[70] A. Kechin, U. Boyarskikh, A. Kel, M. Filipenko, cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing, J Comput Biol 24(11) (2017) 1138-1143.
[71] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat Methods 12(4) (2015) 357-60.
[72] M. Pertea, G.M. Pertea, C.M. Antonescu, T.C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat Biotechnol 33(3) (2015) 290-5.
[73] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26(1) (2010) 139-40.
[74] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C.J. Byrne, M.L. Heuer, E. Larsson, Y. Antipin, B. Reva, A.P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Cancer Discov 2(5) (2012) 401-4.
[75] C. Lee, C.H. Huang, LASAGNA-Search 2.0: integrated transcription factor binding site search and visualization in a browser, Bioinformatics 30(13) (2014) 1923-5.
[76] M.G. Rees, B. Seashore-Ludlow, J.H. Cheah, D.J. Adams, E.V. Price, S. Gill, S. Javaid, M.E. Coletti, V.L. Jones, N.E. Bodycombe, C.K. Soule, B. Alexander, A. Li, P. Montgomery, J.D. Kotz, C.S. Hon, B. Munoz, T. Liefeld, V. Dancik, D.A. Haber, C.B. Clish, J.A. Bittker, M. Palmer, B.K. Wagner, P.A. Clemons, A.F. Shamji, S.L. Schreiber, Correlating chemical sensitivity and basal gene expression reveals mechanism of action, Nat Chem Biol 12(2) (2016) 109-16.
[77] T.-C. Chou, Martin, N., A Computer Program for Quantitation of Synergism and Antagonism in Drug Combinations, and the Determination of IC50 and ED50 Values., CompuSyn For Drug Combinations User's Guide.
[78] U. Sarkans, M. Gostev, A. Athar, E. Behrangi, O. Melnichuk, A. Ali, J. Minguet, J.C. Rada, C. Snow, A. Tikhonov, A. Brazma, J. McEntyre, The BioStudies database-one stop shop for all data supporting a life sciences study, Nucleic Acids Res 46(D1) (2018) D1266-D1270.

Supporting Information

Supplemental Figure S1. Gene Set Enrichment Analysis of Pharmacological or Genetic Induction of NRF2.

Supplemental Figure S2. Literature references and CHIP-Seq analysis of candidate core NRF2 genes as target genes.

Supplemental Figure S3. Induction of candidate core NRF2 gene expression after drug treatment or genetic modification.

Supplemental Figure S4. AKR1C family gene show significant differences in gene induction in drug treatment and genetic modification.

Supplemental Fig S5. Predicted NRF2 binding sites in near universal and conditional NRF2 genes.

Supplemental Fig S6. Predicted NRF2 binding sites in conditional NRF2 genes and candidate downregulated NRF2 genes.

Supplemental Fig S7. Activation of NRF2 by genetic modifications may lead to suppression of inflammatory genes.

Supplemental Fig S8. Validation of the near universal NRF2 Gene Set in Other Published Datasets.

Supplemental Fig S9. Activation of NRF2 is correlated to drug resistance and increased colony formation *in vitro*.

Supplemental Fig S10. Mutations of NFE2L2 and KEAP1 were not significantly associated with prognosis in TCGA datasets.

Supplemental Fig S11. Analysis of TCGA cancers that did not pass training with NRF2 activity score.

Supplemental Fig S12. Analysis of TCGA cancers that passed training with NRF2 activity score but did not have a validation cohort.

Supplemental Fig S13. LOCC cutoff selection of TCGA cancers that passed training with NRF2 activity score and validation.

Supplemental Fig S14. Analysis of TCGA cancers that passed training with NRF2 activity score but not validation.

Supplemental Fig S15. Other NRF2 gene sets is associated with survival for some cancers but not others.

Supplemental Fig. S16. **Exclusion of Irregular Control Samples Affecting NRF2 Target Genes.**

Supplemental Table S1. Core NRF2 Target Gene Fold Changes After CDDO-2P-Im treatment or KEAP1 knockout

Supplemental Table S2. Near Universal NRF2 Target Gene Fold Changes After CDDO-2P-Im treatment or KEAP1 knockout

Supplemental Table S3. Conditional NRF2 Target Gene Fold Changes After CDDO-2P-Im treatment or KEAP1 knockout

Supplemental Table S4. OVCAR8 Low Concentration CDDO-2P-Im Upregulated Gene Fold Changes

Supplemental Table S5. SF8628 Low Concentration CDDO-2P-Im Upregulated Gene Fold Changes

Supplemental Table S6. Primary Dermal Fibroblast Low Concentration CDDO-2P-Im Upregulated Gene Fold Changes

Supplemental Table S7. RPMI-8226 Low Concentration CDDO-2P-Im Upregulated Gene Fold Changes

Supplemental Table S8. ARH-77 KEAP1 Knockout Upregulated Gene Fold Changes

Supplemental Table S9. TCGA LUAD NRF2 Gene Fold Changes

Supplemental Table S10. CCLE NSCLC NRF2 Gene Fold Changes

Supplemental Table S11. Overlapping NRF2 target genes from Published Studies

Supplemental Table S12. NRF2 Activity Score Drug Correlations

Supplemental Table S13. CDDO-2P-Im Drug Concentrations

Supplemental Table S14. NRF2 Validation and Change Score Datasets

Supplemental Table S15. NRF2 Cancer Prognosis Datasets

FIGURE LEGENDS

Figure 1. Generation of Transcriptomic Datasets of NRF2 Activation. A, Schematic overview of the NRF2/KEAP1 pathway and our approach towards understanding transcriptional consequences. Cells are either treated with CDDO-2P-Im to inhibit KEAP1-NRF2 binding or have mutations leading to NRF2 activation. NRF2 moves to the nucleus, forms a heterodimer with small MAF proteins (sMAF), and activate target genes. Transcripts are taken from cells and then analyzed by RNA-Sequencing. **B-E,** *NQO1* expression by RNA-Sequencing is significantly increased after treatment with CDDO-2P-Im in (**B**) SF8628, (**C**) OVCAR8, (**D**) Primary Dermal Fibroblasts, (**E**) RPMI-8226. **F,** DNA of KEAP1 knockout (KO) ARH-77 cells is sequenced following CRISPR-Cas9 editing. **G,** *KEAP1* exon 5 RNA is evaluated by RT-qPCR following CRISPR-Cas9 editing. **H,** Western blot of KEAP1 protein is performed after CRISPR-Cas9 editing. **I,** *NQO1* RNA is evaluated by RT-qPCR following CRISPR-Cas9 editing. **J,** NQO1 protein is evaluated by western blot following CRISPR-Cas9 editing. **K-L**, *NQO1* RNA expression is evaluated between *KEAP1* or *NFE2L2* mutant and WT tumor samples using (**K**) TCGA LUAD database or (**L**) Depmap's NSCLC CCLE Database. Cancer cell lines are from the following cell types: SF8628 – Diffuse Intrinsic Pontine Glioma, OVCAR8 – Ovarian Cancer, RPMI-8226 – Multiple Myeloma, ARH-77 – B cell lymphoma. LUAD – Lung Adenocarcinoma, NSCLC – Non-small cell lung cancer. All RNA-Sequencing experiments were performed once with 3 replicates. Other experiments were performed independently twice with similar results. Values = means ± SD (n = 3). ns, $P \geq 0.05$; *, $P < 0.05$; **, $P < 0.01$.

Figure 2. A Core NRF2 Transcriptional Signature is Present after NRF2 Activation. A, Heatmap of candidate core NRF2 genes following low dose of CDDO-

2P-Im (underlined cell lines) or genetic alteration of NRF2/KEAP1. Highly upregulated in red represents at least a 2 fold change while upregulated represents at least a 1.25 fold change. Not expressed means that gene expression is less than 1 fragments per kilobase of exon per million mapped fragments (FPKM). **B,** Heatmap of near universal upregulated NRF2 genes following NRF2 activation. **C-E**, Gene expression fold changes by RNA-sequencing were evaluated for the following genes: (**C**) *SRXN1*, (**D**) *ABHD4*, and (**E**) *AKR1C3*. **F,** NRF2/ARE (Antioxidant Response Element) binding motif analysis was performed on candidate core NRF2 genes from transcription start site (TSS) to -5000 base pairs using LASAGNA-Search 2.0. ARE numbers are ranked based on highest likelihood of NRF2 binding based on sequence similarity to consensus sequence. Only binding sequences that have a p value of 0.0005 or lower are shown. All RNA-Sequencing experiments were performed once. Abbreviations: C – Control, S_L – SF8268 low dose 2P, S_H – SF8268 high dose 2P, O_L – OVCAR8 low dose 2P, O_H – OVCAR8 high dose 2P, F_L – Primary Dermal Fibroblast low dose 2P, F_H – Primary Dermal Fibroblast high dose 2P, R_L – RPMI-8226 low dose 2P, R_H – RPMI-8226 high dose 2P, A_K – ARH-77 KEAP1 knockout. Values = means ± SD (n = 3). ns, $P \geq 0.05$; *, $P < 0.05$; **, $P < 0.01$. ~ indicates this gene did not pass subsequent validation with other datasets.

Figure 3. Defining Conditionally Upregulated and Novels Targets Genes of NRF2 Activation. A, Heatmap of conditional, novel NRF2 target genes following low dose of CDDO-2P-Im (underlined cells line) or genetic alteration of NRF2/KEAP1. **B-C,** Gene expression changes of primarily drug-induced NRF2 target genes, (**B**) *HMOX1* and (**C**) *AKIRIN2*. **D-G,** Gene expression changes of pentose phosphate pathway enzymes, (**D**)

G6PD, (**E**) *TKT*, (**F**) *TALDO1*, (**G**) *PGD*. **H-I,** Gene expression changes of novel upregulated NRF2 target genes, (**H**) *B4GALNT1* and (**I**) *POPDC3*. All RNA-Sequencing experiments were performed once. Abbreviations: C – Control, S_L – SF8268 low dose 2P, S_H – SF8268 high dose 2P, O_L – OVCAR8 low dose 2P, O_H – OVCAR8 high dose 2P, P_L – Primary Dermal Fibroblast low dose 2P, P_H – Primary Dermal Fibroblast high dose 2P, R_L – RPMI-8226 low dose 2P, R_H – RPMI-8226 high dose 2P, A_K – ARH-77 KEAP1 knockout. Values = means ± SD (n = 3). ns, $P \geq 0.05$; *, $P < 0.05$; **, $P < 0.01$.

Figure 4. Validation of the Core NRF2 Gene Set in Other Published Datasets. A-C, Candidate core NRF2 genes are evaluated in TCGA (**A**) head neck squamous cell carcinoma (HNSC), (**B**) lung squamous cell carcinoma (LUSC), (**C**) liver hepatocellular carcinoma (LIHC) cancer datasets with significant number of *KEAP1*/*NFE2L2* mutations. **D,** Candidate core NRF2 genes are evaluated using published data which uses NRF2 activator, sulforaphane, for 24 hours. Genes that were upregulated more than 25% and had a p value under 0.1 were colored in pink while those that were upregulated by 2 fold or more were colored in red. **E**, Candidate core NRF2 genes were examined for their presence in previously published NRF2 gene sets. Red in the heatmap signifies that they were included in the published gene set. **F,** In a previously published rat model of WT and NRF2 KO, candidate core NRF2 were evaluated to see if gene expression changes were observed after treatment with a NRF2 activator, CDDO-Im, and if NRF2 knockout affect upregulation. **G,** NRF2 change score (N2CS) was evaluated across different treatments of CDDO-2P-Im or KEAP1 knockout. **H**, N2CS was evaluated for other drugs from published datasets. Abbreviations: C – Control, S_L – SF8268 low dose 2P, S_H – SF8268 high dose 2P, O_L – OVCAR8 low dose

2P, O_H – OVCAR8 high dose 2P, P_L – Primary Dermal Fibroblast low dose 2P, P_H – Primary Dermal Fibroblast high dose 2P, R_L – RPMI-8226 low dose 2P, R_H – RPMI-8226 high dose 2P, A_K – ARH-77 KEAP1 knockout, BTZ – Bortezomib, DOXO – doxorubicin, PTX – Paclitaxel, SFN – Sulforaphane. Values = means ± SD (n = 3). ns, P ≥ 0.05; *, P < 0.05; **, P < 0.01. ***, P<0.001 ~ means this gene did not pass validation with other datasets.

Figure 5. A NRF2 Activity Score is Correlated with Resistance to Radiation and Certain Drugs. A, NRF2 Activity Score (N2AS) is calculated for each sample from the TCGA LUAD and color coded by mutation status of NRF2/KEAP1. KEAP1 mutations are colored in blue, NRF2/*NFE2L2* is colored in red, and WT samples is colored in gray. **B-C**, N2AS for each cell line was plotted against sensitivity to (**B**) necrosulfonamide and (**C**) PX-12 represented by area under the curve (AUC). The correlation was calculated for N2AS and the AUC. **D-E,** WT and KEAP1 KO ARH-77 were treated with (**D**) necrosulfonamide or (**E**) PX-12 for 48 hours and evaluated for cell viability by CellTiter-Glo. **F,** Radiation was performed on WT and KEAP1 KO ARH-77 cells at indicated radiation doses. Cells were incubated in 96 well plates for 72 hours and cell viability was measured by CellTiter-Glo. **G-H**, N2AS for each cell line was plotted against sensitivity to (**G**) paclitaxel and (**H**) bardoxolone methyl represented by area under the curve (AUC). The correlation was calculated for N2AS and the AUC. **I-J**, WT and KEAP1 KO ARH-77 were treated with (**I**) paclitaxel or (**J**) bardoxolone methyl for 48 hours and evaluated for cell viability by CellTiter-Glo. Values = means ± SD (n = 3). ns, P ≥ 0.05; *, P < 0.05; **, P < 0.01. All CellTiter-Glo experiments were repeated independently twice with similar results.

Figure 6. NRF2 Activity is Associated with Poor Prognosis in a Variety of Cancers. **A**, Flow chart of the training and validation results using N2AS to evaluate cancer prognosis. **B**, Results for individual TCGA cancers are categorized. Cancers colored in red represent a high N2AS is associated with a poor prognosis while those in green represent a high N2AS is associated with a good prognosis. **C,** TCGA LUAD was trained using LOCC to find the most significant cutoff at N2AS of 0.463. A Kaplan-Meier survival curve was plotted and a p value using cox proportional hazard regression was calculated. **D,** OncoSG LUAD was used as validation with the cutoff of N2AS of 0.463 to determine significance and reproducibility. A Kaplan-Meier survival curve was plotted and a p value using cox proportional hazard regression was calculated. **E**, TCGA KIRC was trained using LOCC to find the most significant cutoff at N2AS of 0.488. A Kaplan-Meier survival curve was plotted and a p value using cox proportional hazard regression was calculated. **F,** MTAB-1980 was used as validation with the cutoff of N2AS of 0.488 to determine significance and reproducibility. A Kaplan-Meier survival curve was plotted and a p value using cox proportional hazard regression was calculated. **G**, TCGA LIHC was trained using LOCC to find the most significant cutoff at N2AS of 0.604. A Kaplan-Meier survival curve was plotted and a p value using cox proportional hazard regression was calculated. **H,** LIKI-JP was used as validation with the cutoff of N2AS of 0.604 to determine significance and reproducibility. A Kaplan-Meier survival curve was plotted and a p value using cox proportional hazard regression was calculated. **I**, TCGA LAML was trained using LOCC to find the most significant cutoff at N2AS of 0.432. A Kaplan-Meier survival curve was plotted and a p value using cox proportional hazard regression was calculated. **J,** OSHU AML was used as validation with the cutoff of N2AS of 0.432

to determine significance and reproducibility. A Kaplan-Meier survival curve was plotted and a p value using cox proportional hazard regression was calculated. Abbreviations: LUAD – Lung adenocarcinoma, KIRC/CCRCC – Kidney renal clear cell carcinoma, LIHC/LIKI – Liver hepatocellular carcinoma, LAML/AML – Acute myeloid leukemia, HNSC – Head and neck squamous cell carcinoma, SKMC – skin cutaneous melanoma, PAAD - pancreatic adenocarcinoma, BLCA - Bladder urothelial carcinoma, KIRP – Kidney renal papillary cell carcinoma, LGG - Brain lower grade glioma, UCEC - Uterine corpus endometrial carcinoma, OV - Ovarian serous cystadenocarcinoma.

Fig 1.

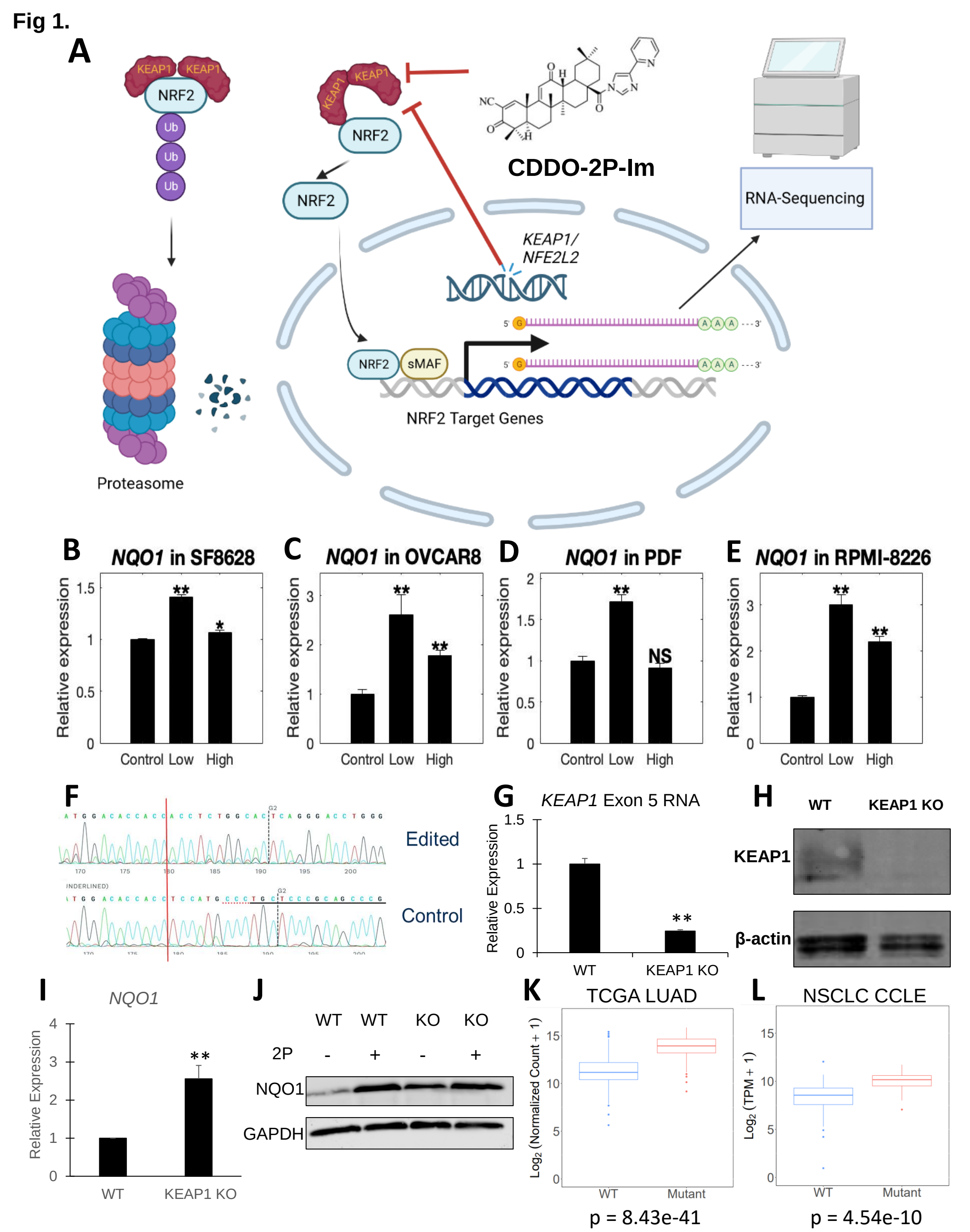

A
KEAP1
KEAP1
NRF2
Ub
Ub
Ub
Proteasome
NRF2
CDDO-2P-Im
RNA-Sequencing
KEAP1/
NFE2L2
NRF2
sMAF
NRF2 Target Genes
B
NQO1 in SF8628
Relative expression
Control Low High
C
NQO1 in OVCAR8
Relative expression
Control Low High
D
NQO1 in PDF
Relative expression
NS
Control Low High
E
NQO1 in RPMI-8226
Relative expression
Control Low High
F
Edited
Control
G
KEAP1 Exon 5 RNA
Relative Expression
WT
KEAP1 KO
H
WT
KEAP1 KO
KEAP1
β-actin
I
NQO1
Relative Expression
WT
KEAP1 KO
J
WT WT KO KO
2P - + - +
NQO1
GAPDH
K
TCGA LUAD
Log2 (Normalized Count + 1)
WT
Mutant
p = 8.43e-41
L
NSCLC CCLE
Log2 (TPM + 1)
WT
Mutant
p = 4.54e-10

Fig 2.

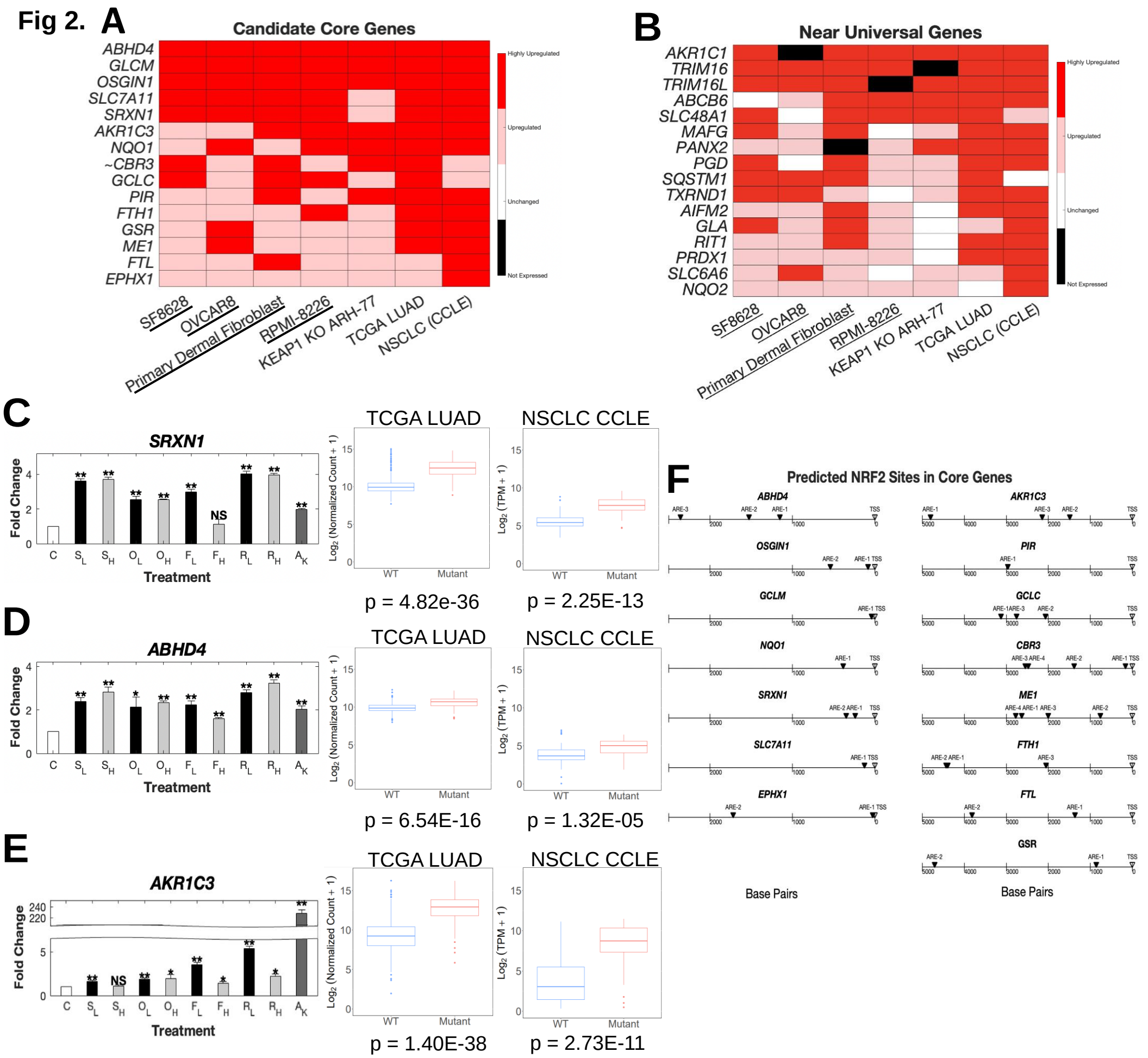

A
Candidate Core Genes
ABHD4
GLCM
OSGIN1
SLC7A11
SRXN1
AKR1C3
NQO1
~CBR3
GCLC
PIR
FTH1
GSR
ME1
FTL
EPHX1
Highly Upregulated
Upregulated
Unchanged
Not Expressed
SF8628
OVCAR8
Primary Dermal Fibroblast
RPMI-8226
KEAP1 KO ARH-77
TCGA LUAD
NSCLC (CCLE)
B
Near Universal Genes
AKR1C1
TRIM16
TRIM16L
ABCB6
SLC48A1
MAFG
PANX2
PGD
SQSTM1
TXRND1
AIFM2
GLA
RIT1
PRDX1
SLC6A6
NQO2
Highly Upregulated
Upregulated
Unchanged
Not Expressed
SF8628
OVCAR8
Primary Dermal Fibroblast
RPMI-8226
KEAP1 KO ARH-77
TCGA LUAD
NSCLC (CCLE)
C
SRXN1
Fold Change
Treatment
C
S_L
S_H
O_L
O_H
F_L
F_H
R_L
R_H
A_K
NS
TCGA LUAD
NSCLC CCLE
Log2 (Normalized Count + 1)
Log2 (TPM + 1)
WT
Mutant
p = 4.82e-36
p = 2.25E-13
D
ABHD4
Fold Change
Treatment
TCGA LUAD
NSCLC CCLE
p = 6.54E-16
p = 1.32E-05
E
AKR1C3
Fold Change
Treatment
NS
TCGA LUAD
NSCLC CCLE
p = 1.40E-38
p = 2.73E-11
F
Predicted NRF2 Sites in Core Genes
ABHD4
AKR1C3
OSGIN1
PIR
GCLM
GCLC
NQO1
CBR3
SRXN1
ME1
SLC7A11
FTH1
EPHX1
FTL
GSR
ARE-1
ARE-2
ARE-3
ARE-4
TSS
Base Pairs

Fig 3.

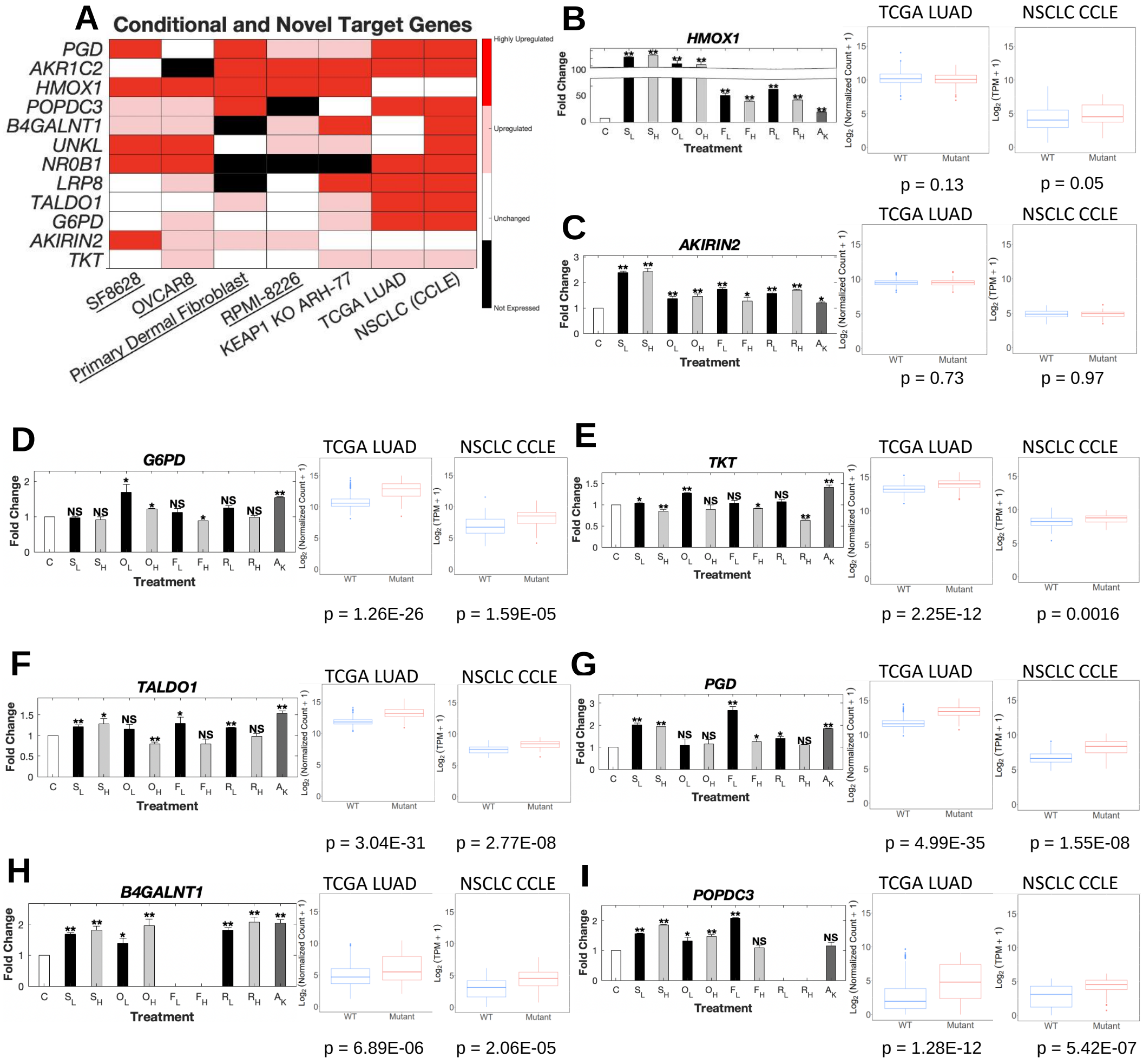

A
Conditional and Novel Target Genes
PGD
AKR1C2
HMOX1
POPDC3
B4GALNT1
UNKL
NR0B1
LRP8
TALDO1
G6PD
AKIRIN2
TKT
SF8628
OVCAR8
Primary Dermal Fibroblast
RPMI-8226
KEAP1 KO ARH-77
TCGA LUAD
NSCLC (CCLE)
Highly Upregulated
Upregulated
Unchanged
Not Expressed
B
HMOX1
Fold Change
Treatment
TCGA LUAD
NSCLC CCLE
p = 0.13
p = 0.05
C
AKIRIN2
p = 0.73
p = 0.97
D
G6PD
p = 1.26E-26
p = 1.59E-05
E
TKT
p = 2.25E-12
p = 0.0016
F
TALDO1
p = 3.04E-31
p = 2.77E-08
G
PGD
p = 4.99E-35
p = 1.55E-08
H
B4GALNT1
p = 6.89E-06
p = 2.06E-05
I
POPDC3
p = 1.28E-12
p = 5.42E-07
WT
Mutant

Fig 4.

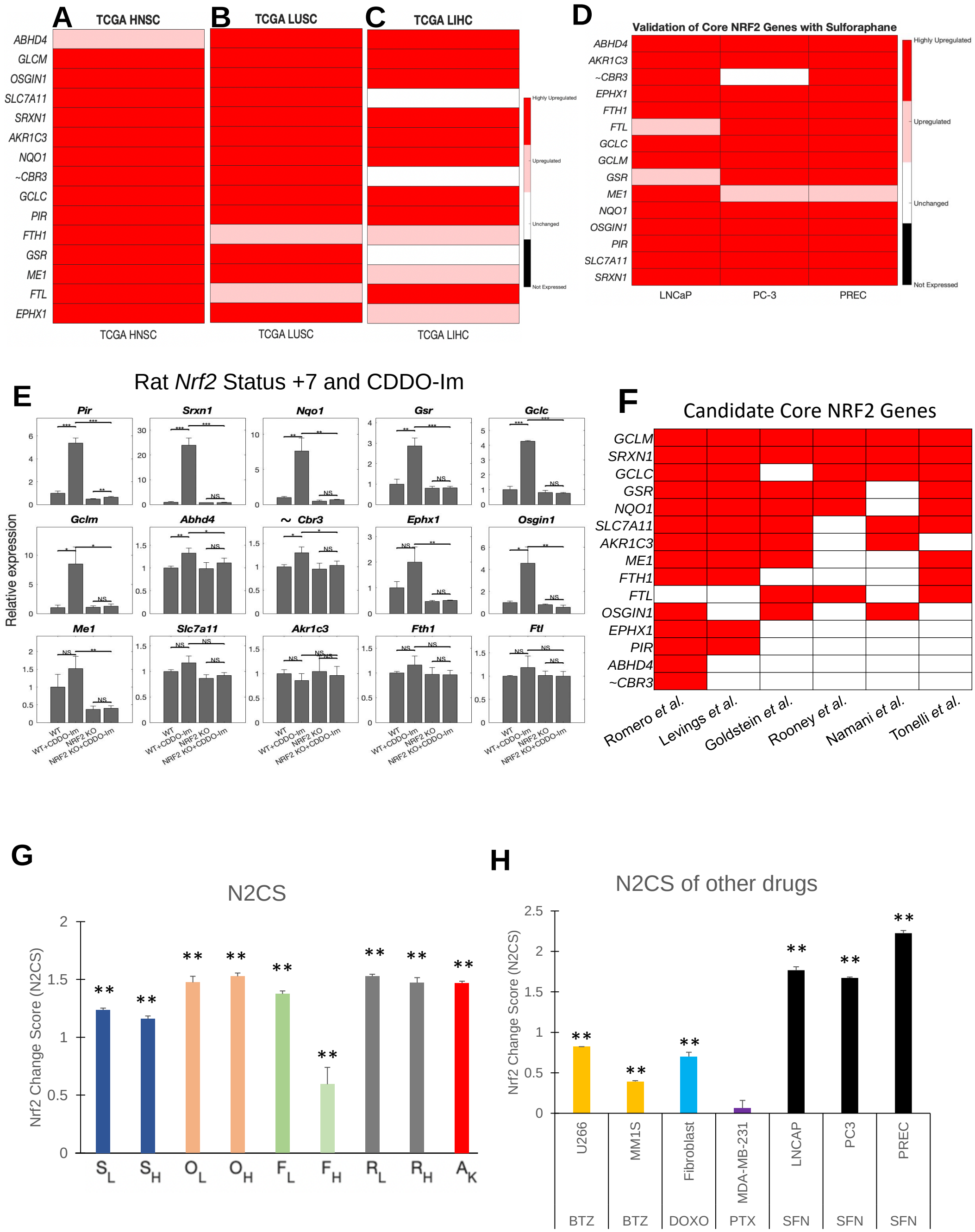

A
TCGA HNSC
B
TCGA LUSC
C
TCGA LIHC
ABHD4
GLCM
OSGIN1
SLC7A11
SRXN1
AKR1C3
NQO1
~CBR3
GCLC
PIR
FTH1
GSR
ME1
FTL
EPHX1
Highly Upregulated
Upregulated
Unchanged
Not Expressed
D
Validation of Core NRF2 Genes with Sulforaphane
LNCaP
PC-3
PREC
E
Rat Nrf2 Status +7 and CDDO-Im
Relative expression
Pir
Srxn1
Nqo1
Gsr
Gclc
Gclm
Abhd4
~ Cbr3
Ephx1
Osgin1
Me1
Slc7a11
Akr1c3
Fth1
Ftl
WT
WT+CDDO-Im
NRF2 KO
NRF2 KO+CDDO-Im
F
Candidate Core NRF2 Genes
GCLM
SRXN1
GCLC
GSR
NQO1
SLC7A11
AKR1C3
ME1
FTH1
FTL
OSGIN1
EPHX1
PIR
ABHD4
~CBR3
Romero et al.
Levings et al.
Goldstein et al.
Rooney et al.
Namani et al.
Tonelli et al.
G
N2CS
Nrf2 Change Score (N2CS)
S_L
S_H
O_L
O_H
F_L
F_H
R_L
R_H
A_K
H
N2CS of other drugs
Nrf2 Change Score (N2CS)
U266
MM1S
Fibroblast
MDA-MB-231
LNCAP
PC3
PREC
BTZ
BTZ
DOXO
PTX
SFN
SFN
SFN

Fig 5.

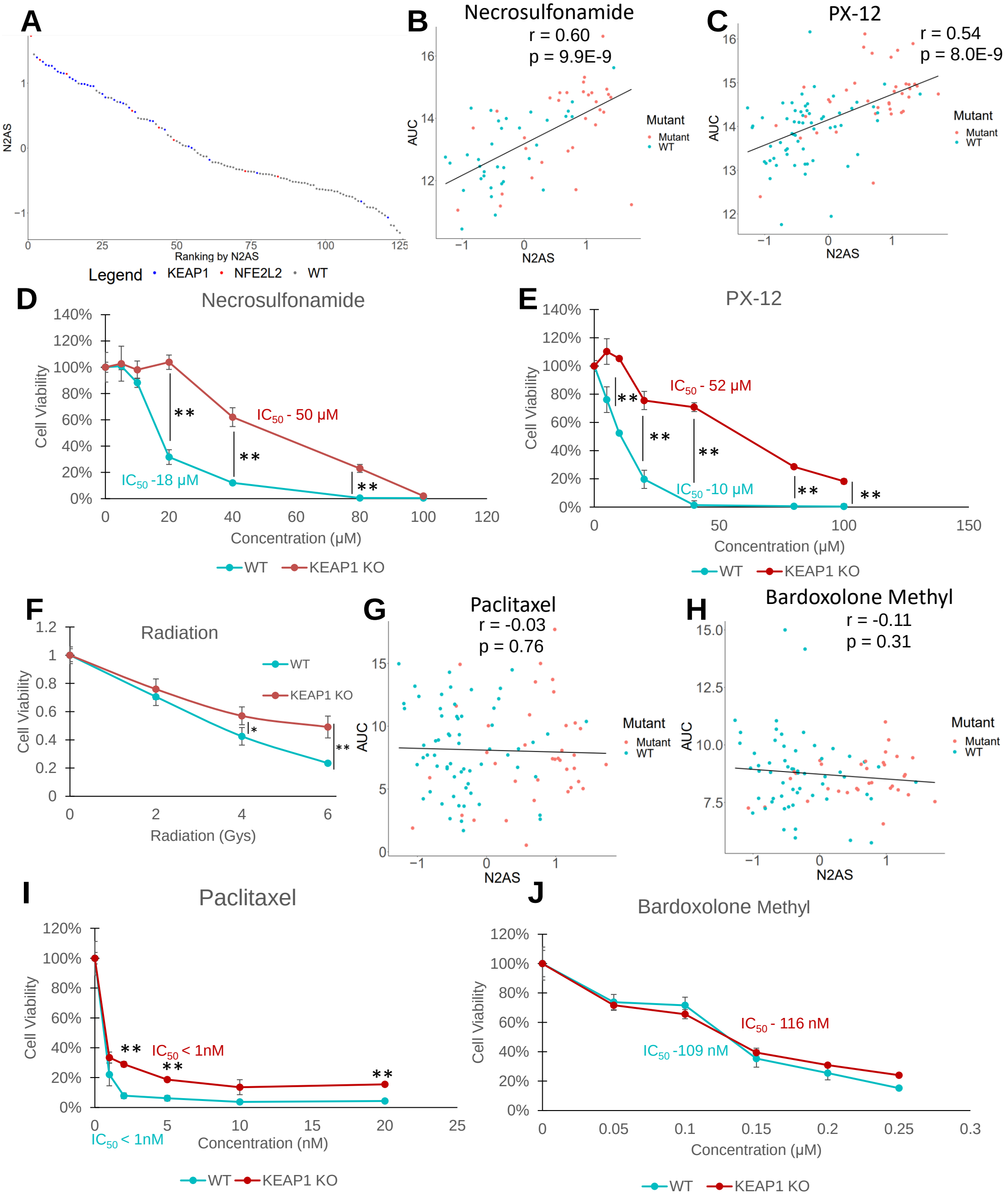
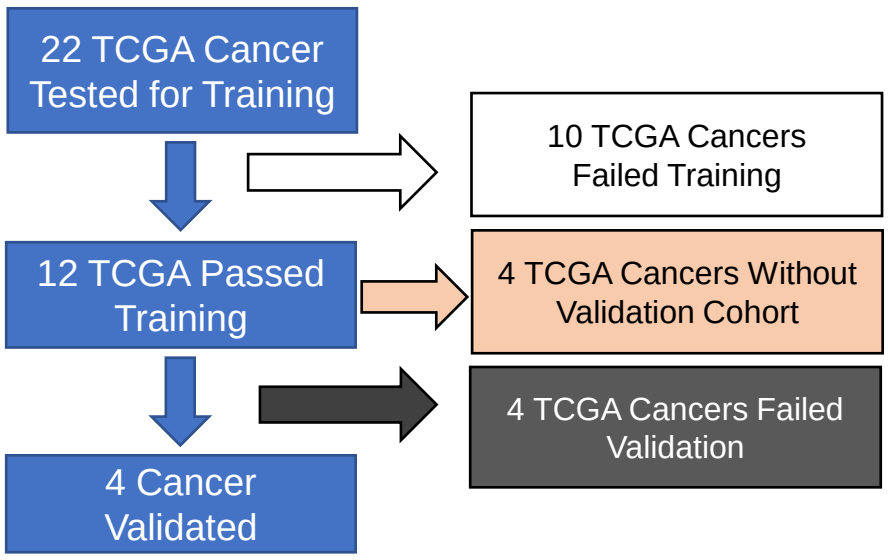

A
N2AS
Ranking by N2AS
Legend KEAP1 NFE2L2 WT
B
Necrosulfonamide
r = 0.60
p = 9.9E-9
AUC
N2AS
Mutant
Mutant
WT
C
PX-12
r = 0.54
p = 8.0E-9
AUC
N2AS
Mutant
Mutant
WT
D
Necrosulfonamide
Cell Viability
Concentration (μM)
IC50 - 50 μM
IC50 -18 μM
WT
KEAP1 KO
E
PX-12
Cell Viability
Concentration (μM)
IC50 - 52 μM
IC50 -10 μM
WT
KEAP1 KO
F
Radiation
Cell Viability
Radiation (Gys)
WT
KEAP1 KO
G
Paclitaxel
r = -0.03
p = 0.76
AUC
N2AS
Mutant
Mutant
WT
H
Bardoxolone Methyl
r = -0.11
p = 0.31
AUC
N2AS
Mutant
Mutant
WT
I
Paclitaxel
Cell Viability
Concentration (nM)
IC50 < 1nM
IC50 < 1nM
WT
KEAP1 KO
J
Bardoxolone Methyl
Cell Viability
Concentration (μM)
IC50 - 116 nM
IC50 -109 nM
WT
KEAP1 KO

Fig 6.

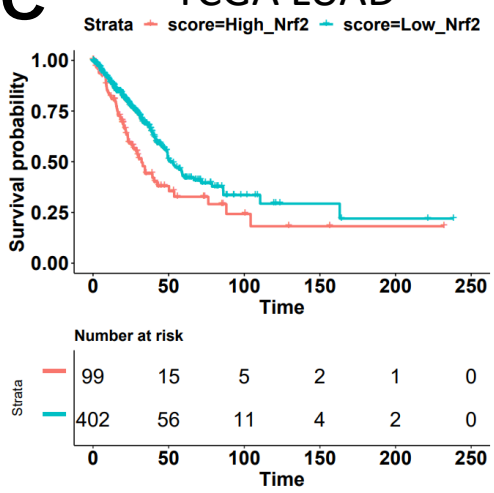
A



B

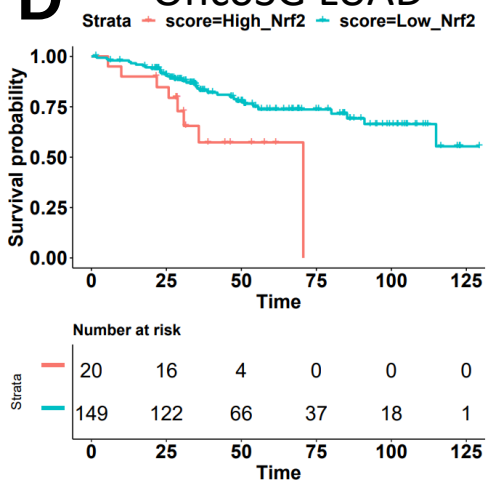
Passed Training but no Validation	Passed Training but failed Validation	**Validated Nrf2 Prognosis Cancers**
KIRP	BLCA	**LUAD**
LGG	SKMC	**KIRC**
UCEC	PAAD	**LIHC**
OV	HNSC	**LAML**

C TCGA LUAD



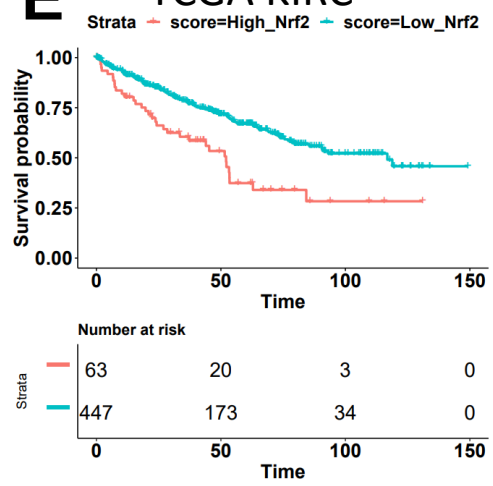
p= 0.003	n	events	median
High_Nrf2	99	52	32.1
Low_Nrf2	402	129	52.6

D OncoSG LUAD



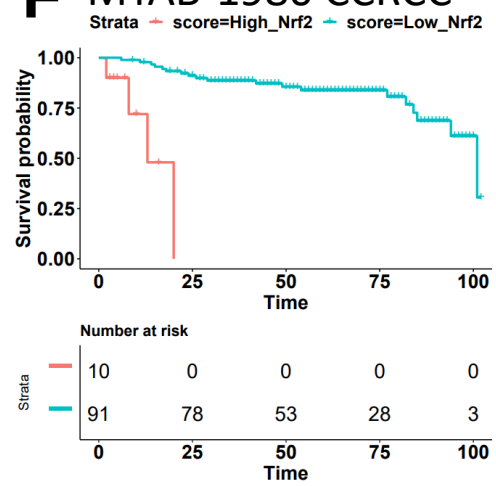
p= 0.01	n	events	median
High_Nrf2	20	8	70.6
Low_Nrf2	149	34	NA

E TCGA KIRC



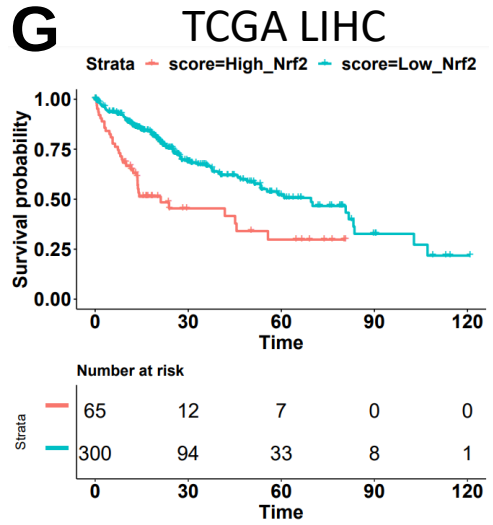
p= 0.00006	n	events	median
High_Nrf2	63	34	52.1
Low_Nrf2	447	134	116.8

F MTAB-1980 CCRCC



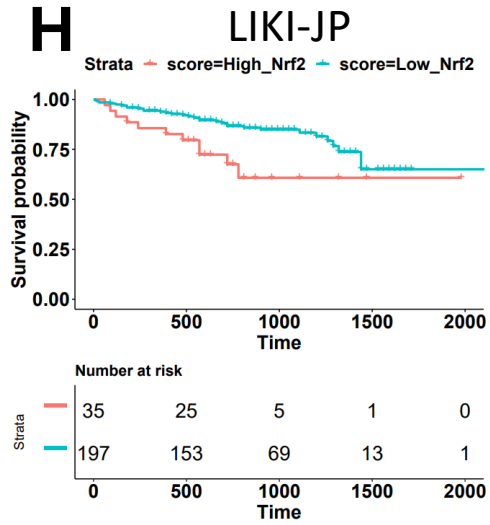
p= 4E-10	n	events	median
High_Nrf2	10	4	13
Low_Nrf2	91	19	101

G TCGA LIHC



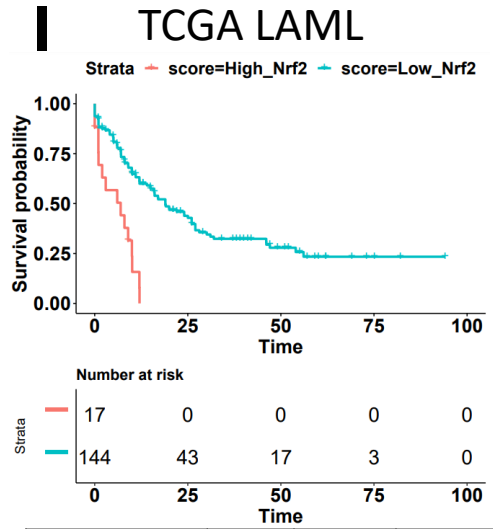
p= 0.0001	n	events	median
High_Nrf2	63	32	23.8
Low_Nrf2	303	97	60.9

H LIKI-JP



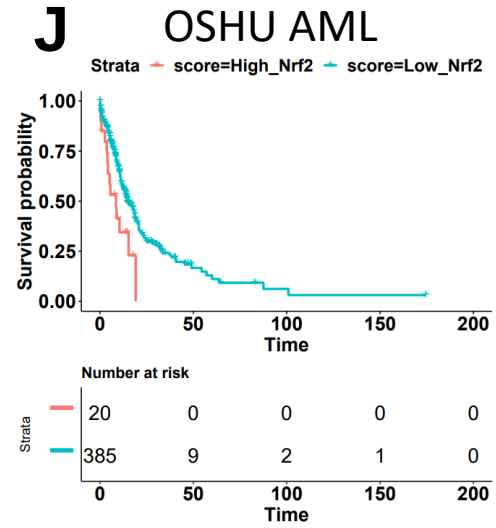
p= 0.01	n	events	median
High_Nrf2	35	11	NA
Low_Nrf2	197	32	NA

I TCGA LAML



p= 0.000001	n	events	median
High_Nrf2	17	15	6.97
Low_Nrf2	144	88	18.97

J OSHU AML



p= 0.005	n	events	median
High_Nrf2	20	14	8.52
Low_Nrf2	385	224	14.77