

<https://doi.org/10.1038/s41698-025-01088-0>

# Integrative analysis of *KEAP1/NFE2L2* alterations across 3600+ tumors reveals an NRF2 expression signature as a prognostic biomarker in cancer

Check for updates

Valentina Crippa<sup>1,3</sup>, Nicoletta Cordani<sup>1,3</sup>, Alberto Maria Villa<sup>1</sup>, Federica Malighetti<sup>1</sup>, Matteo Villa<sup>1</sup>, Luca Sala<sup>2</sup>, Andrea Aroldi<sup>2</sup>, Rocco Piazza<sup>1,2</sup>, Diego Cortinovis<sup>1,2</sup>, Luca Mogni<sup>1</sup> & Daniele Ramazzotti<sup>1</sup> ✉

Non-small cell lung cancer (NSCLC) remains a formidable global health challenge, with heterogeneous molecular characteristics influencing prognosis and treatment response. We present a novel computational framework named ASTUTE (Association of SomaTic mUtaTions to gene Expression profiles), designed to perform genotype-phenotype mapping through the integration of genomic and transcriptomic data. Through the systematic analysis of over 3600 samples from diverse NSCLC datasets and multiple cancer types, we uncovered intricate associations between *KEAP1/NFE2L2* mutations and the NRF2 pathway activation. Our study identified novel NRF2-related functionalities associated with specific genetic alterations and revealed a *KEAP1/NFE2L2* expression signature predictive of prognosis across different cancer types. These findings enhance our understanding of cancer pathogenesis and drug resistance mechanisms mediated by NRF2 activation, paving the way for tailored therapeutic interventions and the development of prognostic biomarkers. Our approach exemplifies the power of integrating genomic and transcriptomic data to elucidate cancer mechanisms, thereby advancing the field of precision oncology.

Lung cancer is the leading cause of cancer-related mortality on a global scale, accounting for an estimated 1.6 million lives lost annually. Among its heterogeneous subtypes, non-small cell lung cancer (NSCLC) represents a significant proportion, with diverse histological manifestations. Notably, NSCLC encompasses two predominant subtypes, lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), each presenting distinct molecular profiles and clinical behaviors<sup>1</sup>.

Over the years, considerable efforts have been devoted to deciphering the intricate genomic landscape of NSCLC, shedding light on the critical role of driver gene mutations in its pathogenesis and progression. Notable among these driver genes we find: the epidermal growth factor receptor (*EGFR*), the Kirsten rat sarcoma viral oncogene homolog (*KRAS*), the tumor protein p53 (*TP53*), and the Kelch-like ECH-associated protein 1 (*KEAP1*), along with its downstream effector, the nuclear factor erythroid 2-related factor 2 (NRF2). Mutations in these key genes orchestrate a complex cascade of events, influencing tumor evolution, therapeutic response, and patient prognosis. While *EGFR* mutations are prevalent in a subset of NSCLC cases and can typically be associated with a more favorable prognosis, also due to

targeted therapies<sup>2</sup>, *KRAS* and *TP53* mutations are associated with aggressive tumor behavior and resistance to most therapies, except for immunotherapy, where these mutations may actually indicate a better response to immune checkpoint blockade<sup>3</sup>. Furthermore, the dysregulation of the *KEAP1*-NRF2 pathway, primarily attributed to somatic mutations in *KEAP1* or Nuclear factor erythroid-derived 2-like 2 (*NFE2L2*), has emerged as a significant determinant of disease progression and therapeutic resistance in NSCLC<sup>4</sup>. Recent evidence indicates that mutations in the *KEAP1* gene, a key regulator of cellular response to oxidative stress, exert a detrimental impact on the prognosis of NSCLC patients<sup>5</sup> and are correlated with resistance to immunotherapy in LUAD<sup>6</sup> and to *KRAS* inhibitors<sup>7</sup>. The *KEAP1* gene plays a critical role in maintaining cellular homeostasis. It oversees the cellular defense against oxidative damage and metabolic stress by modulating the activity of NRF2, a master transcription factor that controls the expression of antioxidant and detoxification enzymes. Mutations in *KEAP1* prevent its binding to the NRF2 degron motifs, thereby impeding the targeting for proteasomal degradation. This results in NRF2 entering the nucleus, where it triggers the expression of target genes

<sup>1</sup>Department of Medicine and Surgery, University of Milano-Bicocca, Milano, Italy. <sup>2</sup>Fondazione IRCCS San Gerardo dei Tintori di Monza, Monza, Italy. <sup>3</sup>These authors contributed equally: Valentina Crippa, Nicoletta Cordani.

✉ e-mail: [daniele.ramazzotti@unimib.it](mailto:daniele.ramazzotti@unimib.it)

containing antioxidant response elements (ARE) in their promoters. Consequently, this facilitates metabolic reprogramming and detoxification<sup>8</sup>.

Similarly, mutations in the *NFE2L2* gene occur in the *KEAP1* binding sites, resulting in the constitutive activation of the NRF2 pathway. This activation contributes to disease progression, metastatic spread, and enhanced resilience against cytotoxic substances<sup>8</sup>.

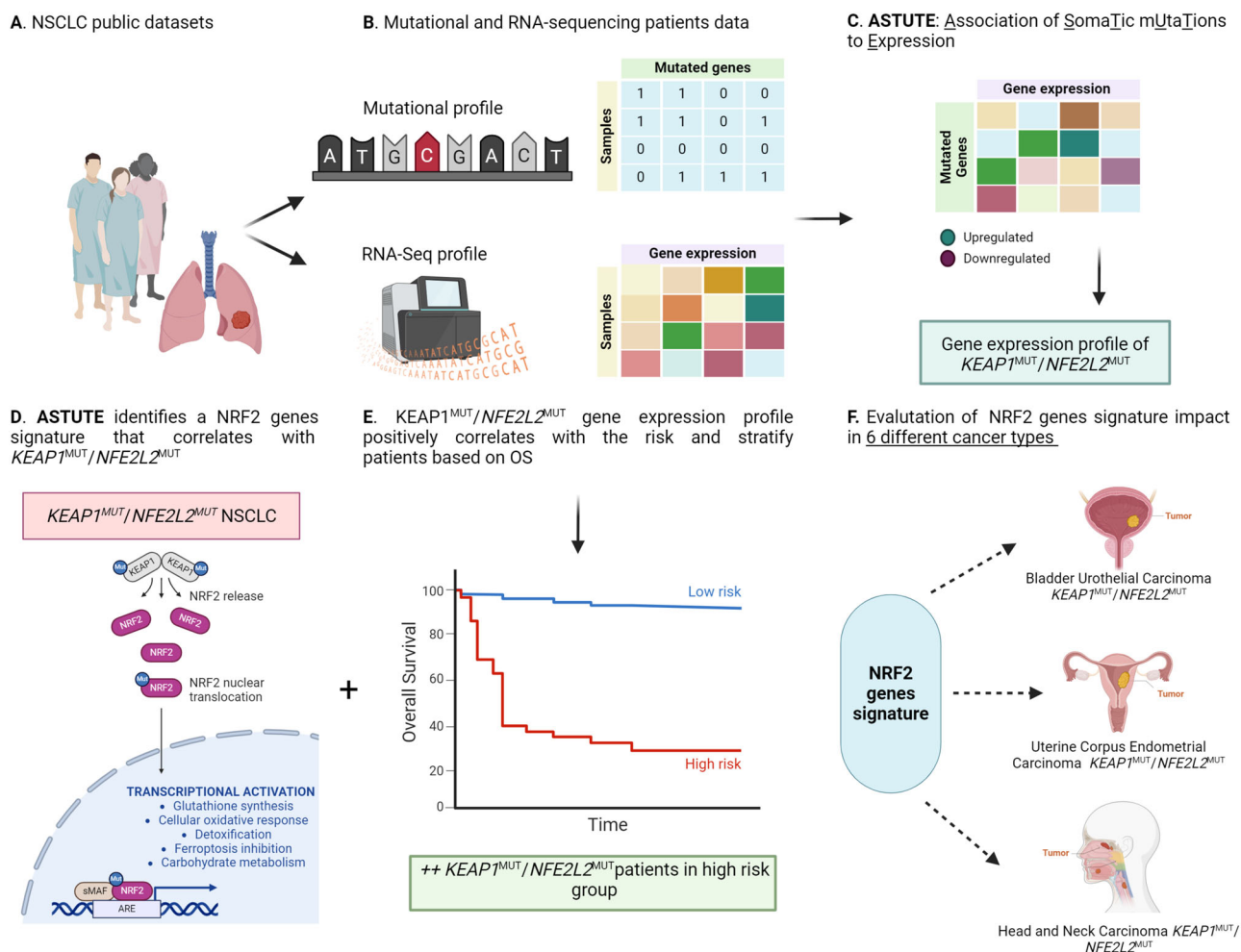
The comprehensive understanding of the intricate genomic landscape of complex cancers such as NSCLC may greatly benefit from the effective integration of both genomic and transcriptomic data. This integrative approach may allow for a deeper examination of the interplay between somatic mutations and gene expression, thereby potentially offering invaluable insights into tumor biology and guiding therapeutic strategies. To address this critical need, we here introduce the Association of SomaTic mUtaTions to gene Expression profiles (ASTUTE) framework, designed to characterize genotype-phenotype associations in cancer. Through the integration of genomic and transcriptomic datasets, ASTUTE provides a sophisticated analytical tool to uncover novel molecular mechanisms driving cancer progression and treatment response, thus contributing to the advancement of precision oncology. Employing ASTUTE, we here systematically analyzed distinct NSCLC datasets, followed by an exploration across different cancer types harboring *KEAP1* or *NFE2L2* mutations. This comprehensive investigation allowed us to elucidate the intricate correlation between *KEAP1/NFE2L2* mutations and the activation of the NRF2 pathway, shedding light on novel NRF2-related functionalities

associated with specific genetic alterations. Additionally, we identified an expression signature associated with mutations in *KEAP1* or *NFE2L2* genes across different cancer types, showing a robust association with prognosis. Our discoveries offer significant insights into the underlying mechanisms of NSCLC pathogenesis and drug resistance. While several NRF2-related gene signatures have been previously reported, our approach differs in that it derives a mutation-driven expression signature through a rigorous statistical framework. By leveraging a robust approach to associate somatic mutations in *KEAP1/NFE2L2* to gene expression, rather than expression-based clustering or pathway annotations, we aim to delineate a mechanistically grounded, reproducible transcriptional program associated with NRF2 pathway activation. These results present a promising avenue for the development of tailored therapeutic interventions and prognostic biomarkers. Furthermore, the relevance of this study becomes even more evident as specific inhibitors of the NRF2 axis become available in clinical practice.

## Results

### Genotype-phenotype mapping in cancer: the ASTUTE framework

Understanding how genetic mutations influence observable traits is crucial. ASTUTE is a novel computational framework capable of performing genotype-phenotype mapping between somatic mutations and expression data (Fig. 1).



**Fig. 1 | Overview of the ASTUTE framework.** In A–C we illustrate how our framework can efficiently integrate mutations with gene expression data to perform the extraction of dysregulated genes associated with *KEAP1* or *NFE2L2* mutations in distinct NSCLC datasets. In D we highlight the consistent association of the identified genes with the NRF2 pathway. In E we show that ASTUTE can stratify patients based on the

identified expression signatures, thus enhancing the prognostic insights returned by the approach. Finally, in F we showcase that ASTUTE could determine a set of genes consistently dysregulated across in NSCLC and other cancer types, emphasizing their role in prognosis at the pan-cancer level. ASTUTE's multidimensional analysis enables a deeper understanding of genotype-phenotype associations in cancer.

Leveraging regularized regression with the LASSO penalty, ASTUTE employs a sophisticated approach based on regularized linear regression and the bootstrap<sup>9</sup> that incorporates a penalty term into the loss function, thus effectively mitigating overfitting and performing feature selection<sup>10</sup>. The LASSO penalty is determined by summing the absolute values of the model coefficients and then multiplying this sum by a regularization parameter, whose selection is optimized through cross-validation to determine the degree of penalty applied to the model. This process induces feature selection by encouraging coefficients associated with the less influential variables to shrink toward zero, resulting in a more interpretable model that emphasizes the most significant variables for predictive purposes. Consequently, the resulting models may include coefficients denoting features lacking a significant association with driver gene mutations, while other genes may exhibit significantly altered expression levels due to the presence of specific mutations. The application of the ASTUTE framework to distinct NSCLC datasets extracted a set of genes consistently upregulated or downregulated in association with mutations in either *KEAP1* or *NFE2L2* genes in lung cancer (Fig. 1A–C). Most of these genes are known to be associated with the NRF2 pathway (Fig. 1D).

Furthermore, ASTUTE offers the capability to estimate baseline gene expression levels, calculate fold changes with respect to the baseline, and compute *p* values using the bootstrapping technique to determine whether a fold change significantly indicates over- or under-expression. The selected features can be exploited for further analyses, such as Gene Set Enrichment Analysis (GSEA), to elucidate the biological implications of genetic somatic mutations (Fig. 1C).

Unlike multi-omics latent factor approaches, ASTUTE is specifically tailored to infer direct genotype-to-phenotype associations. While multi-omics factor analysis methods typically focus on identifying latent variables that explain joint variation across data types, ASTUTE is designed to capture direct, interpretable associations between somatic mutations and gene expression changes. This makes ASTUTE particularly well-suited for mechanistic studies and biomarker discovery, rather than exploratory dimensionality reduction. It employs LASSO regularization to extract sparse, interpretable gene sets whose expression changes are directly linked to specific somatic mutations. This design makes ASTUTE well-suited for biomarker discovery and mechanistic inference, rather than broad unsupervised factor analysis.

Moreover, ASTUTE enables patient stratification based on the identified expression signatures, enhancing prognostic insights (Fig. 1E). The signatures, associated with specific gene mutations, provide a phenotypic characterization of the related somatic mutations, which, in turn, can serve as clinical biomarkers. In the case of the identified NRF2 expression signatures, our approach discovered a set of genes consistently dysregulated in both NSCLC and other cancer types. These genes could effectively stratify patients based on prognosis in all the analyzed cancers, which presented frequent mutations in the *KEAP1/NFE2L2* genes and were consistently overexpressed in cancers with worse prognoses (Fig. 1F). We describe these results in detail in the next sections.

Identification of a NRF2 expression signature associated with *KEAP1* and *NFE2L2* mutations in NSCLC

We applied the ASTUTE framework to analyze five distinct datasets providing both somatic mutations and gene expression data from LUAD and LUSC patients. Specifically, for LUAD, we considered data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) consortium (110 samples)<sup>11</sup>, the Pan-Cancer Atlas (510 samples)<sup>12</sup>, and the study by Chen et al. (169 samples)<sup>13</sup>, for a total of 789 LUAD patients. For LUSC, data from the CPTAC consortium (108 samples)<sup>14</sup> and the Pan-Cancer Atlas (481 samples)<sup>15</sup> were included, comprising 589 LUSC patients. In total, our analysis encompassed 1378 NSCLC samples. ASTUTE was independently executed on each dataset, and the results were compared to identify consistent findings across all datasets (Supplementary Data 1 and 2).

ASTUTE revealed significant upregulation of a specific set of genes strongly associated with the NRF2 pathway activation in LUAD patients

harboring *KEAP1* mutations and in LUSC patients with either *KEAP1* or *NFE2L2* mutations. The low frequency of *NFE2L2* mutations in LUAD patients precluded extensive exploration within this cohort.

These upregulated genes, previously identified as NRF2 targets, fall into three main functional classes crucial for cellular processes (Table 1). First, genes involved in glutathione synthesis, such as *GCLM* and *GCLC*, play

**Table 1 | LUAD/LUSC NRF2 expression signature. Genes identified by the ASTUTE framework in association with the presence of NFE2L2 or KEAP1 mutations in both LUAD and LUSC**

LUAD/LUSC NRF2 signature		
Genes	Functional classes	References
GCLC	Glutathione synthesis	16,17
GCLM		
ABCB6	Cellular oxidative response Detoxification Ferroptosis inhibition	72
AKR1B10		73
AKR1C1		
AKR1C2		74
AKR1C3		
ALDH3A1		
ALDH3A2		75
CBR1		
CBR3		76
CYP4F3		
CYP4F11		77
EPHX1		
GPX2		78
GSR	Carbohydrate metabolism NADPH generation	79
LRP8		76
NQO1		80
OSGIN1		81
PIR		72
PRDX1		82
PTGR1		83
RIT1		20
SRXN1		84
TXN		84
TXNRD1	Other functions	54
UCHL1		85
G6PD		86
ME1		87
PGD		88
TALDO1		89
TKT		90
UGDH		91
B4GALNT1		92
CABYR		20
CBX2		25
JAKMIP3		93
PANX2		94
TRIM16		20
TRIM16L		95

These genes can be categorized into 5 functional classes and are associated with the NRF2 pathway activation.

significant roles in maintaining redox balance<sup>16,17</sup>. Second, a substantial group of genes implicated in cellular oxidative response, detoxification, or inhibition of ferroptosis (Table 1), underscores the critical role of NRF2 in protecting cells from oxidative stress and promoting survival under toxic conditions<sup>18</sup>. Third, genes associated with carbohydrate metabolism and NADPH generation (Table 1), highlight the NRF2 pathway's involvement in sustaining metabolic flexibility and anabolic growth in cancer cells<sup>19</sup>. To experimentally validate these findings, the expression of 10 NRF2 targets was investigated by quantitative PCR in LUAD cells carrying comparable genetic makeup but different NFE2L2 mutational status. Specifically, H2228 cells harbor a gain-of-function mutation in NFE2L2 (G31A), while H3122 cells are wild-type for both KEAP1 and NFE2L2 (see Methods). All tested genes were strongly upregulated in KEAP1-mutated cells compared with KEAP-WT cells (Supplementary Fig. 1).

We further validated the identified signatures using proteomics data from CPTAC. Specifically, we compared protein levels of the genes listed in Table 1 between LUAD (Supplementary Fig. 2) and LUSC (Supplementary Fig. 3) patients, stratified by the presence or absence of KEAP1/NFE2L2 mutations. Patients with mutations consistently showed higher expression of the selected genes at the protein level as well.

In addition to these categories, we identified other NRF2 targets with roles extending beyond these classes (Table 1). *TRIM16* stands out among these genes, which has been reported to associate with the p62-KEAP1-NRF2 complex, suggesting a potential positive feedback loop that may enhance NRF2 signaling activation, offering insights into the regulatory mechanisms driving this pathway<sup>20,21</sup>.

### Association of other driver genes with the LUAD/LUSC NRF2 expression signature

Some genes listed in Table 1 also exhibited significant associations with other driver mutations. Notably, in LUAD, the genes *AKR1C1*, *AKR1C2*, *AKR1C3*, and *ALDH3A1* were positively associated with *STK11* mutations ( $0.2 < \log_2\text{FC} < 1$ ), while *GPX2* showed a similar association with *SMARCA4* mutations ( $0.3 < \log_2\text{FC} < 0.5$ ). In LUSC, the genes *AKR1C1*, *AKR1C2*, *AKR1C3*, and *GPX2* were associated with *KMT2D* mutations ( $0.2 < \log_2\text{FC} < 0.7$ ), and *AKR1C1*, *AKR1C2*, *ALDH3A1*, *CYP4F3*, *GCLC*, *GCLM*, and *GPX2* were associated with *TP53* mutations ( $0.2 < \log_2\text{FC} < 1.3$ ).

These results underscore the varying degrees of association between different mutations and the expression of the NRF2 signature genes in both LUAD and LUSC. While *KEAP1/NFE2L2* mutations are primary drivers of the NRF2 pathway activation, these findings suggest the existence of alternative mechanisms influencing NRF2 expression independently of these mutations.

Interestingly, *ABCC2*<sup>22</sup>, *SLC7A11*<sup>23</sup>, *AIFM2*<sup>24</sup>, and *NEIL3*<sup>25</sup> genes were identified as NRF2 target genes associated with *KEAP1* mutations in LUAD, while in LUSC, these genes were linked with *NFE2L2* mutations. This divergence might reflect distinct molecular pathways underlying NRF2 activation in these cancer types.

Moreover, our analysis revealed two genes involved in the NRF2 pathway selectively upregulated in LUAD but not LUSC in the presence of *KEAP1* mutations: *CPLX2*, known for its role in regulating NRF2 expression in hepatocellular carcinoma (HCC)<sup>26</sup> and recognized as a potential prognostic biomarker in lung cancer<sup>27</sup>, and *KYNU*, extensively linked to NRF2 pathway activation<sup>28,29</sup>. Additionally, we observed other genes associated with promoting lung cancer progression, which exhibited positive associations with *KEAP1*, *SMARCA4*, and *STK11* mutations in LUAD but not LUSC: *S100P*, which encodes a calcium-binding protein implicated in *KEAP1/NRF2* signaling that regulates the mobility of lung cancer cells<sup>30,31</sup>, and *SERPINB5*, acting as a prognostic biomarker and promoter of proliferation in LUAD<sup>32</sup>.

These findings not only deepen our understanding of NRF2 pathway regulation in lung cancer but also highlight potential biomarkers and therapeutic targets for personalized treatment strategies in different molecular subtypes of LUAD and LUSC.

### Identification and characterization of a LUSC-specific NRF2 signature

Furthermore, our analysis has revealed the upregulation of 25 genes (see Table 2) specific for LUSC patients harboring either *KEAP1* or *NFE2L2* mutations. Notably, this upregulation has not been directly observed in association with *KEAP1* mutations in LUAD patients. Particularly noteworthy among these genes are two classes involved in detoxification processes: (i) the Glutathione S-transferase Mu (GSTM) gene family (*GSTM2*, *GSTM3*, and *GSTM4*), critical in eliminating electrophilic compounds by conjugating with glutathione<sup>33,34</sup>, and (ii) the UDP-glucuronosyltransferases family, essential for drug clearance through the glucuronidation process<sup>35,36</sup>. We will refer to the genes listed in Table 1 as *LUSC NRF2* expression signature, to indicate the NRF2 target genes whose expression was identified by ASTUTE to correlate with the presence of *NFE2L2* and *KEAP1* mutations exclusively in LUSC.

As for the NRF2 identified expression signature common to LUAD and LUSC, some of the genes associated with *KEAP1/NFE2L2* mutations selectively in LUSC also exhibit significant associations with other driver mutations. For instance, *UGT1A6*, *ADH7*, and *NDRG4* genes are positively associated with *TP53* mutations ( $\log_2\text{FC} > 0.5$ ), while also *ADAM23*, *UGT1A1*, and *GSTM3* show a positive association but a lower strength ( $\log_2\text{FC} < 0.5$ ). Additionally, *ADH7* correlates strongly ( $\log_2\text{FC} = 0.75$ ) with *KMT2D* mutations, and *UGT1A7* and *NDRG4* are associated with *CDKN2A* mutations ( $0.2 < \log_2\text{FC} < 0.5$ ). *ABCA4* and *NELL1* gene expression is linked with *EGFR* mutations ( $0.2 < \log_2\text{FC} < 0.7$ ) specifically in LUSC. These findings suggest that while LUAD and LUSC patients with *KEAP1* or *NFE2L2* mutations share a common NRF2 pathway activation signature, LUSC also exhibits a distinct NRF2 signature. Notably, many of these genes show increased upregulation in LUSC patients with *TP53*

**Table 2 | LUSC NRF2 expression signature**

LUSC NRF2 signature
ABCA4
ABCC3
ADAM23
ADH7
AKR1B15
ALDH1A1
CES1
GSTM2
GSTM3
GSTM4
NDRG4
NELL1
RAB3B
RAB6B
RPS6KA6
SCN9A
SLC47A1
SOX2
SQSTM1
UGT1A1
UGT1A3
UGT1A4
UGT1A6
UGT1A7
WNT11

Genes identified by the ASTUTE framework in association with the presence of *NFE2L2* or *KEAP1* mutations exclusively in LUSC.



mutations, suggesting a potential synergistic effect between these mutations.

Interestingly, in addition to canonical regulators, we identified several non-canonical alterations—such as *CUL3* mutations, *AKT2* amplifications, and *PTEN* deletions—that are associated with upregulation of NRF2 target genes in specific tumor types. These findings suggest that the NRF2 pathway may also be modulated through broader signaling mechanisms, including ubiquitination, phosphorylation, and metabolic reprogramming. We explored the impact of *CUL3* mutations, detected in approximately 3–4% of both LUSC and LUAD patients, on the NRF2 signature. *CUL3*, a member of the E3 ubiquitin ligase complex, plays a role in NRF2 degradation. Notably, we observed the presence of numerous genes reported in LUAD/LUSC NRF2 expression signature or LUSC NRF2 expression signature (Supplementary Data 1 and 2) in LUSC patients carrying *CUL3* mutations, whereas this signature was not corroborated in LUAD.

### Impact of the NRF2 signatures across cancer types

We applied ASTUTE to other cancer types to investigate whether the NRF2 gene expression signatures identified in NSCLC were also observed in association with *NFE2L2* or *KEAP1* mutations at the pan-cancer level. We considered cancers where either *KEAP1* or *NFE2L2* mutations were present in at least 5% of the patients for a total of 2258 distinct samples. In particular, we selected HCC (361 samples), head and neck squamous cell carcinoma (HNSCC, 507 samples), uterine corpus endometrial carcinoma (UCEC, 515 samples), cervical squamous cell carcinoma (CSCC, 288 samples), bladder urothelial carcinoma (BLCA, 406 samples), and esophageal adenocarcinoma (EAC, 181 samples) from the Pan-Cancer Atlas studies<sup>37</sup>.

In HCC, *KEAP1* and *NFE2L2* mutations occur respectively in about 5% and 3% of the patients. Of note, the activation of the NRF2 pathway has been reported to induce tumor cells to immune escape in HCC<sup>38</sup>. Additionally, NRF2 was identified as a prognostic factor associated with decreased survival in HCC patients<sup>39</sup>. ASTUTE confirmed the correlation between the upregulation of many members of the NRF2 signature and *NFE2L2* mutations in HCC (Supplementary Data 3). In particular, the six genes most upregulated ( $\log_2FC > 1.9$ ) were NRF2 targets: *AKR1B15*, *NQO1*, *AKR1B10*, *CABYR*, *TRIM16L*, and *CPLX2*. Similar results were observed for *KEAP1* mutations, which positive correlated ( $\log_2FC > 1.9$ ) with *CPLX2*, *CABYR*, *TRIM16L*, *TRIM16*, *AKR1B15*, and *AKR1B10*.

In head and neck squamous carcinoma, *KEAP1* and *NFE2L2* mutations are observed, respectively, in about 4% and 5% of patients. It was reported that the activation of NRF2 signaling promotes the acquisition of resistance to cisplatin and metastasis in HNSCC<sup>40</sup>. ASTUTE correlated both mutations with the expression of many genes of the LUAD/LUSC NRF2 expression signature (Supplementary Data 4). Interestingly, these genes appear to be among the most upregulated ones in the presence of these mutations ( $\log_2FC > 1$ ) in this cancer type. We also identified a strong correlation between *KEAP1* and *NFE2L2* mutations and the expression of genes identified in the LUSC NRF2 expression signature.

In CSCC, 1% and 6% of patients harbor, respectively *KEAP1* and *NFE2L2* mutations. The expression of NRF2 was found to be higher in CSCC patients with lymph node metastasis, and in addition NRF2 pathway was positively associated with epithelial to mesenchymal transition<sup>41</sup>.

*KEAP1* and *NFE2L2* mutations were identified to be strongly associated with both the expression of genes within the LUAD/LUSC NRF2 expression signature and the ones related to LUSC NRF2 expression signature (Supplementary Data 5).

About 4% and 8% of the patients with UCEC harbor respectively *KEAP1* or *NFE2L2* mutations. NRF2 overexpression was found to be associated with endometrial neoplasms with serous differentiation<sup>42</sup>. In this tumor type, ASTUTE was able to correlate these mutations to the expression of *ALDH3A1*, *AKR1C2*, *GPX2*, *AKR1C1*, *NQO1*, *TRIM16L*, *CYP4F3*, *JAKMIP3*, *AKR1B10*, *AKR1C3*, and *GCLC* ( $\log_2FC > 0.5$ ) (Supplementary Data 6). Additionally, other LUAD/LUSC NRF2 signature genes were identified to correlate with a major impact with *NFE2L2* mutations in comparison to *KEAP1* mutations. While among the genes identified the

LUSC NRF2 signature, we found only that *CES1*, and *UGT1A6* consistently correlated with *KEAP1/NFE2L2* mutations ( $\log_2FC > 0.7$ ).

In BLCA, *KEAP1* mutations are present in 2% of patients, while *NFE2L2* mutations occur in 6%. NRF2 expression was associated with cisplatin resistance in BLCA<sup>43</sup>. Among the most upregulated genes in the context of *NFE2L2* mutations ( $\log_2FC > 0.8$ ) in bladder cancer, ASTUTE identified 35 genes belonging to the LUAD/LUSC NRF2 expression signature, and 8 genes belonging to the LUSC NRF2 expression signature (Supplementary Data 7). While for the *KEAP1* mutations 13 genes of the LUAD/LUSC NRF2 signature, and 6 genes specific for the LUSC NRF2 signature ( $\log_2FC > 0.8$ ) were found.

Finally, 3% of patients with esophageal adenocarcinoma (EAC) harbor *KEAP1* mutations, while 10% of patients harbor *NFE2L2* mutations. In EAC, NRF2 expression promotes tumor cells' survival, and, in addition, it was demonstrated that NRF2 has a protective role against stress-triggered apoptosis and ferroptosis<sup>44</sup>. ASTUTE revealed the upregulation of 35 genes of the LUAD/LUSC NRF2 signature in association with *NFE2L2* mutations ( $\log_2FC > 0.9$ ), and 13 genes of LUSC NRF2 signature ( $\log_2FC > 1$ ) in the EAC (Supplementary Data 8). We were not able to find similar results for *KEAP1* mutations, in which only the upregulation of *KYNU* and *B4GALNT1* was found.

Overall, ASTUTE's results demonstrate high consistency across cancer types and reveal a widespread expression of the identified NRF2 signatures at the pan-cancer level.

### NRF2 expression as a prognostic biomarker at the pan-cancer level

To assess the prognostic implications of the identified expression signatures, we considered the overall survival (OS) data provided by the Pan-Cancer Atlas studies<sup>37</sup>. We initially focused on genes associated with *KEAP1* or *NFE2L2* mutations in LUAD and LUSC, filtering for  $\log_2$  fold change values greater than 1 and less than  $-1$ . We applied the same filtering criteria to the six other considered cancer types as described before, and we narrowed down our analysis to the genes identified by ASTUTE in LUAD or LUSC and in these cancer types. Subsequently, we conducted standard univariate Cox regression analysis across all eight cancers (LUAD, LUSC, and the six other cancer types), considering the selected genes. Only genes exhibiting consistent correlation with risk, as indicated by ASTUTE, were considered. For instance, if ASTUTE suggested that *KEAP1/NFE2L2* mutations positively impacted gene expression, such genes were classified as oncogenes, and its higher expression had to positively correlate with worse prognosis. This initial screening yielded a list of genes strongly associated with both ASTUTE results and prognosis. We then proceeded to conduct multivariate regularized Cox regression analysis, identifying 14 genes whose expression significantly correlated with negative prognosis in at least one of the analyzed cancers (Table 3 and Supplementary Data 9). Several of the 14 genes comprising the prognostic NRF2 signature, such as *SRXN1* and *CABYR*, are known canonical NRF2 targets, while *TRIM16* has been reported to modulate NRF2 activity via the p62-KEAP1 complex. Notably, these genes were identified through our mutation-centric framework independently of prior pathway annotation, reinforcing the biological validity of the ASTUTE-derived signature and its ability to recover functionally relevant targets. Importantly, this signature was derived through a de novo mutation-centric approach and validated across multiple cancer types, offering both mechanistic insight and potential clinical applicability.

Interestingly, several of the identified genes have been reported as NRF2 target genes in the literature and belong to the two signatures identified in NSCLC. Notable examples include *CABYR*, which is upregulated in HCC and suggested as a cancer-testis antigen in lung cancer<sup>45</sup>. *GCLM* has been implicated as a tumor promoter and immunological biomarker in bladder cancer<sup>46</sup> and linked to cisplatin resistance in NSCLC<sup>47</sup>. *ME1* is associated with poor prognosis in HCC<sup>48</sup> and breast cancer<sup>49</sup>, while *NQO1* is significantly associated with prognosis, immune infiltrates, and drug resistance across multiple cancer types<sup>50</sup>. *SRXN1* is identified as an inducer of hepatocellular carcinogenesis and metastasis, correlating with poor

**Table 3 | NRF2 prognostic biomarkers**

NRF2 prognostic biomarkers
AKR1B15
CABYR
CES1
EGF
GCLM
HHIPL2
ME1
NQO1
PANX2
SLC7A11
SPP1
SRXN1
TRIM16L
TXNRD1

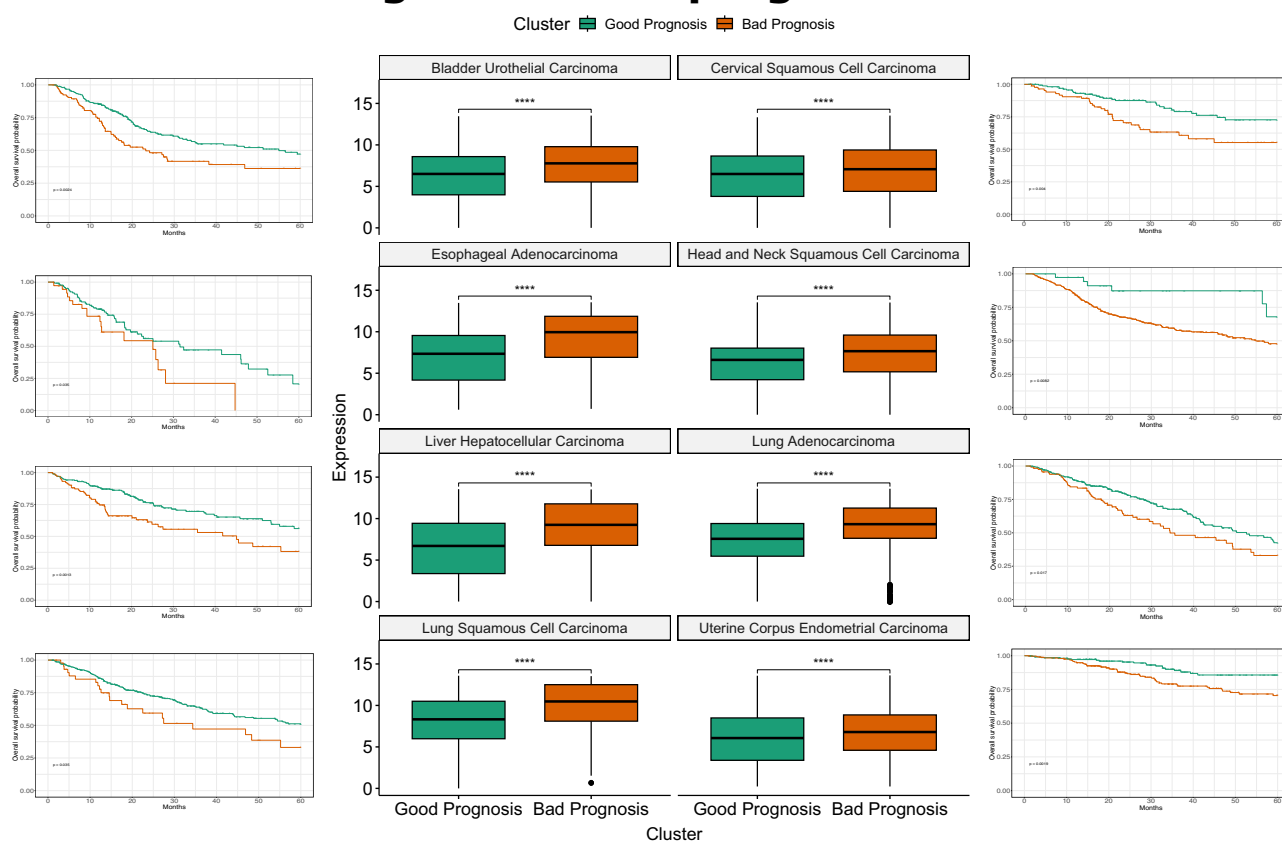
Genes associated to bad prognosis at the pan-cancer level.

prognosis in HCC patients<sup>51</sup>. *TXNRD1* is an unfavorable prognostic biomarker in HCC<sup>52</sup>, breast cancer<sup>53</sup>, and NSCLC<sup>54</sup>. Additionally, *SPP1*, associated with *NFE2L2* mutations in LUSC, is reported as a prognostic biomarker in urothelial bladder cancer<sup>55</sup> and ovarian cancer<sup>56</sup>. These findings highlight the relevance of these genes in NRF2 pathway activation and their potential as significant biomarkers across various cancer types.

We then used the hazard ratios estimated by the regularized Cox multivariate regression for the identified 14 genes to compute a risk score for each patient of the 8 cancer subtypes (see “Methods”). Hierarchical clustering was performed considering the computed risk scores, resulting in the classification of patients into two risk groups within each cancer subtype. Further Kaplan-Meier analysis confirmed significant differences in prognosis for the risk groups in all cancers. In particular, in BLCA ( $p = 0.0024$ ), CSCC ( $p = 0.004$ ), EAC ( $p = 0.035$ ), HNSCC ( $p = 0.0082$ ), HCC ( $p = 0.0013$ ), LUAD ( $p = 0.017$ ), LUSC ( $p = 0.035$ ), and UCEC ( $p = 0.0019$ ). Moreover, all the 14 genes were overexpressed in the patients with worse prognosis in all cancer types (Fig. 2).

We finally validated our findings on external cohorts by replicating the analysis using the LUAD dataset from Chen et al.<sup>13</sup> (Supplementary Fig. 4), primarily comprising patients with *EGFR* mutations, and the one from Pleasance et al.<sup>57</sup> (Supplementary Fig. 5), comprising metastatic lung

## NRF2 signature as a prognostic biomarker



**Fig. 2 | Impact of the NRF2 expression signatures based on the identified 14 prognostic genes across eight cancer types.** Bladder Cancer (BLCA), Cervical Squamous Cell Carcinoma (CSCC), Esophageal Adenocarcinoma (EAC), Head and Neck Squamous Cell Carcinoma (HNSCC), Hepatocellular Carcinoma (HCC), Lung Adenocarcinoma (LUAD), Lung Squamous Cell Carcinoma (LUSC), and Uterine Corpus Endometrial Carcinoma (UCEC). We show boxplots displaying the expression levels of the 14 prognostic genes in patients categorized into two risk groups based on hierarchical clustering of computed risk scores. The higher expression levels of these genes are associated with the worse prognosis group across all cancer types. Kaplan–Meier survival curves illustrate significant differences in

overall survival (OS) between the two risk groups within each cancer type. The  $p$ -values for each cancer type are as follows: BLCA ( $p = 0.0024$ , low-risk group  $n = 246$ , high-risk group  $n = 107$ ), CSCC ( $p = 0.004$ , low-risk group  $n = 173$ , high-risk group  $n = 89$ ), EAC ( $p = 0.035$ , low-risk group  $n = 124$ , high-risk group  $n = 37$ ), HNSCC ( $p = 0.0082$ , low-risk group  $n = 38$ , high-risk group  $n = 448$ ), HCC ( $p = 0.0013$ , low-risk group  $n = 229$ , high-risk group  $n = 102$ ), LUAD ( $p = 0.017$ , low-risk group  $n = 336$ , high-risk group  $n = 117$ ), LUSC ( $p = 0.035$ , low-risk group  $n = 383$ , high-risk group  $n = 45$ ), and UCEC ( $p = 0.0019$ , low-risk group  $n = 220$ , high-risk group  $n = 247$ ).

cancers. Hierarchical clustering was performed based on the risk scores computed from the 14 identified prognostic genes. Kaplan–Meier analysis confirmed the presence of two distinct risk groups. Although the two curves did not reach statistical significance in the dataset from Chen et al. possibly due to the overrepresentation of patients carrying *EGFR* mutations, which correlated with a limited number of patients in the cluster exhibiting the NRF2 signature and the shortest OS ( $p$  value = 0.064), the trend was consistent. Instead, in the dataset from Pleasance et al., including metastatic cancers, the discovered NRF2 prognostic signature identified two clusters with clearly different prognosis ( $p$  = 0.026). In the cluster with the worse prognosis, the 14 genes comprising the NRF2 signature were significantly upregulated, and a significantly higher frequency of *KEAP1* mutations was observed (0% vs 33%,  $p$  value adjusted for false discovery rate = 0.027), hence validating our results.

### Mutational landscape associated with the NRF2 prognostic signature

To further elucidate the mutational landscape underlying the identified NRF2 prognostic signature, we examined the clusters with significantly different survival across the considered eight cancer types. Patients were stratified into two groups in each cancer type—those with better prognosis and those with worse prognosis—based on their pan-cancer NRF2 prognostic signature scores. Consistently, we observed a significantly higher expression of the pan-cancer NRF2 prognostic signature in the worse prognosis clusters across all cancer types. Subsequently, we conducted a differential mutational analysis between the two survival-based clusters within each cancer type. Using a  $z$ -score test for proportions, corrected for false discovery rate ( $p$  < 0.05), we identified specific genetic alterations that were significantly enriched in the worse prognosis clusters. These alterations include mutations in key driver genes and other genetic events potentially contributing to the aggressive phenotypes observed (Fig. 3). Detailed results of the differential mutational analysis are provided in Supplementary Data 10.

In LUAD, mutations in *KEAP1* (5.8% vs 53.4%) and *NFE2L2* (2.1% vs 6.9%) are enriched in the cluster with higher NRF2 prognostic signature, thus playing a central role in activating the NRF2 pathway, potentially disrupting its regulatory mechanisms. This disruption likely leads to enhanced antioxidant responses, contributing to drug resistance and metabolic reprogramming that supports tumor cell survival under oxidative stress. Additionally, enriched mutations in *STK11* (10.1% vs 26.7%) and *ROS1* (4.3% vs 12.1%) may further contribute to NRF2 pathway activation, although the precise mechanisms require further investigation. Furthermore, we found enrichment of copy number gains in *IDH2* (9.8% vs 22.4%) in the LUAD patients with high NRF2 expression, a gene potentially impacting tumor metabolism and survival<sup>58,59</sup>. Similarly, in LUSC, we observed mutations in *KEAP1* (9% vs 23.3%) and *NFE2L2*. This activation enhances antioxidant defenses, which could support tumor progression. Concurrent observed enriched mutations in genes like *ARID5B* (2.2% vs 11.6%) and *TAF1* (4.6% vs 16.3%) might augment NRF2 activity, facilitating tumor growth and resistance to therapeutic interventions<sup>60,61</sup>.

Other cancer types, such as BLCA, CESC, and EAC, also exhibit NRF2 activation influenced by specific mutational profiles. In BLCA, mutations in *NFE2L2* (2.9% vs 15.1%) are enriched, suggesting a potential enhancement of antioxidant defenses. Additionally, enriched copy number loss in *PPARG* (9.9% vs 20.8%) may influence NRF2 activity indirectly through its regulatory networks, affecting metabolic and stress response pathways<sup>62,63</sup>. In CESC, enriched mutations in *ADAMTS9* (0.6% vs 6.2%) and *AFF1* (0% vs 6.2%) might augment NRF2-mediated antioxidant responses, aiding in cell survival under oxidative stress conditions. The enriched demethylation of *CST6* (9.9% vs 21%) could potentially affect NRF2 regulation, although its specific interaction requires further investigation. In EAC, *NFE2L2* (3.2% vs 37.8%) mutations are also enriched. Additionally, enriched mutations in *MST1R* (0% vs 8.1%) and *ZNF217* (0.8% vs 10.8%) might suggest potential pathways through which NRF2 activity could be modulated, influencing cellular proliferation and survival mechanisms<sup>64</sup>.

In HNSCC patients with high NRF2 signature show higher frequencies of point mutations in *CDKN2A* (5.7% vs 22.4%) and *TP53* (34.3% vs 74.4%). While in HCC, the enriched mutations in *KEAP1* (1.4% vs 12.5%) and *NFE2L2* (1.4% vs 8.3%), along with copy number loss in *ARID2* (6.9% vs 20.8%), suggest a significant association with NRF2 activation and its regulatory pathways<sup>65,66</sup>.

Finally, in UCEC, several key genetic alterations are enriched, underscoring the potential implications of NRF2 pathway activation in this cancer. Mutations in *TP53* (24.2% vs 51.5%) are known to disrupt redox homeostasis, leading to increased oxidative stress within cells. This can activate NRF2 as a protective mechanism, enhancing antioxidant defenses and promoting cell survival. Concurrently, amplification of *AKT2* (6.6% vs 21.5%), a regulator of growth factor signaling, may directly phosphorylate NRF2, stabilizing its protein levels and promoting transcription of anti-oxidant genes<sup>67</sup>. Additionally, copy number loss in *AXIN1* (5.7% vs 21.9%), associated to the activation of the Wnt signaling pathways, might suggest a potential indirect modulation of NRF2 activity through crosstalk mechanisms<sup>68</sup>. Furthermore, methylation of *TRIM31* (10% vs 20.7%), through its direct interaction with NRF2, requires further elucidation, likely influences gene expression profiles involved in oxidative stress responses<sup>69</sup>. Together, these genetic alterations in UCEC collectively highlight diverse pathways through which NRF2 overactivation may contribute to tumor progression, emphasizing the need for targeted therapeutic strategies aimed at disrupting NRF2-dependent oncogenic processes.

Overall, while NRF2 pathway activation appears to be a common feature across multiple cancer types, the specific genetic alterations influencing its activity can vary. These findings highlight the complexity of NRF2-mediated mechanisms in cancer and underscore the need for further research to elucidate these pathways fully, potentially informing targeted therapeutic strategies aimed at disrupting NRF2-dependent oncogenic processes.

### Discussion

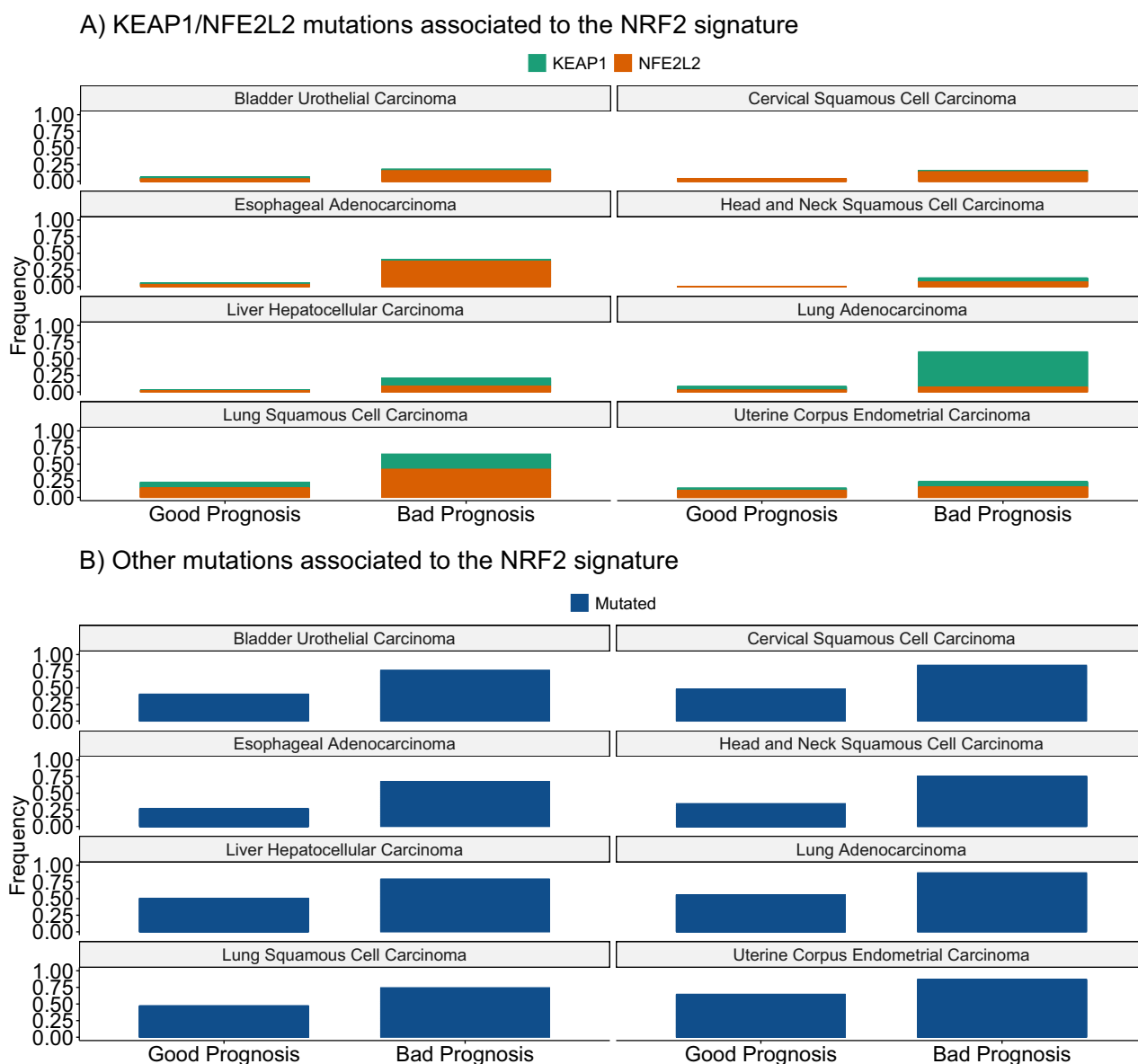
In this study, we employed the ASTUTE framework to conduct a comprehensive analysis of *KEAP1* and *NFE2L2* mutations in NSCLC, focusing on their influence on NRF2 expression and their potential as prognostic biomarkers. Our findings suggest a significant association between genetic alterations in the *KEAP1*/NRF2 pathway and NSCLC progression, offering valuable insights into the underlying mechanisms of pathogenesis and drug resistance.

Our analysis revealed a robust association between *KEAP1*/*NFE2L2* mutations and the upregulation of NRF2 target genes critical in antioxidant response, detoxification, and metabolism. Notably, *KEAP1* mutations were prevalent in both LUAD and LUSC subtypes, whereas *NFE2L2* mutations were more frequent in LUSC. This highlights the necessity of subtype-specific considerations in understanding NRF2's role in NSCLC and suggests implications for personalized therapeutic interventions.

Furthermore, we identified specific gene signatures associated with *KEAP1*/*NFE2L2* mutations that could serve as prognostic biomarkers in NSCLC. For instance, upregulation of the GSTM family members and UDP-glucuronosyltransferases in LUSC patients harboring these mutations suggests enhanced detoxification capacity, potentially influencing treatment response and outcomes. Moreover, genes selectively upregulated in LUAD or LUSC due to *KEAP1* mutations, such as *CPLX2* and *KYNU*, underscore the subtype-specific activation of NRF2 and its implications for disease prognosis.

We also unveiled intriguing associations between NRF2 expression signatures and other key genetic alterations, notably *TP53* mutations. Several genes upregulated in LUSC patients with *KEAP1*/*NFE2L2* mutations showed positive associations with *TP53* mutations as well, hinting at potential crosstalk between the NRF2 and *TP53* pathways in driving tumorigenesis and drug resistance. Understanding these interactions could unveil novel therapeutic targets and combination strategies for NSCLC treatment.

Importantly, we conducted a comprehensive survival analysis revealing a set of 14 genes consistently associated with prognosis across different



**Fig. 3 | Differential mutation frequencies associated with the NRF2 prognostic signature across eight cancer types.** In **A** we show barplots representing the frequency of KEAP1 and NFE2L2 mutations in patients with worse prognosis clusters compared to those with better prognosis clusters across the eight studied cancer types: Lung Adenocarcinoma (LUAD), Lung Squamous Cell Carcinoma (LUSC), Bladder Cancer (BLCA), Cervical Squamous Cell Carcinoma (CSCC), Esophageal Adenocarcinoma (EAC), Head and Neck Squamous Cell Carcinoma (HNSCC), Hepatocellular Carcinoma (HCC), and Uterine Corpus Endometrial Carcinoma (UCEC). KEAP1 and NFE2L2 mutations were particularly enriched in LUAD, LUSC, and EAC, indicating a significant role in activating the NRF2 pathway and

contributing to worse prognosis in these cancer types. In **B** we show the aggregated frequency of other significant genetic alterations in the worse prognosis clusters compared to better prognosis clusters in the same eight cancer types. These mutations include, but are not limited to, STK11, ROS1, IDH2, ARID5B, TAF1, ADAMTS9, AFF1, MST1R, ZNF217, CDKN2A, TP53, ARID2, AKT2, AXIN1, and TRIM31. These alterations might suggest alternative mechanisms of NRF2 activation and regulation, highlighting the complexity of NRF2 pathway dysregulation across different cancers and emphasizing the necessity for diverse therapeutic strategies.

cancers, underscoring the NRF2 signature's robustness as a prognostic biomarker at the pan-cancer level.

While our pan-cancer analysis demonstrates that the NRF2 expression signature identified by ASTUTE is consistently associated with prognosis across diverse tumor types, we acknowledge that NRF2 signaling might be context- and tissue-specific. Previous studies have shown that super-enhancer regulation might play a critical role in modulating NRF2 transcriptional activity, with implications for lineage-restricted gene expression programs. Additionally, hypoxia and other tumor microenvironmental factors are known to influence NRF2 activity and its downstream effects.

Although our current analysis does not incorporate enhancer-state data or microenvironmental features such as oxygen tension, the reproducibility of our mutation-derived signature across squamous and adenocarcinoma subtypes suggests that it captures a broadly conserved core of NRF2 pathway activation. Future studies incorporating chromatin accessibility and single-cell resolution datasets could further refine the context-specific components of NRF2 signaling.

Several NRF2-related gene expression signatures have been proposed in previous studies, often derived via clustering or curated pathway annotations. In contrast, our work builds a prognostic signature through a mutation-centric



approach that directly links *KEAP1*/*NFE2L2* alterations to expression profiles, enabling mechanistic attribution. The identification of well-established *NRF2* targets such as *SRXN1* and *CABYR*, as well as *NRF2* modulators like *TRIM16*, within our expression signature, derived solely from mutation-based associations, supports the capacity of *ASTUTE* to capture mechanistically grounded transcriptional programs without relying on prior pathway knowledge. This methodological distinction enhances the mechanistic interpretability and relevance of the signature for precision oncology.

The mutational landscape associated with the *NRF2* prognostic signature revealed specific genetic alterations across various cancer types that significantly enrich *NRF2* activity, promoting tumor progression, drug resistance, and metabolic reprogramming. These findings not only deepen our understanding of *KEAP1* and *NFE2L2* mutations in NSCLC but also hold promise for guiding therapeutic strategies and advancing precision oncology tailored to individual patient subtypes. Further validation in external patient cohorts is warranted to substantiate these observations and elucidate the molecular mechanisms driving *NRF2* activation in NSCLC.

It is important to note that in certain tumor types, such as EAC, *KEAP1* and *NFE2L2* mutations are relatively infrequent. This low prevalence poses challenges in terms of statistical power and increases the risk of spurious associations. To address this, we applied bootstrapped regularized regression and false discovery rate (FDR) correction to reduce false positives. Moreover, gene-level associations identified in low-frequency mutation contexts were interpreted cautiously and primarily emphasized when supported by consistent trends across multiple datasets or cancer types. This limitation underscores the need for validation in larger or prospectively curated cohorts.

Another limitation of our study is its reliance on retrospective, publicly available datasets, which may introduce confounding variables such as heterogeneity in sequencing technologies, treatment regimens, and clinical annotations. These sources of variability may affect the generalizability of our findings and should be carefully considered when interpreting the results.

Moreover, while our 14-gene *NRF2* prognostic signature demonstrates robust associations with patient outcomes across multiple cancer types, its clinical utility must ultimately be evaluated in the context of established biomarkers such as PD-L1 expression and tumor mutational burden. Unfortunately, the retrospective datasets used in this study lack standardized and complete annotations for these variables, limiting our ability to perform comparative multivariate Cox regression models. Future prospective studies with harmonized clinical, immunological, and molecular profiling will be essential to assess the incremental prognostic value and potential complementarity of our *NRF2* signature relative to existing biomarkers. Such efforts will also help determine its cost-effectiveness and practical feasibility for integration into clinical decision-making.

In conclusion, our study demonstrates the clinical relevance of *NRF2* activation pathways in NSCLC and highlights the potential for personalized treatment approaches based on subtype-specific molecular profiles. By elucidating the complex interplay between genetic alterations and the *NRF2*-driven pathways, our findings pave the way for innovative therapeutic avenues and enhance our broader understanding of cancer biology and precision medicine.

## Methods

### Input datasets

We considered genomic and transcriptomic data from five distinct datasets comprising NSCLC patients, specifically LUAD and LUSC. For LUAD, datasets included genomic profiles from the CPTAC (110 samples)<sup>11</sup>, the Pan-Cancer Atlas (510 samples)<sup>12</sup>, and Chen et al. (169 samples)<sup>13</sup>, totaling 789 patients. LUSC datasets encompassed samples from CPTAC (108 samples)<sup>14</sup> and the Pan-Cancer Atlas (481 samples)<sup>14</sup>, comprising 589 patients. To minimize batch effects, each dataset was analyzed independently using *ASTUTE* to identify associations between somatic mutations and gene expression profiles, without direct data merging or cross-cohort normalization. This ensured that dataset-specific technical variability did not confound the genotype-to-expression associations.

Additionally, *ASTUTE* was also utilized to investigate the pan-cancer implications of *NRF2* gene expression signatures in cancers where *KEAP1* or *NFE2L2* mutations were observed in at least  $\geq 5\%$  of the patients, including HCC, HNSCC, UCEC, CSCC, BLCA, and EAC from the Pan-Cancer Atlas studies<sup>37</sup>.

To validate our findings, we conducted a further analysis using the LUAD datasets from Chen et al.<sup>13</sup> and from Pleasance et al.<sup>57</sup>. The data are publicly available and were downloaded from cBioPortal<sup>70</sup>.

### The *ASTUTE* framework

*ASTUTE* is a computational framework designed to integrate somatic mutation data with gene expression profiles to elucidate their functional implications in cancer biology. It employs a regularized regression model using the LASSO penalty within a linear regression framework<sup>10</sup>. The LASSO penalty introduces a regularization parameter ( $\lambda$ ), which is optimized through  $k$ -fold cross-validation, typically with  $k = 10$ , to control model complexity, promote feature selection, and minimize prediction error. In particular, the regularization parameter used in the LASSO model by *ASTUTE* was selected via 10-fold cross-validation. All regularized regression procedures were implemented using the *glmnet* R package.

To enhance robustness, *ASTUTE* repeats this procedure multiple times (e.g., 100 iterations) using bootstrap resampling of the input data. The algorithm is reapplied in each iteration, and the final results are computed through bootstrap aggregation. This approach enables *ASTUTE* to identify gene expression features most significantly associated with somatic mutations in driver genes. Additionally, *ASTUTE* estimates baseline expression levels for each gene and calculates fold changes to quantify the impact of specific mutations. Finally, the fold changes computed at each bootstrap iteration are used to estimate whether they are significantly lower or greater than 1, enabling the calculation of  $p$  values that provide confidence estimates of the associated mutations either increasing or decreasing the expression of specific genes. In particular, for each gene,  $p$  values were derived from the distribution of fold changes obtained across bootstrap iterations. These  $p$  values were adjusted for multiple testing using the Benjamini-Hochberg FDR correction method. The R *glmnet* packages was used to implement the regularized regression procedures.

### Survival analysis

Prior to survival modeling, we applied a preliminary log<sub>2</sub> fold-change threshold ( $\pm 1$ ) as a conservative filter to retain genes with biologically meaningful expression changes associated with *KEAP1*/*NFE2L2* mutations. This step reduced noise and focused the analysis on robust candidates. We then employed regularized Cox regression with LASSO penalty to investigate the prognostic impact of gene expression profiles on OS in various cancer types. We utilized the R *glmnet* package to implement regularized Cox regression. This technique facilitates variable selection by shrinking the coefficients of the less relevant predictors to zero, thereby highlighting the most significant associations between gene expression and survival outcomes. Cross-validation was employed to determine the optimal penalization parameter ( $\lambda$ ) that minimized prediction error and improved the robustness of our survival predictions. For each patient, we calculated risk scores based on the estimated hazard ratios of gene expression changes identified through LASSO regression. These risk scores were used to stratify patients into different risk groups, enabling the identification of high-risk subgroups with poorer survival outcomes. The clustering analysis was performed by hierarchical clustering using the dynamic tree cut approach<sup>71</sup> implemented by the *dynamictreecut* R package to group patients based on their risk scores. Kaplan-Meier survival curves were generated to assess the differences in survival between risk groups, and statistical significance was determined using the log-rank test.

### Mutations enrichment analysis

We conducted an enrichment analysis to assess differences in the proportions of mutation occurrences across predefined clusters. We consider the clusters obtained from the survival analysis and computed the proportion of

mutations in the set of genes of interest. To determine statistical significance, we employed a  $z$ -test to compare these proportions, adjusted for multiple comparisons using false discovery rate correction. Genes were considered significantly enriched if they exhibited a  $q$  value  $< 0.05$ , indicating significance after controlling for false discovery. We further assessed the alterations that can directly influence gene expression, such as copy number gains and demethylations, which can result in increased gene expression and copy number losses and methylations, leading to decreased gene expression. To validate their impact on gene expression levels, we conducted  $t$ -tests ( $p < 0.05$ ) to confirm the significant impact of the enriched alterations to expression. This approach allowed us to systematically evaluate and compare mutation patterns across distinct clusters associated to the NRF2 pathway expression.

### RNA extraction, reverse transcription, and quantitative real-time PCR

H2228 and H3122 NSCLC cell lines were cultured in biological triplicates and seeded at a density of  $3 \times 10^6$  cells per T75 flask. Cells were harvested at ~70% confluence. Total RNA was extracted using the RNeasy Mini Kit (QIAGEN, Germany) following the manufacturer's instructions. Two micrograms of RNA were reverse-transcribed using the LunaScript RT SuperMix (Euroclone, Italy) in a final reaction volume of 40  $\mu$ L. Quantitative PCR was performed in technical duplicates using 2  $\mu$ L of cDNA and 2X Mastermix (GeneSpin, Italy) on a QuantStudio™ Real-Time PCR system (Life Technologies). Gene expression levels were normalized to GAPDH and calculated using the  $2^{-\Delta\Delta Ct}$  method. Target genes included CYP4F11, AKR1C1, AKR1C2, AKR1C3, AKR1B10, CYP4F3, GPX2, CABYR, JAKMIP3, and UCHL1, using commercially available TaqMan assays.

### Software and statistical analysis

ASTUTE is available as an R package and can be installed from GitHub (<https://github.com/ramazzottitlab/ASTUTE>). All computational analyses were conducted using R (version 4.4.1). Key R packages included *glmnet* for the regularized regression analyses, *dynamicTreeCut* for the clustering analysis, and the *survival* package for survival analysis. Statistical significance was determined at  $p < 0.05$ , unless otherwise specified.

### Data availability

Data is provided within the manuscript or supplementary information files. ASTUTE is available as an R package and can be installed from GitHub (<https://github.com/ramazzottitlab/ASTUTE>). All computational analyses were conducted using R (version 4.4.1). Key R packages included *glmnet* for the regularized regression analyses, *dynamicTreeCut* for the clustering analysis, and the *survival* package for survival analysis.

Received: 23 February 2025; Accepted: 9 August 2025;

Published online: 18 August 2025

### References

- Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **553**, 446–454 (2018).
- Mitsudomi, T. & Yatabe, Y. Mutations of the epidermal growth factor receptor gene and related genes as determinants of epidermal growth factor receptor tyrosine kinase inhibitors sensitivity in lung cancer. *Cancer Sci.* **98**, 1817–1824 (2007).
- Budczies, J. et al. KRAS and TP53 co-mutation predicts benefit of immune checkpoint blockade in lung adenocarcinoma. *Br. J. Cancer* **131**, 524–533 (2024).
- Campbell, J. D. et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).
- Frank, R. et al. Clinical and pathological characteristics of. *Clin. Cancer Res.* **24**, 3087–3096 (2018).
- Zavitsanou, A. M. et al. KEAP1 mutation in lung adenocarcinoma promotes immune evasion and immunotherapy resistance. *Cell Rep.* **42**, 113295 (2023).
- Negrao, M. V. et al. Comutations and KRASG12C inhibitor efficacy in advanced NSCLC. *Cancer Discov.* **13**, 1556–1571 (2023).
- Scalera, S. et al. KEAP1-mutant NSCLC: the catastrophic failure of a cell-protecting hub. *J. Thorac. Oncol.* **17**, 751–757 (2022).
- Henderson, A. R. The bootstrap: a technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clin. Chim. Acta* **359**, 1–26 (2005).
- Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
- Gillette, M. A. et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *e35* **182**, 2020–225.e35 (2020).
- Network CGAR. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
- Chen, J. et al. Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* **52**, 177–186 (2020).
- Satpathy, S. et al. A proteogenomic portrait of lung squamous cell carcinoma. *Cell* **184**, 4348–4371.e40 (2021).
- Network CGAR. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
- Yang, H. et al. Nrf1 and Nrf2 regulate rat glutamate-cysteine ligase catalytic subunit transcription indirectly via NF-kappaB and AP-1. *Mol. Cell Biol.* **25**, 5933–5946 (2005).
- Tonelli, C., Chio, I. I. C. & Tuveson, D. A. Transcriptional regulation by Nrf2. *Antioxid. Redox Signal.* **29**, 1727–1745 (2018).
- Ma, Q. Role of Nrf2 in oxidative stress and toxicity. *Annu. Rev. Pharm. Toxicol.* **53**, 401–426 (2013).
- He, F., Antonucci, L. & Karin, M. NRF2 as a regulator of cell metabolism and inflammation in cancer. *Carcinogenesis* **41**, 405–416 (2020).
- Lacher, S. E. & Slattery, M. Gene regulatory effects of disease-associated variation in the NRF2 network. *Curr. Opin. Toxicol.* **1**, 71–79 (2016).
- Jena, K. K. et al. TRIM16 controls assembly and degradation of protein aggregates by modulating the p62-NRF2 axis and autophagy. *EMBO J.* **37**, e98358 (2018).
- Vollrath, V., Wielandt, A. M., Iruretagoyena, M. & Chianale, J. Role of Nrf2 in the regulation of the Mrp2 (ABCC2) gene. *Biochem. J.* **395**, 599–609 (2006).
- Feng, L. et al. SLC7A11 regulated by NRF2 modulates esophageal squamous cell carcinoma radiosensitivity by inhibiting ferroptosis. *J. Transl. Med.* **19**, 367 (2021).
- Shakya, A., McKee, N. W., Dodson, M., Chapman, E. & Zhang, D. D. Anti-ferroptotic effects of Nrf2: beyond the antioxidant response. *Mol. Cells* **46**, 165–175 (2023).
- Namani, A., Zheng, Z., Wang, X. J. & Tang, X. Systematic identification of multi omics-based biomarkers in. *J. Cancer* **10**, 6813–6821 (2019).
- Li, H. et al. CPLX2 regulates ferroptosis and apoptosis through nrf2 pathway in human hepatocellular carcinoma cells. *Appl. Biochem. Biotechnol.* **195**, 597–609 (2023).
- Komatsu, H. et al. Complexin-2 (CPLX2) as a potential prognostic biomarker in human lung high grade neuroendocrine tumors. *Cancer Biomark.* **13**, 171–180 (2013).
- León-Letelier, R. A. et al. Kynureninase upregulation is a prominent feature of NRF2-activated cancers and is associated with tumor immunosuppression and poor prognosis. *Cancers* **15**, <https://doi.org/10.3390/cancers15030834> (2023).
- Fahrmann, J. F. et al. Mutational Activation of the NRF2 pathway upregulates kynureninase, resulting in tumor immunosuppression and poor outcome in lung adenocarcinoma. *Cancers* **14**, <https://doi.org/10.3390/cancers14102543> (2022).

30. Chien, M. H. et al. Keap1-Nrf2 interaction suppresses cell motility in lung adenocarcinomas by targeting the S100P protein. *Clin. Cancer Res.* **21**, 4719–4732 (2015).
31. Hsu, Y. L. et al. S100P interacts with integrin  $\alpha 7$  and increases cancer cell migration and invasion in lung cancer. *Oncotarget* **6**, 29585–29598 (2015).
32. He, X. et al. SERPINB5 is a prognostic biomarker and promotes proliferation, metastasis and epithelial-mesenchymal transition (EMT) in lung adenocarcinoma. *Thorac. Cancer* **14**, 2275–2287 (2023).
33. Nebert, D. W. & Vasiliou, V. Analysis of the glutathione S-transferase (GST) gene family. *Hum. Genom.* **1**, 460–464 (2004).
34. Chanas, S. A. et al. Loss of the Nrf2 transcription factor causes a marked reduction in constitutive and inducible expression of the glutathione S-transferase Gsta1, Gsta2, Gstm1, Gstm2, Gstm3 and Gstm4 genes in the livers of male and female mice. *Biochem. J.* **365**, 405–416 (2002).
35. Rowland, A., Miners, J. O. & Mackenzie, P. I. The UDP-glucuronosyltransferases: their role in drug metabolism and detoxification. *Int. J. Biochem. Cell Biol.* **45**, 1121–1132 (2013).
36. Moinova, H. R. & Mulcahy, R. T. Up-regulation of the human gamma-glutamylcysteine synthetase regulatory subunit gene involves binding of Nrf-2 to an electrophile responsive element. *Biochem. Biophys. Res. Commun.* **261**, 661–668 (1999).
37. Blum, A., Wang, P. & Zenklusen, J. C. SnapShot: TCGA-analyzed tumors. *Cell* **173**, 530 (2018).
38. Li, C., Liang, G., Yan, K. & Wang, Y. NRF2 mutation enhances the immune escape of hepatocellular carcinoma by reducing STING activation. *Biochem. Biophys. Res. Commun.* **698**, 149536 (2024).
39. Zhang, M. et al. Nrf2 is a potential prognostic marker and promotes proliferation and invasion in human hepatocellular carcinoma. *BMC Cancer* **15**, 531 (2015).
40. Osman, A. A. et al. Dysregulation and epigenetic reprogramming of NRF2 signaling axis promote acquisition of cisplatin resistance and metastasis in head and neck squamous cell carcinoma. *Clin. Cancer Res.* **29**, 1344–1359 (2023).
41. Zhang, M. et al. The promoting effect and mechanism of Nrf2 on cell metastasis in cervical cancer. *J. Transl. Med.* **21**, 433 (2023).
42. Chen, N. et al. Nrf2 expression in endometrial serous carcinomas and its precancers. *Int. J. Clin. Exp. Pathol.* **4**, 85–96 (2010).
43. Hayden, A. et al. The Nrf2 transcription factor contributes to resistance to cisplatin in bladder cancer. *Urol. Oncol.* **32**, 806–814 (2014).
44. Zhu, J. et al. Ferroptosis: a new mechanism of traditional Chinese medicine for cancer treatment. *Front. Pharmacol.* **15**, 1290120 (2024).
45. Li, H., Fang, L., Xiao, X. & Shen, L. The expression and effects the CABYR-c transcript of CABYR gene in hepatocellular carcinoma. *Bull. Cancer* **99**, E26–E33 (2012).
46. Wang, S., Wang, H., Zhu, S. & Li, F. Systematical analysis of ferroptosis regulators and identification of GCLM as a tumor promotor and immunological biomarker in bladder cancer. *Front. Oncol.* **12**, 1040892 (2022).
47. Fujimori, S. et al. The subunits of glutamate cysteine ligase enhance cisplatin resistance in human non-small cell lung cancer xenografts in vivo. *Int. J. Oncol.* **25**, 413–418 (2004).
48. Wen, D. et al. Malic enzyme 1 induces epithelial-mesenchymal transition and indicates poor prognosis in hepatocellular carcinoma. *Tumour Biol.* **36**, 6211–6221 (2015).
49. Hu, W. C. et al. Overexpression of malic enzyme is involved in breast cancer growth and is correlated with poor prognosis. *J. Cell Mol. Med.* **28**, e18163 (2024).
50. Shen, L. et al. Pan-cancer and single-cell analysis reveal the prognostic value and immune response of NQO1. *Front. Cell Dev. Biol.* **11**, 1174535 (2023).
51. Lv, X. et al. SRXN1 stimulates hepatocellular carcinoma tumorigenesis and metastasis through modulating ROS/p65/BTG2 signalling. *J. Cell Mol. Med.* **24**, 10714–10729 (2020).
52. Fu, B. et al. TXNRD1 is an unfavorable prognostic factor for patients with hepatocellular carcinoma. *Biomed. Res. Int.* **2017**, 4698167 (2017).
53. Patwardhan, R. S., Rai, A., Sharma, D., Sandur, S. K. & Patwardhan, S. Txnrd1 as a prognosticator for recurrence, metastasis and response to neoadjuvant chemotherapy and radiotherapy in breast cancer patients. *Heliyon* **10**, e27011 (2024).
54. Delgobo, M. et al. Thioredoxin reductase-1 levels are associated with NRF2 pathway activation and tumor recurrence in non-small cell lung cancer. *Free Radic. Biol. Med.* **177**, 58–71 (2021).
55. Nedjadi, T., Ahmed, M. E., Ansari, H. R., Aouabdi, S. & Al-Maghribi, J. Identification of SPP1 as a prognostic biomarker and immune cells modulator in urothelial bladder cancer: a bioinformatics analysis. *Cancers* **15**, 5704 (2023).
56. Gao, W. et al. SPP1 is a prognostic related biomarker and correlated with tumor-infiltrating immune cells in ovarian cancer. *BMC Cancer* **22**, 1367 (2022).
57. Pleasance, E. et al. Pan-cancer analysis of advanced patient tumors reveals interactions between therapy and genomic landscapes. *Nat. Cancer* **1**, 452–468 (2020).
58. Liu, Y. et al. Targeting IDH1-mutated malignancies with NRF2 blockade. *J. Natl. Cancer Inst.* **111**, 1033–1041 (2019).
59. Liu, Y. et al. Protein kinase B (PKB/AKT) protects IDH-mutated glioma from ferroptosis via Nrf2. *Clin. Cancer Res.* **29**, 1305–1316 (2023).
60. Cheng, I. H. et al. TAF2, within the TFIID complex, regulates the expression of a subset of protein-coding genes. *Cell Death Discov.* **10**, 244 (2024).
61. Gouge, J. et al. Redox signaling by the RNA polymerase III TFIIB-related factor Brf2. *Cell* **163**, 1375–1387 (2015).
62. Lu, Y. et al. Oridonin exerts anticancer effect on osteosarcoma by activating PPAR- $\gamma$  and inhibiting Nrf2 pathway. *Cell Death Dis.* **9**, 15 (2018).
63. Lee, C. Collaborative power of Nrf2 and PPAR. *Oxid. Med. Cell Longev.* **2017**, 1378175 (2017).
64. He, K. et al. Genomic profiling reveals novel predictive biomarkers for chemo-radiotherapy efficacy and thoracic toxicity in non-small-cell lung cancer. *Front. Oncol.* **12**, 928605 (2022).
65. Nguyen, V., Schrank, T. P., Major, M. B. & Weissman, B. E. ARID1A loss is associated with increased NRF2 signaling in human head and neck squamous cell carcinomas. *PLoS ONE* **19**, e0297741 (2024).
66. Song, S. et al. Loss of SWI/SNF chromatin remodeling alters NRF2 signaling in non-small cell lung carcinoma. *Mol. Cancer Res.* **18**, 1777–1788 (2020).
67. Cahuzac, K. M. et al. AKT activation because of PTEN loss upregulates xCT via GSK3 $\beta$ /NRF2, leading to inhibition of ferroptosis in PTEN-mutant tumor cells. *Cell Rep.* **42**, 112536 (2023).
68. Rada, P. et al. WNT-3A regulates an Axin1/NRF2 complex that regulates antioxidant metabolism in hepatocytes. *Antioxid. Redox Signal.* **22**, 555–571 (2015).
69. Deng, N. H., Tian, Z., Zou, Y. J. & Quan, S. B. E3 ubiquitin ligase TRIM31: a potential therapeutic target. *Biomed. Pharmacother.* **176**, 116846 (2024).
70. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
71. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. *Bioinformatics* **24**, 719–720 (2008).
72. Kerins, M. J. & Ooi, A. The roles of NRF2 in modulating cellular iron homeostasis. *Antioxid. Redox Signal.* **29**, 1756–1773 (2018).
73. Jung, K. A. et al. Identification of aldo-keto reductases as NRF2-target marker genes in human cells. *Toxicol. Lett.* **218**, 39–49 (2013).
74. Chen, Y. et al. Hypoxia-induced ALDH3A1 promotes the proliferation of non-small-cell lung cancer by regulating energy metabolism reprogramming. *Cell Death Dis.* **14**, 617 (2023).



75. Miura, T., Taketomi, A., Nishinaka, T. & Terada, T. Regulation of human carbonyl reductase 1 (CBR1, SDR21C1) gene by transcription factor Nrf2. *Chem. Biol. Interact.* **202**, 126–135 (2013).
76. Namani, A. et al. NRF2-regulated metabolic gene signature as a prognostic biomarker in non-small cell lung cancer. *Oncotarget* **8**, 69847–69862 (2017).
77. Su, S., Yang, X. & Omiecinski, C. J. Intronic DNA elements regulate Nrf2 chemical responsiveness of the human microsomal epoxide hydrolase gene (EPHX1) through a far upstream alternative promoter. *Biochim. Biophys. Acta* **1839**, 493–505 (2014).
78. Singh, A. et al. Glutathione peroxidase 2, the major cigarette smoke-inducible isoform of GPX in lungs, is regulated by Nrf2. *Am. J. Respir. Cell Mol. Biol.* **35**, 639–650 (2006).
79. Harvey, C. J. et al. Nrf2-regulated glutathione recycling independent of biosynthesis is critical for cell survival during oxidative stress. *Free Radic. Biol. Med.* **46**, 443–453 (2009).
80. Nioi, P., McMahon, M., Itoh, K., Yamamoto, M. & Hayes, J. D. Identification of a novel Nrf2-regulated antioxidant response element (ARE) in the mouse NAD(P)H:quinone oxidoreductase 1 gene: reassessment of the ARE consensus sequence. *Biochem. J.* **374**, 337–348 (2003).
81. Brennan, M. S., Matos, M. F., Richter, K. E., Li, B. & Scannevin, R. H. The NRF2 transcriptional target, OSGIN1, contributes to monomethyl fumarate-mediated cytoprotection in human astrocytes. *Sci. Rep.* **7**, 42054 (2017).
82. Kim, Y. J. et al. Human prx1 gene is a target of Nrf2 and is up-regulated by hypoxia/reoxygenation: implication to tumor biology. *Cancer Res.* **67**, 546–554 (2007).
83. Sánchez-Rodríguez, R. et al. Ptgr1 expression is regulated by NRF2 in rat hepatocarcinogenesis and promotes cell proliferation and resistance to oxidative stress. *Free Radic. Biol. Med.* **102**, 87–99 (2017).
84. Zhou, J., Jiang, G., Xu, E., Liu, L. & Yang, Q. Identification of *SRXN1* and *KRT6A* as key genes in smoking-related non-small-cell lung cancer through bioinformatics and functional analyses. *Front. Oncol.* **11**, 810301 (2021).
85. Li, N. & Zhan, X. Machine learning identifies pan-cancer landscape of Nrf2 oxidative stress response pathway-related genes. *Oxid. Med. Cell Longev.* **2022**, 8450087 (2022).
86. Heiss, E. H., Schachner, D., Zimmermann, K. & Dirsch, V. M. Glucose availability is a decisive factor for Nrf2-mediated gene expression. *Redox Biol.* **1**, 359–365 (2013).
87. Dinkova-Kostova, A. T. & Abramov, A. Y. The emerging role of Nrf2 in mitochondrial function. *Free Radic. Biol. Med.* **88**, 179–188 (2015).
88. Zhao, J. et al. Nrf2 mediates metabolic reprogramming in non-small cell lung cancer. *Front. Oncol.* **10**, 578315 (2020).
89. Malhotra, D. et al. 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 (2010).
90. Xu, I. M. et al. Transketolase counteracts oxidative stress to drive cancer development. *Proc. Natl. Acad. Sci. USA* **113**, E725–E734 (2016).
91. Loboda, A., Damulewicz, M., Pyza, E., Jozkowicz, A. & Dulak, J. Role of Nrf2/HO-1 system in development, oxidative stress response and diseases: an evolutionarily conserved mechanism. *Cell Mol. Life Sci.* **73**, 3221–3247 (2016).
92. Song, M. O., Mattie, M. D., Lee, C. H. & Freedman, J. H. The role of Nrf1 and Nrf2 in the regulation of copper-responsive transcription. *Exp. Cell Res.* **322**, 39–50 (2014).
93. DeNicola, G. M. et al. NRF2 regulates serine biosynthesis in non-small cell lung cancer. *Nat. Genet.* **47**, 1475–1481 (2015).
94. Liao, D. et al. Identification of Pannexin 2 as a novel marker correlating with ferroptosis and malignant phenotypes of prostate cancer cells. *OncoTargets Ther.* **13**, 4411–4421 (2020).
95. Park, H. R. et al. Identification of novel NRF2-dependent genes as regulators of lead and arsenic toxicity in neural progenitor cells. *J. Hazard Mater.* **463**, 132906 (2024).

## Acknowledgements

L.M. acknowledges funding from the Italian Foundation for Cancer Research (AIRC) under IG 2020, ID 24828 project. Figure 1 was created with BioRender.com.

## Author contributions

Conceptualization and methodology: V.C., D.R. Software: D.R. Investigation: V.C., N.C., A.V., F.M., M.V., L.S., A.A., D.R. Visualization: V.C., N.C., F.M., D.R. Funding acquisition: L.M., D.R. Supervision: R.P., D.C., L.M., D.R. Writing—original draft: V.C., D.R. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-025-01088-0>.

**Correspondence** and requests for materials should be addressed to Daniele Ramazzotti.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2025