Article

Title: NRF2-driven transcription and KEAPness predict survival outcomes in lung cancer treated with immunotherapy.

Authors

Stefano Scalera [1, †], Biagio Ricciuti [2, †], Daniele Marinelli [3, 4, 5, †], Marco Mazzotta [6], Laura Cipriani [1], Giulia Bon [7], Beatrice Messina [1], Giulia Schiavoni [1], Alessandro Di Federico [2], Joao V. Alessi [2], Maurizio Fanciulli [8], Ludovica Ciuffreda [8], Francesca De Nicola [8], Frauke Goeman [8], Giulio Caravagna [9], Daniele Santini [10], Ruggero De Maria [11], Federico Cappuzzo [12], Gennaro Ciliberto [13], Mariam Jamal-Hanjani [4], Mark M. Awad [2], Nicholas McGranahan [5, *], and Marcello Maugeri-Saccà [1, 12, *]

[†] These authors contributed equally to this work.

* These authors share senior authorship.

Affiliations

1. Clinical Trial Center, Biostatistics and Bioinformatics Division, IRCCS Regina Elena National Cancer Institute, Rome, Italy.

2. Lowe Center for Thoracic Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA.

3. Department of Experimental Medicine, Sapienza University, Rome, Italy.

4. Cancer Metastasis Laboratory, University College London Cancer Institute, London, UK.

5. Cancer Genome Evolution Research Group, University College London Cancer Institute, London, UK

6. Medical Oncology Unit, Sandro Pertini Hospital, Rome, Italy.

7. Cellular Network and Molecular Therapeutic Target Unit, IRCCS Regina Elena National Cancer Institute, Rome, Italy.

8. SAFU Laboratory, Department of Research, Advanced Diagnostic, and Technological Innovation, IRCCS Regina Elena National Cancer Institute, Rome, Italy.

9. Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy.

10. Department of Medical-Surgical Sciences and Biotechnologies, Università degli Studi di Roma La Sapienza, Rome, Italy.

11. Dipartimento di Medicina e Chirurgia Traslazionale, Università Cattolica del Sacro Cuore, Rome, Italy; Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy.

12. Division of Medical Oncology 2, IRCCS Regina Elena National Cancer Institute, Rome, Italy.

13. Scientific Direction, IRCCS Regina Elena National Cancer Institute, Rome, Italy.

Corresponding author: Dr. Marcello Maugeri-Saccà, Clinical Trial Center, Biostatistics and Bioinformatics Division, IRCCS Regina Elena National Cancer Institute, Via Elio Chianesi 53, 00144, Roma, Italy. email: marcello.maugerisacca@ifo.it, phone +39-0652662724, fax +39-0652665523.

Abstract

KEAP1 mutations have emerged as determinants of survival outcomes in non-small cell lung cancer (NSCLC) patients treated with immune checkpoint inhibitors (ICIs). Given that mutational inactivation of *KEAP1* leads to NRF2 transcription, we hypothesized that gene expression can determine immune responsiveness with higher precision. Using the TCGA, we identified a NRF2 gene signature shared by NSCLC with and without KEAP1-NRF2 pathway mutations. NRF2-enriched tumors were associated with poor survival outcomes and immune exclusion in two independent cohorts of ICI-treated NSCLC patients (SU2C n=153; OAK/POPLAR n=439). Patients with KEAPness (NRF2-enriched and *KEAP1* wild-type) tumors had survival outcomes comparable to their *KEAP1*-mutant counterparts. In the TRACERx 421, NRF2-enriched tumors exhibited limited transcriptional intratumoral heterogeneity and an immune-excluded microenvironment. Moreover, the comparison between *KEAP1*-mutant and KEAPness NSCLC revealed shared and private cancer genes under positive selection. This study demonstrates that NRF2 transcriptional activity and *KEAP1*-loss phenocopies predict immunotherapy outcomes and outperforms mutational models.

Introduction

Immune checkpoint inhibitors targeting the programmed cell death protein 1/programmed death-ligand 1 (PD-1/PD-L1) pathway have reshaped the therapeutic landscape of advanced non-small-cell lung cancer (NSCLC) [1]. Nevertheless, the limited accuracy of companion biomarkers in predicting clinical outcomes is a critical hurdle towards precision immuno-oncology. While immunohistochemical assessment of PD-L1 is routinely performed in the clinical setting, and generally correlates with increased benefit from PD-(L)1-based therapies, responses are also seen in tumor lacking PD-L1 expression [1]. Similarly, tumor mutational burden (TMB) has been intensively investigated as predictive biomarker. However, its use has not yet been implemented in clinical practice [2].

The Kelch-like ECH-associated protein 1 (KEAP1)-Nuclear factor erythroid-2-related factor 2 (NRF2) pathway is a core defensive mechanism against a variety of harmful cues, regulating cellular redox homeostasis and protecting cells against xenobiotics [3-5]. In unstressed cells, KEAP1 triggers NRF2 proteasomal degradation through the CUL3-RBX1 E3 ubiquitin ligase complex. Oxidative and electrophilic stressors modify KEAP1 sensor cysteines, hampering its capability to bind the transcription factor NRF2. NRF2-driven transcription promotes metabolic rewiring and ensures protection against reactive oxygen species, xenobiotics, and ferroptosis [3-5].

KEAP1 loss-of-function (LOF) mutations occur in approximately 15% of NSCLC [6,7], leading to uncontrolled NRF2 transcriptional activity and enhanced cytoprotection. In the specific context of non-squamous NSCLC, mostly consisting of lung adenocarcinoma (LUAD), mutations in the *KEAP1* tumor-suppressor gene are associated with resistance to chemotherapy, radiotherapy, and targeted agents, including KRAS-G12C inhibitors [8-11]. To a similar extent, *KEAP1* mutations have been associated with inferior survival outcomes among non-squamous NSCLC patients treated with ICIs [12-15]. Studies exploring the connection between *KEAP1* mutations and immunotherapy efficacy

capitalized on the concept of anti-immune epistatic interactions, a genetic framework envisioning that cooperating genomic events elicit non-linear effects on cancer cell fitness and foster immune evasion. Consistently, co-existing *KEAP1* and *KRAS*/*STK11* mutations defined an uncommon subset of patients experiencing rapid disease progression during immunotherapy [12,13]. Conversely, *KEAP1*/*TP53* co-mutations resulted in intermediate outcomes, denoting a neutral interaction [14]. We have also described the deleterious nature of *KEAP1* mutations when associated with loss of heterozygosity (LOH) [15]. However, the same considerations do not extend to *NFE2L2* (the gene encoding for NRF2). Indeed, in squamous cell lung carcinoma (LUSC), where *NFE2L2* is more commonly mutated than *KEAP1* [7], evidence on the association between pathway mutations and inferior survival outcomes is still lacking. As a result, clinical outcomes of NSCLC patients receiving ICIs remain largely unpredictable beyond the limited fraction of patients whose tumors harbor specific mutational configurations, or that exhibit very high PD-L1/TMB levels.

It has recently been observed that the magnitude of NRF2 transcription predicts both *KEAP1*/*NFE2L2* mutations and mutation pathogenicity (deleterious versus tolerated) [16]. We hypothesized that NRF2 transcriptional activity may accurately predict immunotherapy efficacy by intercepting both negative/positive outliers (rapid disease progression/durable clinical benefit) and the large proportion of patients who do not obtain significant benefit from ICIs despite their tumors lack putative biomarkers of efficacy/resistance. We envisioned that hallmarks of lethal *KEAP1* mutations are seeded in tumors with intact KEAP1-NRF2 pathway, and that this KEAPness phenotype shares the anti-immune properties of the canonical *KEAP1*-mutant counterpart. On this basis, we conceived the following workflow: i) The TCGA NSCLC cohort was used to identify NRF2 target genes that are expressed in both *KEAP1*/*NFE2L2*-mutant and wild-type tumors [6,7,17,18], and that were used for generating a NRF2 transcriptional signature; ii) The impact of the NRF2 signature on survival outcomes and immune phenotype was tested in two independent cohorts of metastatic NSCLC profiled

by RNA-Seq and treated with ICIs: The Stand Up To Cancer-Mark Foundation (SU2C) NSCLC cohort (discovery cohort, n=153) and the OAK/POPLAR cohort (validation cohort, n=439) [19-22]. This latter consists of patients enrolled in the phase II/III trials investigating second-line atezolizumab versus docetaxel in advanced NSCLC [20,21]; iii) The pooled SU2C/OAK/POPLAR cohort with available *KEAP1* mutation data was analyzed to evaluate whether KEAPness tumors (NRF2-enriched but *KEAP1* wild-type) are poorly responsive to ICIs, similarly to those carrying *KEAP1* mutations; iv) The prospective TRACERx421 multi-region sequencing study, which contains matched RNA-seq and whole-exome sequencing (WES) data from 347 non-metastatic NSCLC patients (947 tumor regions) [23], was used to infer the oncogenic drivers of NRF2-enriched NSCLC, NRF2 transcriptional heterogeneity, and immunological correlates of NRF2-driven transcription.

Results

Identification of the NRF2 signature in the TCGA NSCLC cohort

The workflow of the present study is illustrated in Figure 1. To provide conceptual ground to our hypothesis, we first performed a differential gene expression analysis between *KEAP1-NFE2L2*-mutant and wild-type cases in the TCGA NSCLC cohort (n=943). This approach retrieved a set of 220 differentially expressed genes between mutant and wild-type tumors (Figure 2a, Supplementary Table 1). Unsupervised hierarchical clustering revealed that while most of the differentially expressed genes are hallmark of *KEAP1-NFE2L2*-mutant tumors, a pool of positively correlated NRF2 target genes are also expressed in a non-negligible fraction of NSCLC lacking pathway mutations (Figure 2b-c). These NRF2-associated genes, expressed in both *KEAP1-NFE2L2*-mutant and wild-type tumors, were therefore exploited to generate the NRF2 signature. To better characterize the transcriptional repertoire associated with NRF2 activity, we investigated its relationship with both histology and cancer-associated pathways inferred from perturbation experiments (PROGENy) [24]. We noticed that LUSC

had significantly higher NRF2 signature levels than LUAD (Extended Data Figure 1). Moreover, tumors with higher NRF2 signature scores exhibited a distinct profile of cancer-related pathway activity, e.g., p53, WNT, and JAK-STAT, compared to NRF2-depleted samples (Figure 2d; and Supplementary Figure 1). Differences in the activity of the immune-related transcription factors STAT1 and STAT3 were confirmed with DoRothEA, a method that estimates transcription factor-target gene interactions (Figure 2e) [25]. Next, we exploited an established method for immune subtyping from bulk RNA-Seq to explore differences in the immune microenvironment (http://science.bostongene.com/tumor-portrait) [26]. We noted that tumors with high NRF2 signature scores prevalently had an immune-excluded microenvironment (Figure 2f; and Extended Data Figure 2). Overall, hypothesis-generating analyses in the TCGA indicated that a specific set of NRF2 target genes are also expressed in NSCLC with wild-type KEAP1-NRF2 pathway, and are associated with a transcriptomic background potentially denoting increased resistance to ICIs (e.g., immune-related features). On this basis, the NRF2 signature identified in the TCGA was used for survival and correlative molecular analyses in the two immunotherapy-treated cohorts (SU2C and OAK/POPLAR).

The NRF2 gene signature predicts survival outcomes in the SU2C identification cohort

Having identified NRF2 target genes with broad expression in NSCLC, we next tested the impact of the NRF2 signature on the efficacy of ICIs in the SU2C cohort. Using the receiver operator characteristic (ROC) curve and bootstrap resampling for predicting a lack of clinical benefit (disease progression within six months from the beginning of ICIs), we defined the most accurate signature cut point (cut-off 2.1678, area under the curve 0.67, Extended Data Figure 3). To avoid confounding factors related to histology and previous therapies, this analysis was carried out in non-squamous NSCLC patients who did not receive prior tyrosine kinase inhibitors (TKIs) (n=93). This threshold was

used for all subsequent survival analyses and immune subtyping in the two independent clinical cohorts (SU2C and OAK/POPLAR), as well as the TRACERx 421 cohort.

Baseline clinical characteristics of patients included in the SU2C cohort (n=153), and their association with the NRF2 signature, are reported in Supplementary Table 2 and 3. Patients with NRF2-enriched tumors had shorter progression-free survival (PFS) and overall survival (OS) compared to those with NRF2-depleted NSCLC (PFS log-rank p=0.042; OS log-rank p=0.008, Figure 3a-b). Multivariate Cox regression models for PFS and OS confirmed the robustness of the NRF2 signature, indicating that the performances of the model were independent of established prognostic factors (PFS: p=0.018, HR 1.76, 95% CI 1.10-2.8; OS: p=0.019, HR 1.96, 95% CI 1.11-3.4; figure 3c-d). We observed a similar association between the NRF2 signature and overall response rate (ORR), being progressive disease (PD)/stable disease (SD) more commonly reported among patients with NFR2-enriched tumors (chi-squared test p=0.037; Figure 3e). Moreover, LUSC had significantly higher NRF2 signature levels than non-squamous NSCLC, as observed in the TCGA, a finding that is consistent with the lower efficacy of ICIs in LUSC (Extended Data Figure 4). Considering the limited role of ICIs in oncogene-addicted NSCLC, we carried out a sensitivity analysis upon exclusion of patients who received prior TKIs. In this case, the performances of the NRF2 classifier were further improved (Extended Data Figure 5). Of note, all the 12 patients treated with prior TKIs had tumors classified as NRF2-depleted, consistently with the mutual exclusivity between *KEAP1* and *EGFR* mutations [12]. PD-L1 expression levels were available for a minority of patients (n=71). NRF2-enriched tumors had lower PD-L1 expression levels, even though this difference was not significant due to the limited size of the subgroups compared (Supplementary Figure 2). Lastly, we confirmed that NRF2-enriched tumors were prevalently characterized by an immune-excluded phenotype, as already noted in the TCGA (chi-squared test p=0.001, Figure 3f-g).

The NRF2 gene signature predicts survival outcomes in the OAK/POPLAR validation cohort

The independent OAK/POPLAR cohort (atezolizumab, n=439) was used to confirm the findings from the SU2C study. To ensure reproducibility, the same NRF2 signature cut-off identified in the SU2C study was adopted. The available clinical features and their association with the NRF2 signature are reported as Supplementary Table 4. In the OAK/POPLAR cohort, patients with NRF2-enriched tumors had inferior PFS and OS compared to those with NRF2-depleted NSCLC (PFS log-rank p=0.001; OS log-rank p<0.001; Figure 4a-b). The distribution of NRF2-enriched and depleted cases was comparable to what is reported in the SU2C cohort. Also in the OAK/POPLAR study, squamous tumors exhibited significant higher NRF2 signature levels than non-squamous NSCLC (Extended Data Figure 6). Despite the limited availability of clinical features, multivariate Cox regression models for PFS and OS were consistent with Kaplan-Meier survival curves (PFS: p<0.001, HR 1.47, 95% CI 1.18-1.8; OS: p=0.002, HR 1.5, 95% CI 1.16-1.9; Figure 4c-d). We also confirmed that NRF2-enriched tumors are associated with lower ORR and are characterized by an immune-excluded phenotype (Figure 4e-g).

Next, we explored the association between the NRF2 signature and survival outcomes in a control cohort of chemotherapy-treated patients in the OAK/POPLAR trials (docetaxel, n=452). Also in this setting, NSCLC patients with NRF2-enriched tumors had shorter PFS and OS, as well as lower ORR (log-rank p=0.019 and p<0.001 for PFS and OS, respectively; chi-squared test for ORR p=0.005; Supplementary Figure 3).

Collectively, these data indicate that NRF2 transcriptional activity efficiently predicts immunotherapy efficacy and immune exclusion in advanced NSCLC patients, and is associated with worse outcomes to chemotherapy suggesting a potential prognostic role. This classifier divides the population in two groups of comparable size, which is consistent with the expected benefit from ICIs, thus indicating that the model does not exclusively identify outlier patients.

Pooled analysis of the SU2C and OAK/POPLAR cohorts

Having validated the model in the two independent cohorts, we used the merged SU2C/OAK/POPLAR dataset to further assess the differences between NRF2-enriched and NRF2-depleted NSCLC and investigate the impact of *KEAP1* mutations.

As expected, differences in survival outcomes between the NRF2-enriched and depleted NSCLC were more pronounced than in the individual cohorts (Supplementary Figure 4). Immune-subtyping, cancer-associated signatures (PROGENy), and STAT1/STAT3 activity (DoRothEA) were consistent with results from the TCGA (Supplementary Figure 5). Having noticed that LUSC had significantly higher levels of the NRF2 signature across all the cohorts considered, we carried out survival analysis in the subset of non-squamous NSCLC. In this subgroup analysis, we confirmed the impact of the NRF2 signature on survival outcomes among patients with non-squamous lung cancer (Supplementary Figure 6). Moreover, the clinical significance of the NRF2 transcriptional signature was independent of histology, even though survival analyses in squamous tumors were limited by the relatively low sample size (Supplementary Figure 6).

Data on *KEAP1* mutations were available from 65 patients in the SU2C cohort and 251 patients in the OAK/POPLAR studies. While the majority of *KEAP1* mutations (n=61) were seen in NRF2-enriched NSCLCs, a limited number of *KEAP1* mutations occurred in the NRF2-depleted context (Supplementary Figure 7). In survival analyses, we confirmed the hypothesis that patients with NRF2-enriched/*KEAP1* wild-type (KEAPness) NSCLC had shorter PFS and OS compared to those with NRF2-depleted tumors (Figure 5a-b). Importantly, survival outcomes of the KEAPness group were comparable to those of the NRF2-enriched/*KEAP1*-mutant group, indicating the mutation-independent significance of the NRF2 signature (Figure 5a-b). Similarly, KEAPness tumors had lower ORR

compared to NRF2-depleted tumors (Figure 5c). In terms of immune-related features, NSCLC with KEAPness more frequently exhibited an immune-excluded microenvironment compared to NRF2-depleted tumors, having an immunological phenotype reminiscent of that of *KEAP1*-mutant tumors (Figure 5d-e). Again, KEAPness tumors were characterized by higher p53 and WNT signature enrichment scores, and lower JAK-STAT activity, compared to NRF2-depleted NSCLC (Extended Data Figure 7).

Overall, data from the larger pooled cohort indicated that NRF2-enriched tumors are poorly responsive to ICIs, regardless of the presence of concurrent *KEAP1* mutations. In addition, this transcriptional model outperforms a mutation-only approach by identifying a large population of NSCLC patients who do not obtain significant clinical benefit from ICIs.

NRF2 transcriptional heterogeneity, genetic selection, and immune infiltration in the TRACERx 421 cohort

We lastly investigated the transcriptional heterogeneity and genetic correlates of our model. To this end, we exploited multi-region WES and transcriptome sequencing data from the TRACERx 421 cohort (previously untreated, early-stage NSCLC) [23, 27].

First, the majority of NRF2-enriched tumors were uniformly enriched, given that only 24% of NRF2-enriched tumors had at least one sample classified as NRF2-depleted (33/139, Figure 6a), denoting limited intra-patient heterogeneity, namely, a phenotype which is consistent across time and space in tumor evolution. The frequency of NRF2 heterogeneity (i.e. NRF2-enriched tumors with at least one sample classified as NRF2-depleted) was independent of tumor histology (Extended Data Figure 8). *KEAP1* mutations and copy number alterations (*KEAP1* CNAs in *KEAP1* wild-type tumors) were frequently, but not exclusively, seen among NRF2-enriched NSCLC (Supplementary Figure 8)

To further explore the genetic background of NRF2-enriched tumors, we investigated genetic selection among more than 700 cancer genes using the dndscv method [28]. We identified a set of driver genes under positive selection (q-value < 0.05), with differences in relation to histology (Figure 6b-c). In non-squamous NSCLC, NRF2-enriched tumors with and without *KEAP1* mutations shared positive selection for *TP53*, *KRAS* and *STK11* (Figure 6b). However, while *SMARCA4* was under positive selection exclusively in the *KEAP1*-mutant background, *CDKN2A* (p14(ARF) was identified as a private driver in non-squamous NSCLC with KEAPness (Figure 6b). In squamous tumors, *TP53* and *CDKN2A* were identified as shared drivers in the NRF2-enriched context (Figure 6c). Conversely, squamous tumors with KEAPness exhibited positive selection for PI3K pathway genes (*PIK3CA* and *PTEN*), the tumor suppressor gene *RB1*, and *NFE2L2* (Figure 6c). Most non-silent mutations in these cancer-associated genes were of clonal nature (Supplementary Figure 9).

Next, we explored the immunological features associated with our model. As already reported in clinical cohorts, NRF2-enriched NSCLC were more commonly immune-excluded compared to NRF2-depleted tumors, regardless of co-existing *KEAP1* mutations (Figure 6d, and Extended Data Figure 9). Among NRF2 heterogeneous tumors, namely, NRF2-enriched tumors with at least one NRF2-depleted region, we observed a mixed immune-related pattern (Figure 6e). While some tumors were immunologically stable having the same immune subtype in all the tumor regions, in other tumors NRF2 enrichment was associated with changes in the tumor microenvironment (i.e., from immune-enriched to immune-excluded), suggesting tumor-microenvironment co-evolution (Figure 6e). Lastly, we leveraged T cell ExTRECT, a method that infers T cell infiltration from DNA sequencing, rather than RNA-Seq, for orthogonal validation of immune-related features [29]. NRF2-enriched tumors had lower, and *KEAP1*-independent, T-cell infiltration compared to the NRF2-depleted counterpart (Figure 6f).

Collectively, the analyses carried out in the TRACERx 421 highlighted limited intratumoral heterogeneity of the NRF2 signature, clarified the nature of cancer genes under positive selection in the NRF2-enriched background, and strengthened the relationship between NRF2 transcription and the immune microenvironment.

Discussion

Accumulating evidence links *KEAP1*-based mutational contexts to survival outcomes in NSCLC patients treated with immunotherapy [12-15]. However, dissecting the impact of *KEAP1* mutations on immunotherapy outcomes is challenging, given that co-mutations in *KEAP1*/*KRAS*/*STK11* occur in approximately 10-15 % of non-squamous NSCLC patients [12, 13]. In addition, these models cannot be extended to squamous tumors, which have a distinct genomic profile characterized by *TP53* alterations and lack of *KRAS*/*STK11* mutations [7]. Biomarkers associated with DCB represent the other side of the challenging task of predicting ICIs efficacy, considering the limited accuracy of PD-L1 and TMB. High PD-L1 levels (≥ 50%) are detected in approximately 30% of NSCLC [30], and are associated with greater benefit from immunotherapy. However, primary and acquired resistance is also commonly seen among patients with PD-L1-expressing NSCLC. Similarly, tumor responses are often observed in PD-L1 low/negative NSCLC. The use of TMB has also been difficult to implement in clinical practice. Beyond the discrepancies among the various pipelines [2], we have witnessed a wealth of studies which progressively raised the TMB cut-off, from ≥ 10 non-synonymous mutations per megabase (mut/Mb) to more extreme values (e.g., ≥ 19 mut/MB) [2, 31-35]. Intuitively, while raising the TMB is expected to improve its predictive ability, very high TMB thresholds significantly lower the TMB-high population, which drops down to 10-15% of the whole population [35]. Thus, there is a pressing need for improved biomarkers which can predict ICIs efficacy.

KEAP1 LOF mutations lead to aberrant NRF2 transcriptional activity. Here, we hypothesized that analyzing the KEAP1/NRF2 pathway from a transcriptional perspective may help overcome the many limitations of current mutation-based biomarkers. By integrating survival analyses and computational studies, we reported that: i) A set of NRF2 target genes are also expressed in *KEAP1-NFE2L2* wild-type NSCLCs; ii) A transcriptional NRF2 gene signature conferred inferior survival outcomes, was associated with lower response to immunotherapy, and was characterized by a cold immune microenvironment in two independent clinical cohorts. Importantly, our model divided the population into two equally sized subgroups; iii) The shorter survival outcomes charactering NRF2-enriched tumors were independent of underlying *KEAP1* mutations, indicating the clinical relevance of the KEAPness phenotype; iv) The analysis of the TRACERx 421multi-region sequencing cohort revealed limited, but not negligible, NRF2 transcriptional heterogeneity. Furthermore, while NRF2-enriched tumors with and without *KEAP1* mutations shared a set of driver mutations under positive selection, KEAPness tumors also had some private genomic traits driving tumor evolution.

Our study has important strengths, holding the potential to represent a turning point in the search of biomarkers for immune-oncology. First, the pool of genes included in the NRF2 signature was identified in the TCGA NSCLC cohort, and then applied to the clinical cohorts. Second, we used the same NRF2 signature threshold for the survival analyses, which yielded comparable results in terms of survival outcomes, distribution of cases, and immunological correlates in independent cohorts. Likewise, the same cut-off was also applied to the TRACERx 421, where immunological features of NRF2-enriched tumors were comparable to those observed in the other cohorts, and further confirmed by a different method leveraging DNA sequencing (T cell ExTRECT) [29]. Third, we used a streamlined method to classify samples as NRF2-enriched or NRF2-depleted, which could be easily exploited in the clinical setting. Importantly, we pursued an identification-validation approach, evaluating data from patients enrolled into a randomized phase III trial (OAK). In the search for cancer biomarkers, these

represent important methodological added values. The distribution of NRF2-enriched and -depleted cases deserves further mention. While current biomarkers mostly focus on outlier patients, the split achieved by the NFR2 signature suggests that this model better delineates the actual population of immunotherapy responsive/resistant NSCLC.

From a clinical perspective, this study has several implications. We currently have multiple first-line treatment options for patients with metastatic non-small cell lung cancer, including PD-(L)1 monotherapy alone or in combination with CTLA-4 inhibitors and/or chemotherapy, but we do not have robust biomarkers to individualize patient selection. These data suggest that patients with NRF2-enriched tumors have worse outcomes to PD-(L)1-based monotherapies, and that intensifying treatment with the addition of CTLA-4 blockade or chemotherapy may be a reasonable strategy, since the addition of CTLA-4 inhibitors and chemotherapy can be beneficial in immune-excluded tumors with low/negative PD-L1 expression [36-38]. Importantly, an increasing number of commercial assays are now exploiting WES and RNA-seq, highlighting how gene expression signature can be incorporated in clinical decision making. Our data also have implication for patients with early-stage NSCLC. PD-(L)1 blockade is now approved as adjuvant and neoadjuvant therapy [39-41], but we currently lack biomarkers that can help predicting who will need the addition of peri-operative immunotherapies. PD-L1 expression and TMB have been inconsistently associated with the risk of recurrence in these settings. Therefore, there are opportunities for the use of other biomarkers, such as mutational profiles and gene expression signatures, to identify early-stage NSCLC patients who will benefit from the addition of immunotherapy.

Although our data were consistent between two independent clinical cohorts, we acknowledge that our study has some limitations. The main limitation is a lack of data on PD-L1 and TMB, apart from a subset of patients in the SU2C cohort. Thus, we were unable to include these features in multivariate

Cox regression models. Likewise, a limited number of clinical parameters were available in the OAK/POPLAR cohort. Nevertheless, the reproducibility of the results at multiple levels (PFS, OS, ORR, immune subtyping) with the same NRF2 signature cut-off, in two independent cohorts containing data from nearly 600 immunotherapy-treated NSCLC patients, indicate the robustness of the model. A second limitation is the relatively small sample size of patients with *KEAP1* mutation data. Given that *KEAP1* status was available for approximately half of the whole population, our study was not powered to explore the impact of *KEAP1* in individual cohorts. Indeed, *KEAP1* mutation status was available for only 61 patients in the SU2C study. However, the reproducibility of the data across all the clinical and molecular endpoints (e.g., immune subtyping) in the pooled SU2C/OAK/POPLAR cohort indicated that the NRF2 gene signature is independent of co-existing *KEAP1* mutations. Lastly, we were unable to investigate the impact of other driver mutations, particularly *NFE2L2*. The low mutational frequency of *NFE2L2* in non-squamous NSCLC, which represented most of our population, and the lack of evidence connecting *NFE2L2* mutations to reduced immunotherapy efficacy, mitigate this issue. To a similar extent, lack of data on *KRAS* mutations in the OAK/POPLAR cohort hindered subgroup analyses in non-squamous tumors, considering the relationship between *KEAP1* and *KRAS* mutations [13].

In conclusion, results from this study indicate that a transcriptional signature denoting high NRF2 activity efficiently predicts survival outcomes and an immune-excluded phenotype in NSCLC patients treated with ICIs. The adverse survival outcomes conferred by NRF2 transcriptional activity extended beyond the presence of *KEAP1* mutations, indicating that KEAPness tumors behave in a *KEAP1*-mutant-like manner. Lastly, our data provide solid ground to the concept that transcriptomics, and integrated transcriptomics-mutational data, can significantly foster the identification of biomarkers predicting immunotherapy efficacy, and that this approach can be transferred to other driver mutations associated with deregulation of transcription factors.

Methods

Cohorts and patients

The TCGA NSCLC cohort, containing data from 943 tumor samples profiled by WES and RNA-Seq, was used for the identification of NRF2 target genes. Mutational and RNA-Seq data (RSEM TPM) were downloaded from cBioPortal (https://www.cbioportal.org/) [17,18].

The SU2C identification cohort contains data from NSCLC patients treated with immunotherapy. For this study, we selected those with available RNA-Seq data (n=153). Among them, 61 had matched WES. Survival outcomes, namely, PFS, OS, and ORR, were available for 136, 143, and 140 patients, respectively. Detailed information on sequencing methods are available in the original publication [19]. Baseline clinical features are reported as Supplementary Table 2, and included patient age at diagnosis, sex, histology, stage at diagnosis, treatment type, line of therapy, prior use of TKIs, and smoking status. PD-L1 status, assessed by the tumor proportion score (TPS), was available for 71 patients.

The OAK/POPLAR validation cohort consists of 891 patients enrolled into the phase II POPLAR trial and the phase III OAK trial, both comparing second-line atezolizumab (1200 mg IV every 3 weeks until disease progression) versus docetaxel (75 mg/m2 IV every 3 weeks) monotherapy. The study protocols and sequencing methods are detailed in the original publications [20,21]. In the available datasets, 439 patients were treated with atezolizumab, whereas 452 patients received mono-chemotherapy. Regarding clinical data, histology (squamous versus non-squamous) and sex were the only clinical features reported in the datasets. Data on *KEAP1* mutations were available from 517 patients, of whom 251 treated with atezolizumab and 266 treated with docetaxel [22]. Given that histology was reported as squamous and non-squamous in the OAK/POPLAR cohort, this definition was also applied to the SU2C and TRACERx 421 cohorts.

Genomic correlates of NRF2 transcription were investigated in the TRACERx 421 cohort, containing matched RNA-seq and WES data from 347 non-metastatic NSCLC patients (947 tumor regions). Methods of data collection and sequencing are extensively described in the related publications [23, 27].

Statistical analyses

Time to event endpoints (PFS and OS) were estimated using the Kaplan-Meier product-limit method, using the log-rank test for subgroups comparison. Multivariate Cox regression models for PFS and OS were performed by taking into account all the relevant patient- and treatment-related features reported in the original datasets. The related estimates were reported as hazard ratio (HR) and 95% confidence interval (CI). Differences in median values were estimated with the Wilcoxon test when two groups were compared. For the pooled SU2C/OAK/POPLAR cohort, where we investigated differences among three molecular subgroups (NRF2-enriched/*KEAP1*-mutant, KEAPness, and NRF2-depleted), we used the Kruskal-Wallis test followed by the Dunn's test for pairwise comparisons, and the Benjamini-Hochberg method to control the False Discovery Rate (FDR). When we evaluated the relationship between categorical variables (e.g., differences in ORR and immune subtyping between NRF2-enriched and NRF2-depleted NSCLC), the Pearson Chi-square test of independence was used. The Pearson's correlation test was used to assess the connection among the twelve genes included in the NRF2 signature. SPSS v21 and R ("survival" and "survminer" packages) were used for statistical analyses. The level of significance was defined as $p<0.05$.

Calculation of the NRF2 signature score

The NRF2 signature score was defined by the mean expression value of the 12 NRF2-linked genes identified upon differential gene expression and unsupervised hierarchical clustering in the TCGA NSCLC study (*AKR1C1*, *AKR1C2*, *AKR1C3*, *AKR1B10*, *AKR1B15*, *ALDH3A1*, *CYP4F3*, *CYP4F11*, *GPX2*, *PPP2R2C*, *UGT1A9*, *UGT1A6*). In the TCGA, we used the median value to classify samples as

NRF2-enriched and NRF2-depleted. In order to identify a NRF2 signature cut-off for survival analyses, log2(TPM + 1) RNA-seq data were used for both clinical cohorts (SU2C and OAK/POPLAR) and the TRACERx 421. The optimal cut-point of the NRF2 signature in predicting no durable clinical benefit (PFS<6 months) was identified with the "cutpointr" R package, using the receiver operator characteristic (ROC) curve and optimizing the Youden index via bootstrap resampling (n=2,000). For this purpose, we used a subset of patients included in the SU2C cohort, namely, non-squamous NSCLC patients who did not receive prior TKIs (n=93). The same cutoff (NRF2 signature score ≥ 2.1678 versus <2.1678) was adopted for all the analyses in the SU2C, OAK/POPLAR, pooled SU2C/OAK/POPLAR, and TRACERx 421.

Bioinformatic analyses

For differential gene expression, carried out in the TCGA NSCLC study with the goal of identifying differentially expressed gene between *KEAP1-NFE2L2*-mutant and wild-type samples, we used DESeq2 with the following parameters: Log2FC ≥ 1.2 and Log2FC ≤ -1.2; padj ≤ 0.05. Unsupervised hierarchical clustering was performed with "ComplexHeatmap", using the following settings: clustering_method_columns = "euclidean" (ward.D), column_km (k-means) = 4, row_split = 15. The study of cancer-associated pathway signatures was performed with PROGENy (Pathway RespOnsive GENes; https://saezlab.github.io/progeny/), a method that infers the activity of cancer pathways from perturbation experiments [24], whereas the immune-related transcription factors STAT1 and STAT3 were investigated with DoRothEA (https://saezlab.github.io/dorothea/) [25]. For immune subtyping, we exploited a method that generates four subtypes (immune-depleted and non-fibrotic, immune-depleted and fibrotic, immune-enriched and non-fibrotic, immune-enriched and fibrotic) on the basis of 29 microenvironment-related signatures estimated from RNA-Seq (http://science.bostongene.com/tumor-portrait; http://science.bostongene.com/tumor-portrait). [26] In the TRACERx 421, the dN/dS method

was used to detect positive selection in established cancer genes [28]. The dndscv function from the dNdScv R package was run on genes from the COSMIC Cancer Gene Census (v. 98, https://cancer.sanger.ac.uk/census) using the gene_list argument; the run was performed separately on patients grouped by NRF2 signature classes, histology, and presence of a non-silent *KEAP1* mutation. Only genes with global q-values < .05 and evidence of positive selection (i.e. total number of non-synonymous mutations ≥ total number of synonymous mutations) were classified as significant. T cell infiltration was estimated from WES data using the T cell ExTRECT method [29].

Data availability

WES and RNA-Seq data related to the TCGA NSCLC study are freely available at https://www.cbioportal.org/. Details regarding data availability of the SU2C cohort are provided in the original publication [19]. For this study, an additional patient treated at the Dana-Farber Cancer Institute (Boston, MA, USA) was included. Data related to the OAK/POPLAR trials are deposited to the European Genome-phenome Archive (EGA) (Study ID EGAS00001005013. Datasets: EGAD00001007703, EGAD00001008550, EGAD00001008390, EGAD00001008391, EGAD00001008548, EGAD00001008549). Data access was granted by Genentech to Dr. M. Maugeri-Saccà for the specific purpose of this study (Project Title: Mapping NRF2 activity in non-small cell lung cancer treated with immune checkpoint inhibitors). TRACERx 421 data are available from the original publication [23].

Acknowledgements

M. Maugeri-Saccà is supported by the Italian Association for Cancer Research (AIRC) under MFAG 2019 - project ID. 22940, and the Italian Ministry of Health (MoH)- project ID. GR-2016-02362025. N.M. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (Grant Number 211179/Z/18/Z), and also receives funding from Cancer Research UK Lung Cancer Centre of

Excellence, Rosetrees, and the NIHR BRC at University College London Hospitals. B. Ricciuti is supported by The Conquer Cancer Foundation of the American Society for Clinical Oncology and the Society for Immunotherapy of Cancer. G. Caravagna is supported by the Italian Association for Cancer Research (AIRC) under MFAG 2020- ID. 24913. D. Marinelli is a fellow of the PhD Network Oncology and Precision Medicine, Department of Experimental Medicine, Sapienza University of Rome. The research leading to these results has received funding from the European Union - NextGenerationEU through the Italian Ministry of University and Research under PNRR - M4C2-I1.3 Project PE_00000019 "HEAL ITALIA" to M. Maugeri-Saccà and G. Ciliberto, CUP H83C22000550006. The views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Competing interests

R.D.M. reports serving as a scientific advisory board member at Exosomics SpA (Siena IT), HiberCell Inc. (New York, NY), Kiromic Inc. (Houston, TX) and Exiris Inc. (Rome, IT). **FC** reports personal fees from Roche/Genentech, AstraZeneca, Takeda, Pfizer, Bristol-Myers Squibb, Merck Sharp & Dohme, Lilly, and Bayer. **M.M.A.** reported serving as a consultant for Achilles, AbbVie, Neon, Maverick, Nektar, and Hegrui; receiving grants and personal fees from Genentech, Bristol-Myers Squibb, Merck, AstraZeneca, and Lilly; and receiving personal fees from Maverick, Blueprint Medicine, Syndax, Ariad, Nektar, Gritstone, ArcherDx, Mirati, NextCure, Novartis, EMD Serono, and NovaRx. **N.M.** has received consultancy fees and has stock options in Achilles Therapeutics. He holds European patents relating to targeting neoantigens (PCT/EP2016/ 059401), identifying clinical response to immune checkpoint blockade (PCT/ EP2016/071471), determining HLA loss of heterozygosity

(PCT/GB2018/052004) and predicting survival rates of patients with cancer (PCT/GB2020/050221). The remaining authors declare no conflict of interest.

References

1. Grant, M.J. et al. Selecting the optimal immunotherapy regimen in driver-negative metastatic NSCLC. Nat Rev Clin Oncol. 18, 625-644 (2021).

2. Jardim, D.L. et al. The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker. Cancer Cell. 39, 154-173 (2021).

3. Pillai, R. et al. NRF2: KEAPing tumors protected. Cancer Discov. 12, 625-643 (2022).

4. Hellyer, J.A. et al. Clinical implications of KEAP1-NFE2L2 mutations in NSCLC. J Thorac Oncol. 16, 395-403 (2021).

5. Scalera, S. et al. KEAP1-Mutant NSCLC: The catastrophic failure of a cell-protecting hub. J Thorac Oncol. 17, 751-757 (2022).

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

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

8. Goeman, F. et al Mutations in the KEAP1-NFE2L2 pathway define a molecular subset of rapidly progressing lung adenocarcinoma. J Thorac Oncol. 14, 1924-1934 (2019).

9. Binkley, M.S. et al. KEAP1/NFE2L2 Mutations Predict Lung Cancer Radiation Resistance That Can Be Targeted by Glutaminase Inhibition. Cancer Discov. 10, 1826-1841 (2020).

10. Foggetti, G. et al. Genetic Determinants of EGFR-Driven Lung Cancer Growth and Therapeutic Response In Vivo. Cancer Discov. 11, 1736-1753 (2021).

11. Negrao, M.V. Comutations and KRASG12C Inhibitor Efficacy in Advanced NSCLC. Cancer Discov. 13, 1556-1571 (2023).

12. Marinelli, D. et al. KEAP1-driven co-mutations in lung adenocarcinoma unresponsive to immunotherapy despite high tumor mutational burden. Ann Oncol. 31, 1746-1754(2020).

13. Ricciuti, B. et al. Diminished Efficacy of Programmed Death-(Ligand)1 Inhibition in STK11- and KEAP1-Mutant Lung Adenocarcinoma Is Affected by KRAS Mutation Status. J Thorac Oncol. 17, 399-410 (2022).

14. Scalera, S. et al. KEAP1 and TP53 Frame Genomic, Evolutionary, and Immunologic Subtypes of Lung Adenocarcinoma With Different Sensitivity to Immunotherapy. J Thorac Oncol. 16, 2065-2077 (2021).

15. Scalera, S. et al. Clonal KEAP1 mutations with loss of heterozygosity share reduced immunotherapy efficacy and low immune cell infiltration in lung adenocarcinoma. Ann Oncol. 34, 275-288 (2023).

16. Arolt, C. et al. KEAP1/NFE2L2 pathway signature outperforms KEAP1/NFE2L2 mutation status and reveals alternative pathway-activating mutations in Non-Small Cell Lung Cancer. J Thorac Oncol. S1556-0864(23)00684-6 (2023).

17. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 6:pl1 (2013)

18. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2, 401□404 (2012).

19. Ravi, A. et al. Genomic and transcriptomic analysis of checkpoint blockade response in advanced non-small cell lung cancer. Nat Genet. 55, 807-819 (2023).

20. Fehrenbacher, L. et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. Lancet. 387, 1837-1846 (2016).

21. Rittmeyer, A. et. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet. 389, 255-265 (2017).

22. Patil, N.S. et al. Intratumoral plasma cells predict outcomes to PD-L1 blockade in non-small cell lung cancer. Cancer Cell. 40, 289-300 (2022).

23. Martínez-Ruiz, C. Genomic-transcriptomic evolution in lung cancer and metastasis. Nature. 616, 543-552 2023.

24. Schubert, M. et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun. 9, 20 (2018).

25. Garcia-Alonso, L. et al. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 29,1363-1375 (2019).

26. Bagaev, a. et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell. 39, 845-865 (2021).

27. Frankell, A.M. et al. The evolution of lung cancer and impact of subclonal selection in TRACERx. Nature. 616, 525-533 (2023).

28. Martincorena, I. et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 171, 1029-1041 (2017).

29. Bentham, R. et al. Using DNA sequencing data to quantify T cell fraction and therapy response. Nature. 597, 555-560 (2021) Sep; (7877).

30. Reck, M. et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med. 375, 1823-1833 (2016).

31. Marabelle, A. et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. Lancet Oncol. 21, 1353-1365 (2020).

32. Hellmann, M.D. et al. Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. Cancer Cell. 33, 843-852 (2018).

33. Rizvi, H. et al. Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. J Clin Oncol. 36, 633-641 (2018).

34. Gandara, D.R. et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. Nat Med. 24, 1441-1448 (2018)

35. Ricciuti, B. et al. Association of High Tumor Mutation Burden in Non-Small Cell Lung Cancers With Increased Immune Infiltration and Improved Clinical Outcomes of PD-L1 Blockade Across PD-L1 Expression Levels. JAMA Oncol. 8, 1160-1168 (2022).

36. Hellmann, M.D. et al. Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. N Engl J Med. 22, 2093-2104 (2018).

37. Paz-Ares, L. et al. First-line nivolumab plus ipilimumab combined with two cycles of chemotherapy in patients with non-small-cell lung cancer (CheckMate 9LA): an international, randomized, open-label, phase 3 trial. Lancet Oncol. 2, 198-211 (2021).

38. Gandhi, L. et al. Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. N Engl J Med. 378, 2078-2092 (2018).

39. Forde, P.M. et al. Neoadjuvant Nivolumab plus Chemotherapy in Resectable Lung Cancer. N Engl J Med. 386, 1973-1985 (2022).

40. Felip, E. et al. Adjuvant atezolizumab after adjuvant chemotherapy in resected stage IB-IIIA non-small-cell lung cancer (IMpower010): a randomised, multicentre, open-label, phase 3 trial. Lancet. 398, 1344-1357 (2021).

41. Wakelee, H. et al. Perioperative Pembrolizumab for Early-Stage Non-Small-Cell Lung Cancer. N Engl J Med. 389, 491-503 (2023).

Figure Legends

Figure 1. Study workflow.

The TCGA NSCLC cohort was used to identify NRF2-associated genes shared by *KEAP1-NFE2L2*-mutant and wild-type cases. The Stand Up To Cancer-Mark Foundation (SU2C identification cohort, n=153) and OAK/POPLAR (validation cohort, n=439) cohorts were used to investigate the impact of the transcriptional NRF2 gene signature on survival outcomes and immune subtyping in advanced NSCLC patients treated with immune checkpoint inhibitors. The impact of *KEAP1* mutations on the performances of the NRF2 signature was assessed in the pooled cohort (SU2C/OAK/POPLAR). The TRACERx 421 multi-region sequencing study was analyzed to investigate the genetic drivers

associated with KEAPness, NRF2 transcriptional heterogeneity, and the immunological features associated with NRF2 transcription.

Figure 2. Identification of the NRF2 gene signature in the TCGA NSCLC study.

a, Volcano plot showing differentially expressed genes between *KEAP1-NFE2L2*-mutant and wild-type NSCLC (right: up-regulated, left: down-regulated; log2 fold change ≥1.2 and ≤ -1.2, adjusted p-value ≤ 0.05). **b**, Heatmap showing unsupervised hierarchical clustering of differentially expressed genes, with magnification of the gene set of interest. **c**, Heatmap illustrating Pearson correlation coefficients for the genes included in the NRF2 gene signature. **d**, Box plots illustrating the differences in the p53, WNT, and JAK-STAT PROGENy signatures between NRF2-enriched and depleted cases (Wilcoxon test). **e**, Box plots illustrating DoRothEA-predicted activity of the immune-related transcriptional factors STAT1 and STAT3 in NRF2-enriched and depleted NSCLC (Wilcoxon test). **f**, Stacked bar chart displaying the distribution of immune subtypes in NFR2-enriched and depleted NSCLC (χ^2). D: desert, immune-depleted and non-fibrotic, F: fibrotic, immune-depleted and fibrotic, IE: immune-enriched and non-fibrotic, IE/F: immune-enriched and fibrotic. The NRF2 signature enrichment score was obtained by the mean expression value of the 12 NRF2-associated genes. For TCGA data analyses, NRF2-enriched and NRF2-depleted cases were classified on the basis of the median value of the NRF2 signature.

Figure 3. Survival analyses and immune subtyping in the SU2C identification cohort.

a, b Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS) comparing NRF2-enriched and NRF2-depleted NSCLC. **c**, **d** Forest plots showing multivariate Cox regression models for PFS and OS. **e**, Stacked bar chart illustrating the differences in the overall response rate (ORR) between patients with NRF2-enriched and NRF2-depleted NSCLC (χ^2). CR: complete response, PR: partial response, SD: stable disease, PD: progressive disease. **f**, Stacked bar

chart showing the distribution of immune subtypes in NRF2-enriched and NRF2-depleted NSCLC (χ^2). D: desert, immune-depleted and non-fibrotic, F: fibrotic, immune-depleted and fibrotic, IE: immune-enriched and non-fibrotic, IE/F: immune-enriched and fibrotic. **g**, Heatmap illustrating the 29 tumor immune microenvironment (TIME) signatures used for immune subtyping (BostonGene) in relation to the NRF2 model.

Figure 4. Survival analyses and immune subtyping in the OAK/POPLAR validation cohort.

a, b, Kaplan-Meier survival curves of progression-free survival (PFS) and overall survival (OS) comparing NRF2-enriched and NRF2-depleted NSCLC. **c**, **d**, Forest plots illustrating multivariate Cox regression models for PFS and OS. **e**, Stacked bar chart showing tumor response to immunotherapy (overall response rate, ORR) in patients with NRF2-enriched and NRF2-depleted NSCLC (χ^2). CR: complete response, PR: partial response, SD: stable disease, PD: Progressive Disease. **f**, Stacked bar chart showing differences in immune subtypes between NRF2-enriched and NRF2-depleted NSCLC (χ^2). D: desert, immune-depleted and non-fibrotic, F: fibrotic, immune-depleted and fibrotic, IE: immune-enriched and non-fibrotic, IE/F: immune-enriched and fibrotic. **g**, Heatmap showing the 29 tumor immune microenvironment (TIME) signatures exploited for immune subtyping (BostonGene) in relation to the NRF2 signature.

Figure 5. Survival analyses in the pooled SU2C/OAK/POPLAR cohort to explore the impact of *KEAP1* mutations on the NRF2 signature.

a, b, Kaplan-Meier survival curves of progression-free survival (PFS) and overall survival (OS) comparing NRF2-enriched/*KEAP1*-mutant, NRF2-enriched/*KEAP1* wild-type (KEAPness), and NRF2-depleted NSCLC. **c**, Stacked bar chart displaying the differences in terms of overall response rate (ORR) across the subgroups compared (χ^2). **d**, Stacked bar chart for immune subtyping across the three molecular subgroups considered. **e**, Heatmap illustrating the 29 tumor immune microenvironment

(TIME) signatures used for immune subtyping (BostonGene) across the three subgroups compared (NRF2-enriched/*KEAP1*-mutant, NRF2-enriched/*KEAP1* wild-type, and NRF2-depleted).

Figure 6. Genomic and immunological correlates of NRF2-enriched tumors in the TRACERx 421 multi-region sequencing cohort.

a, Dumbbell plot showing the NRF2 signature at the patient-level. Each dot represents a tumor region (NRF2-enriched and depleted regions are indicated in red and water green, respectively). The dotted line represents the NRF2 signature cutoff identified in clinical cohorts. Tumors with NRF2 transcriptional heterogeneity are those with dots of both colors. **b, c,** COSMIC cancer genes under positive selection in NRF2-enriched non-squamous (b) and squamous (c) tumors with and without *KEAP1* mutations. **d,** Stacked bar chart displaying the sample-level distribution of the four immune subtypes across NRF2-depleted, NRF2-enriched/*KEAP1* wild-type (KEAPness), and NRF2-enriched/*KEAP1*-mutant NSCLC. D: desert, immune-depleted and non-fibrotic, F: fibrotic, immune-depleted and fibrotic, IE: immune-enriched and non-fibrotic, IE/F: immune-enriched and fibrotic. **e,** Dumbbell plot showing the NRF2 signature at the patient level across NRF2 heterogeneous tumors, and the related immune subtyping (D-F: desert/fibrotic, IE-IE/F immune-enriched, either fibrotic or not). **f,** Box plot illustrating T cell infiltration estimated by T cell ExTRECT across the three subgroups compared: NRF2-depleted, NRF2-enriched/*KEAP1* wild-type (KEAPness), and NRF2-enriched/*KEAP1*-mutant tumors.

Extended data figures

Extended Data Figure 1. The NRF2 signature in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) in the TCGA study. a, Box plot for the NRF2 signature in LUAD and LUSC (Wilcoxon test).

Extended Data Figure 2. The NRF2 signature and the tumor immune microenvironment (TIME) in the TCGA study. a, Heatmap illustrating the 29 TIME signatures exploited for generating four immune subtypes (BostonGene) in relation to the NRF2 signature.

Extended Data Figure 3. Optimal cutoff of the NRF2 signature. a, b ROC-AUC curve and bootstrap resampling to identify the optimal cutoff of the NRF2 signature in predicating lack of clinical benefit in a subset of patients included in the SU2C identification cohort (n=93, lung adenocarcinoma patients who did not receive prior tyrosine kinase inhibitors). Lack of durable clinical benefit was defined as a progression-free survival shorter than 6 months.

Extended Data Figure 4. The NRF2 signature in relation to histology in the SU2C identification cohort. a, Box plot showing the difference in the NRF2 signature between squamous and non-squamous NSCLC (Wilcoxon test).

Extended Data Figure 5. Sensitivity analyses in the SU2C cohort upon exclusion of patients who received prior targeted agents. a, b Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS). **c, d** Forest plots illustrating multivariate Cox regression models for PFS and OS. **e**, Stacked bar chart for overall response rate (ORR) (χ^2).

Extended Data Figure 6. The NRF2 signature in relation to histology in the OAK/POPLAR validation cohort. a, Box plot showing the difference in the NRF2 signature between squamous and non-squamous NSCLC (Wilcoxon test).

Extended Data Figure 7. The NRF2 signature reclassified according to *KEAP1* mutations and p53, WNT, and JAK-STAT PROGENy signatures. a, b, c Box plots showing the differences in the p53, WNT, and JAK-STAT PROGENy signatures across the three subgroups compared: NRF2-enriched/*KEAP1*-mutant, NRF2-enriched/*KEAP1* wild-type (KEAPness), and NRF2-depleted.

Extended Data Figure 8: NRF2 transcriptional heterogeneity in non-squamous and squamous tumors in the TRACERx 421. a, b, Dumbbell plots showing NRF2 transcriptional heterogeneity in non-squamous (a) and squamous (b) lung cancer. Each dot represents a tumor region (water green: NRF2-depleted, red: NRF2-enriched), whereas the dotted line indicates the NRF2 signature cutoff.

Extended Data Figure 9: Immune-related processes and the NRF2 signature in the TRACERx 421 study. a, Heatmap illustrating the 29 TIME signatures used for immune subtyping (BostonGene) in relation to the NRF2 signature.

Supplementary figures and tables

Supplementary Figure 1. Relationship between the NRF2 signature and PROGENy core cancer pathway signatures in the TCGA NSCLC study. a-i, Box plots illustrating the relationship between PROGENy signatures (EGFR, hypoxia, MAPK, NFKB, PI3K, TGFb, TNFa, TRAIL, VEGF) and the NRF2 signature (Wilcoxon test). p53, WNT, and JAK-STAT are presented in Figure 2.

Supplementary Figure 2. The NRF2 signature and PD-L1 tumor proportion score (TPS) in the SU2C identification cohort. a, Stacked bar chart showing the distribution of PD-L1 TPS (<50% and ≥50%) in NRF2-enriched and NRF2-depleted NSCLC.

Supplementary Figure 3: The NRF2 signature in advanced NSCLC treated with chemotherapy (docetaxel) in the OAK/POPLAR trials. a, b, Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS). **c**, Stacked bar chart illustrating the differences in overall response rate (ORR) between NRF2-enriched and NRF2-depleted NSCLC (χ^2).

Supplementary Figure 4: Survival analyses in the pooled SU2C/OAK/POPLAR cohort. a, b, Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS). **c**, Stacked bar chart for overall response rate (ORR).

Supplementary Figure 5: Immune subtyping, cancer-associated PROGENy signatures, and transcription factors (DoRothEA) in the pooled SU2C/OAK/POPLAR cohort. a, Stacked bar chart illustrating immune subtypes in NRF2-enriched and NRF2-depleted NSCLC (χ^2). D: desert, immune-depleted and non-fibrotic, F: fibrotic, immune-depleted and fibrotic, IE: immune-enriched and non-fibrotic, IE/F: immune-enriched and fibrotic. **b**, **c**, **d**, Box plots showing the differences in the p53, WNT, and JAK-STAT PROGENy signatures between NRF2-enriched and NRF2-depleted NSCLC (Wilcoxon test). **e**, **f,** Box plots illustrating DoRothEA-predicted activity of the immune-related transcriptional factors STAT1 and STAT3 in NRF2-enriched and NRF2-depleted NSCLC (Wilcoxon test).

Supplementary Figure 6: The NRF2 signature and tumor histology in the pooled SU2C/OAK/POPLAR cohort. a, b, Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS) comparing NRF2-enriched and NRF2-depleted non-squamous NSCLC. **c, d,** Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS) comparing NRF2-enriched and NRF2-depleted squamous lung cancer. **e, f,** Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS) comparing NRF2-enriched squamous and non-squamous tumors. **g, h,** Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS) comparing NRF2-depleted squamous and non-squamous tumors.

Supplementary Figure 7: *KEAP1* mutations and the NRF2 signature in the pooled SU2C/OAK/POPLAR cohort. a, Stacked bar chart illustrating the distribution of *KEAP1* mutations in NRF2-enriched and NRF2-depleted tumors.

Supplementary Figure 8. Distribution of *KEAP1* mutations and copy number alterations (CNAs) according to the NRF2 signature in the TRACERx 421 study. a, Stacked bar chart illustrating the distribution of *KEAP1* clonal mutations in NRF2-enriched and NRF2-depleted NSCLC. **b**, Stacked bar

chart illustrating the distribution of *KEAP1* driver mutations in NRF2-enriched and NRF2-depleted NSCLC. **c,** Stacked bar chart illustrating the distribution of *KEAP1* CNAs (loss of heterozygosity) in NRF2-enriched and NRF2-depleted *KEAP1* wild-type NSCLC.

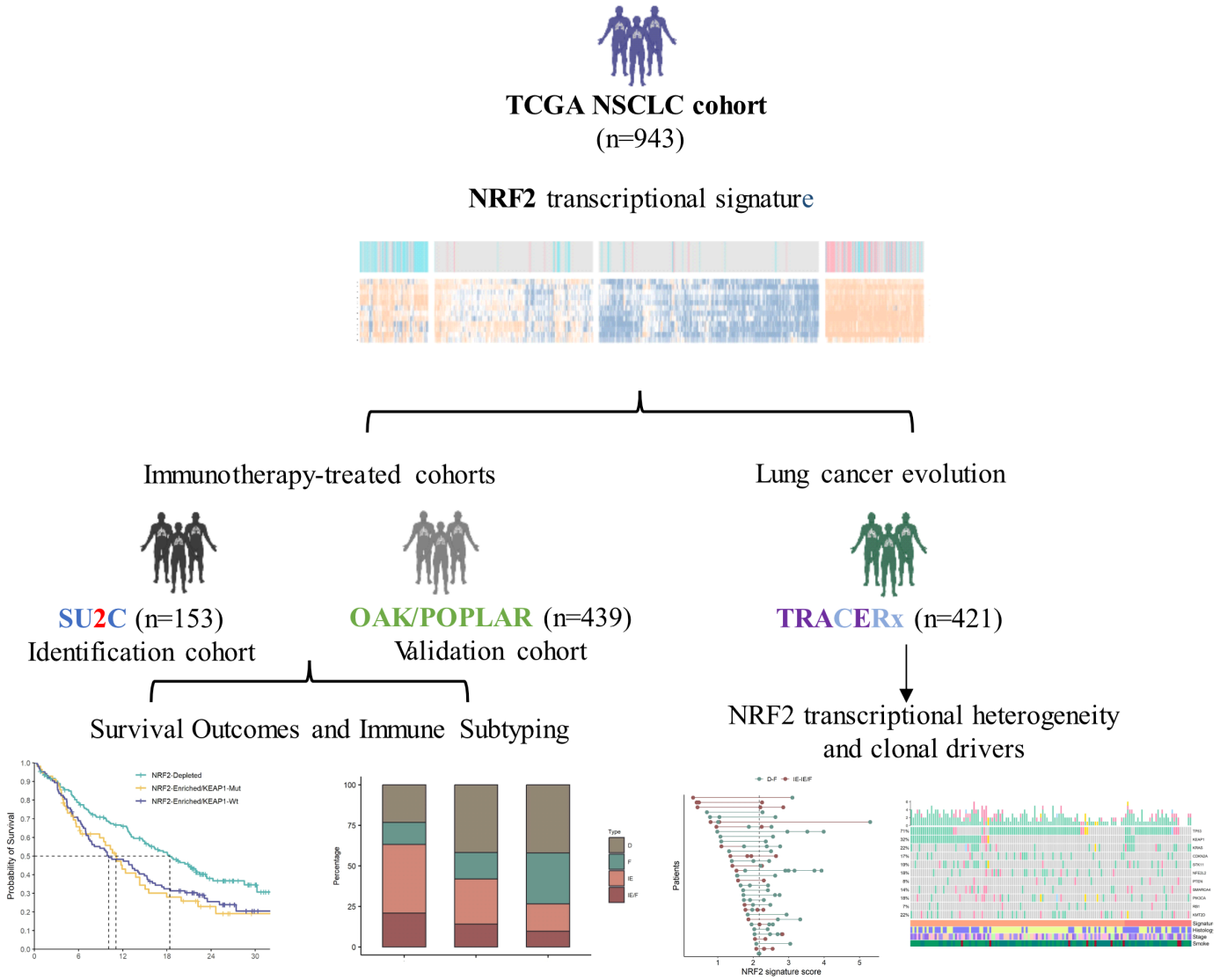
Supplementary Figure 9. NRF2-enriched tumors and clonal status of mutations under positive selection identified in the TRACERx 421. a, Oncoprint illustrating the mutational frequency and clonal status of genes under positive selection in NRF2-enriched NSCLC. Clonal mutations (C) are indicated in water green in the Oncoprint.

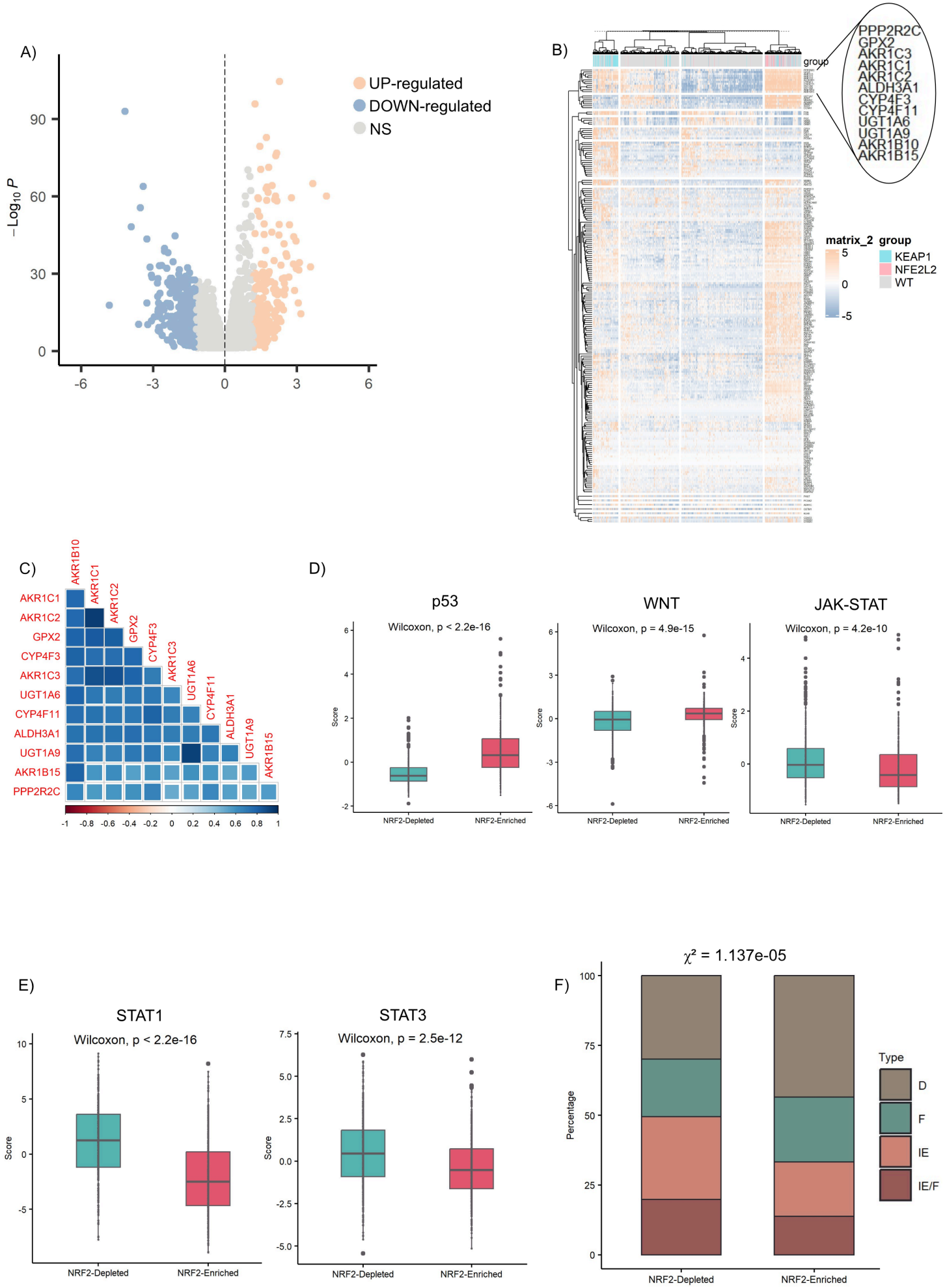
Supplementary Table 1. List of up- and down-regulated genes identified by differential gene expression when comparing *KEAP1-NFE2L2*-mutant and wild-type NSCLC in the TCGA study.

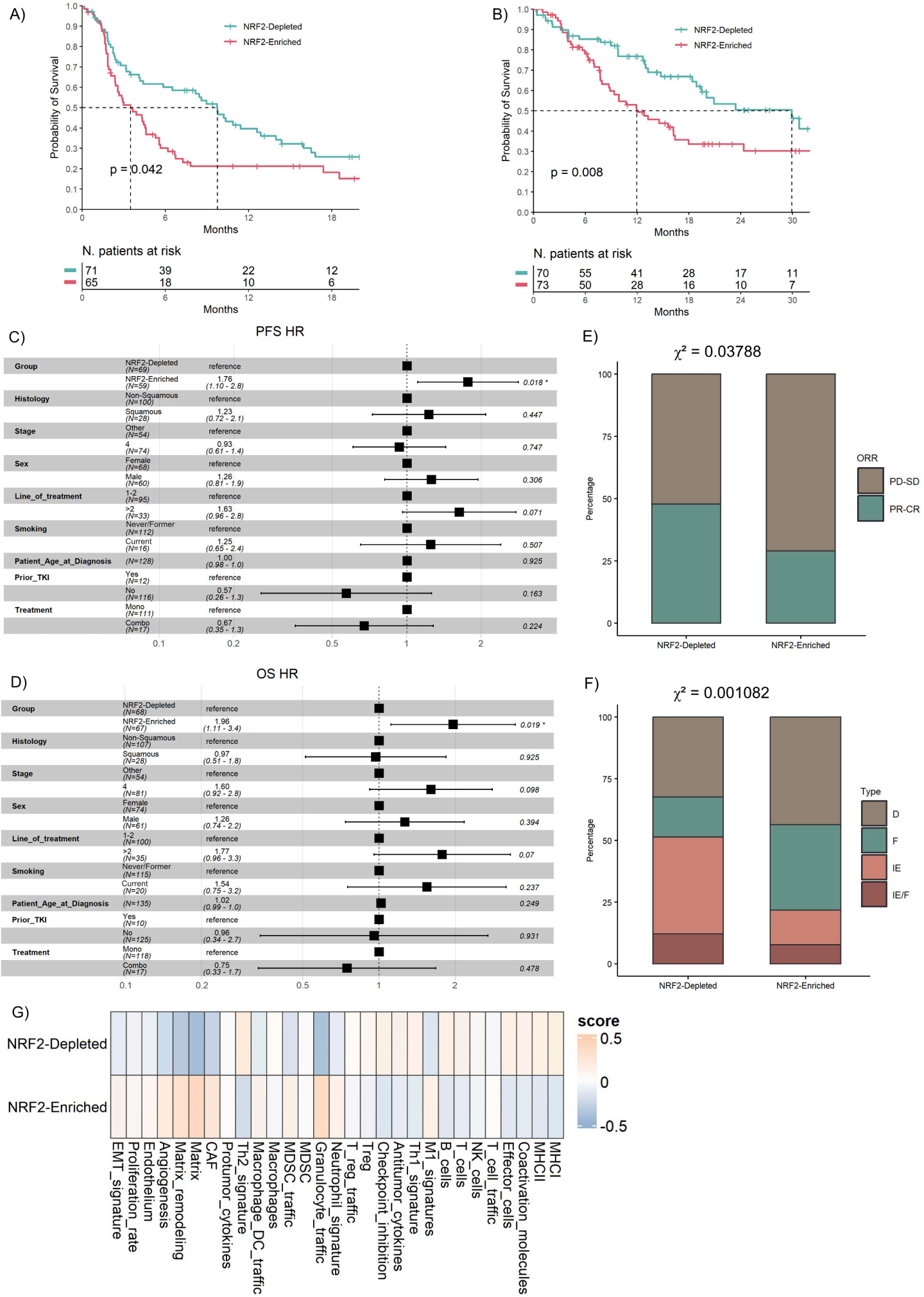
Supplementary Table 2. Clinical features of the patients included in the SU2C identification cohort.

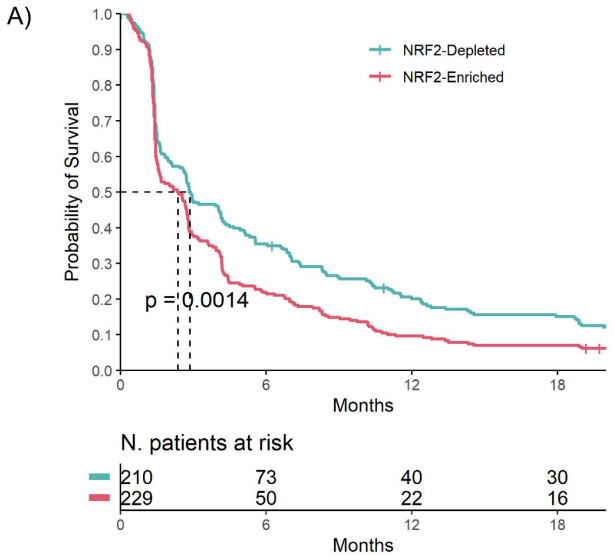
Supplementary Table 3. Association between the NRF2 signature and baseline characteristics of NSCLC patients included in the SU2C identification cohort.

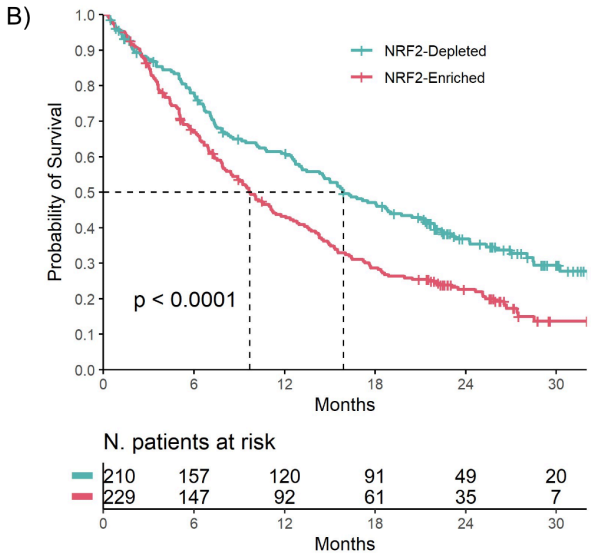
Supplementary Table 4. Clinical features of the patients included in the OAK/POPLAR validation cohort and association with the NRF2 signature.

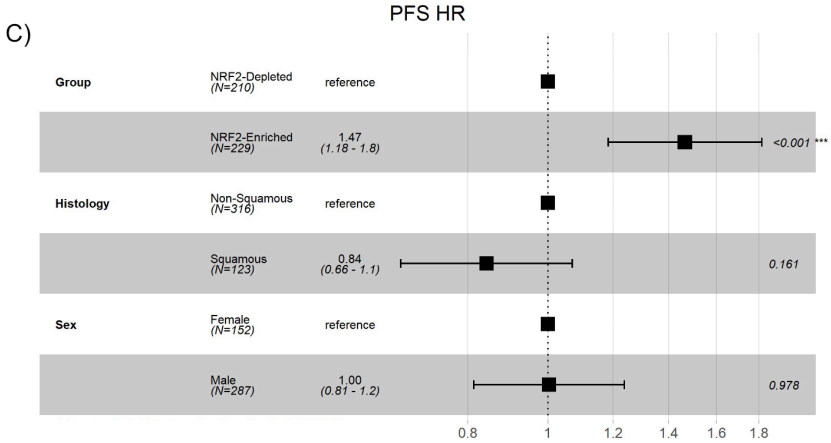

TCGA NSCLC cohort
(n=943)
NRF2 transcriptional signature
Immunotherapy-treated cohorts
Lung cancer evolution
SU2C (n=153)
Identification cohort
OAK/POPLAR (n=439)
Validation cohort
TRACERx (n=421)
Survival Outcomes and Immune Subtyping
NRF2 transcriptional heterogeneity
and clonal drivers
NRF2-Depleted
NRF2-Enriched/KEAP1-Mut
NRF2-Enriched/KEAP1-Wt
Probability of Survival
Percentage
Type
D
F
IE
IE/F
D-F
IE-IE/F
Patients
NRF2 signature score

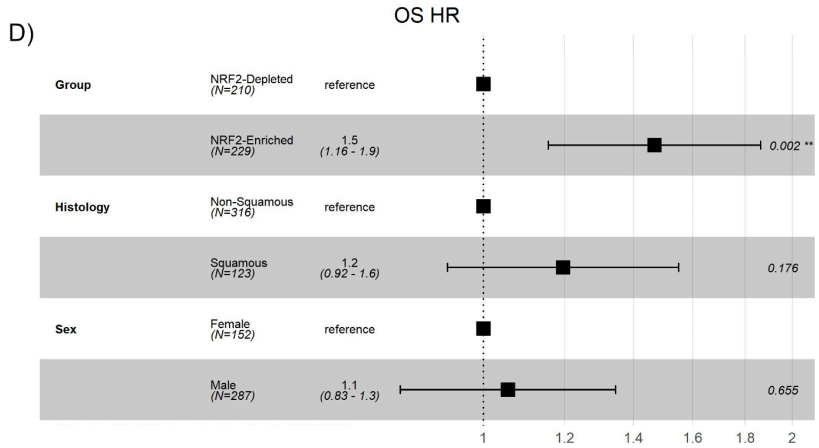

A)
UP-regulated
DOWN-regulated
NS
−Log10 P
B)
group
PPP2R2C
GPX2
AKR1C3
AKR1C1
AKR1C2
ALDH3A1
CYP4F3
CYP4F11
UGT1A6
UGT1A9
AKR1B10
AKR1B15
matrix_2
group
KEAP1
NFE2L2
WT
C)
AKR1B10
AKR1C1
AKR1C2
GPX2
CYP4F3
AKR1C3
UGT1A6
CYP4F11
ALDH3A1
UGT1A9
AKR1B15
PPP2R2C
D)
p53
Wilcoxon, p < 2.2e-16
WNT
Wilcoxon, p = 4.9e-15
JAK-STAT
Wilcoxon, p = 4.2e-10
Score
NRF2-Depleted
NRF2-Enriched
E)
STAT1
Wilcoxon, p < 2.2e-16
STAT3
Wilcoxon, p = 2.5e-12
Score
NRF2-Depleted
NRF2-Enriched
F)
χ² = 1.137e-05
Percentage
Type
D
F
IE
IE/F
NRF2-Depleted
NRF2-Enriched

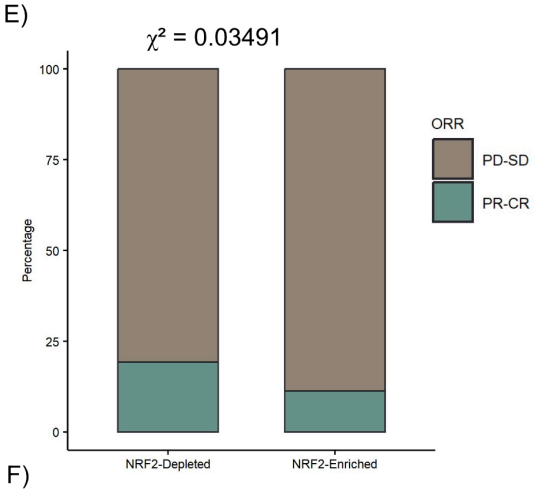

A)
NRF2-Depleted
NRF2-Enriched
Probability of Survival
p = 0.042
Months
N. patients at risk
71 39 22 12
65 18 10 6
B)
p = 0.008
N. patients at risk
70 55 41 28 17 11
73 50 28 16 10 7
C)
PFS HR
Group
NRF2-Depleted (N=69) reference
NRF2-Enriched (N=59) 1.76 (1.10 - 2.8) 0.018 *
Histology
Non-Squamous (N=100) reference
Squamous (N=28) 1.23 (0.72 - 2.1) 0.447
Stage
Other (N=54) reference
4 (N=74) 0.93 (0.61 - 1.4) 0.747
Sex
Female (N=68) reference
Male (N=60) 1.26 (0.81 - 1.9) 0.306
Line_of_treatment
1-2 (N=95) reference
>2 (N=33) 1.63 (0.96 - 2.8) 0.071
Smoking
Never/Former (N=112) reference
Current (N=16) 1.25 (0.65 - 2.4) 0.507
Patient_Age_at_Diagnosis (N=128) 1.00 (0.98 - 1.0) 0.925
Prior_TKI
Yes (N=12) reference
No (N=116) 0.57 (0.26 - 1.3) 0.163
Treatment
Mono (N=111) reference
Combo (N=17) 0.67 (0.35 - 1.3) 0.224
D)
OS HR
Group
NRF2-Depleted (N=68) reference
NRF2-Enriched (N=67) 1.96 (1.11 - 3.4) 0.019 *
Histology
Non-Squamous (N=107) reference
Squamous (N=28) 0.97 (0.51 - 1.8) 0.925
Stage
Other (N=54) reference
4 (N=81) 1.60 (0.92 - 2.8) 0.098
Sex
Female (N=74) reference
Male (N=81) 1.26 (0.74 - 2.2) 0.394
Line_of_treatment
1-2 (N=100) reference
>2 (N=35) 1.77 (0.96 - 3.3) 0.07
Smoking
Never/Former (N=115) reference
Current (N=20) 1.54 (0.75 - 3.2) 0.237
Patient_Age_at_Diagnosis (N=135) 1.02 (0.99 - 1.0) 0.249
Prior_TKI
Yes (N=10) reference
No (N=125) 0.96 (0.34 - 2.7) 0.931
Treatment
Mono (N=118) reference
Combo (N=17) 0.75 (0.33 - 1.7) 0.478
E)
χ² = 0.03788
Percentage
ORR
PD-SD
PR-CR
NRF2-Depleted
NRF2-Enriched
F)
χ² = 0.001082
Type
D
F
IE
IE/F
G)
score
NRF2-Depleted
NRF2-Enriched
EMT_signature
Proliferation_rate
Endothelium
Angiogenesis
Matrix_remodeling
Matrix
CAF
Protumor_cytokines
Th2_signature
Macrophage_DC_traffic
Macrophages
MDSC_traffic
MDSC
Granulocyte_traffic
Neutrophil_signature
T_reg_traffic
Treg
Checkpoint_inhibition
Antitumor_cytokines
Th1_signature
M1_signatures
B_cells
T_cells
NK_cells
T_cell_traffic
Effector_cells
Coactivation_molecules
MHCII
MHCI

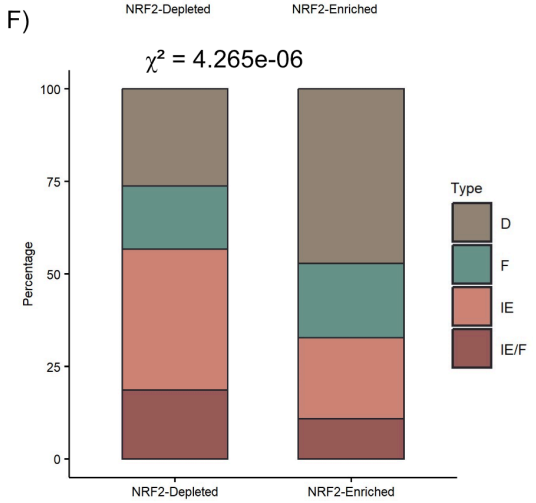

A)
NRF2-Depleted
NRF2-Enriched
Probability of Survival
p = 0.0014
Months
N. patients at risk
210 73 40 30
229 50 22 16
Months

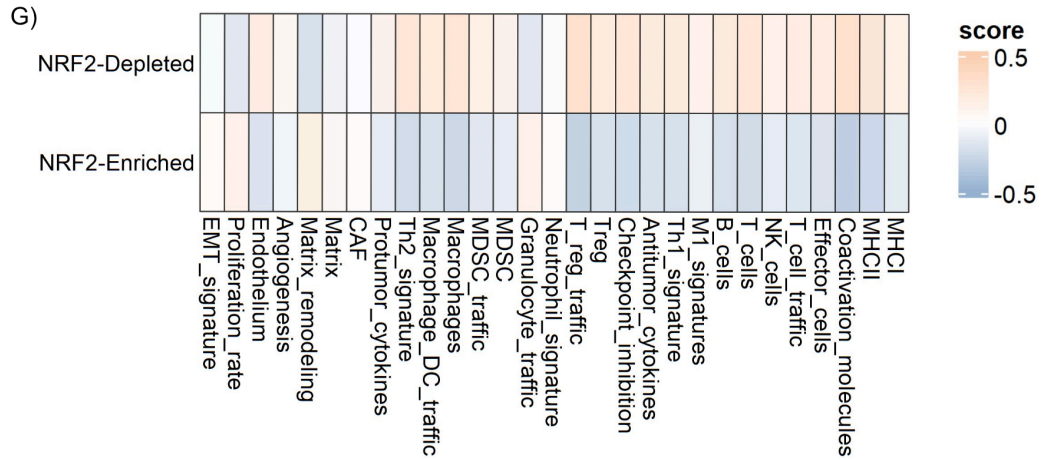

B)
NRF2-Depleted
NRF2-Enriched
Probability of Survival
p < 0.0001
Months
N. patients at risk
210 157 120 91 49 20
229 147 92 61 35 7
Months

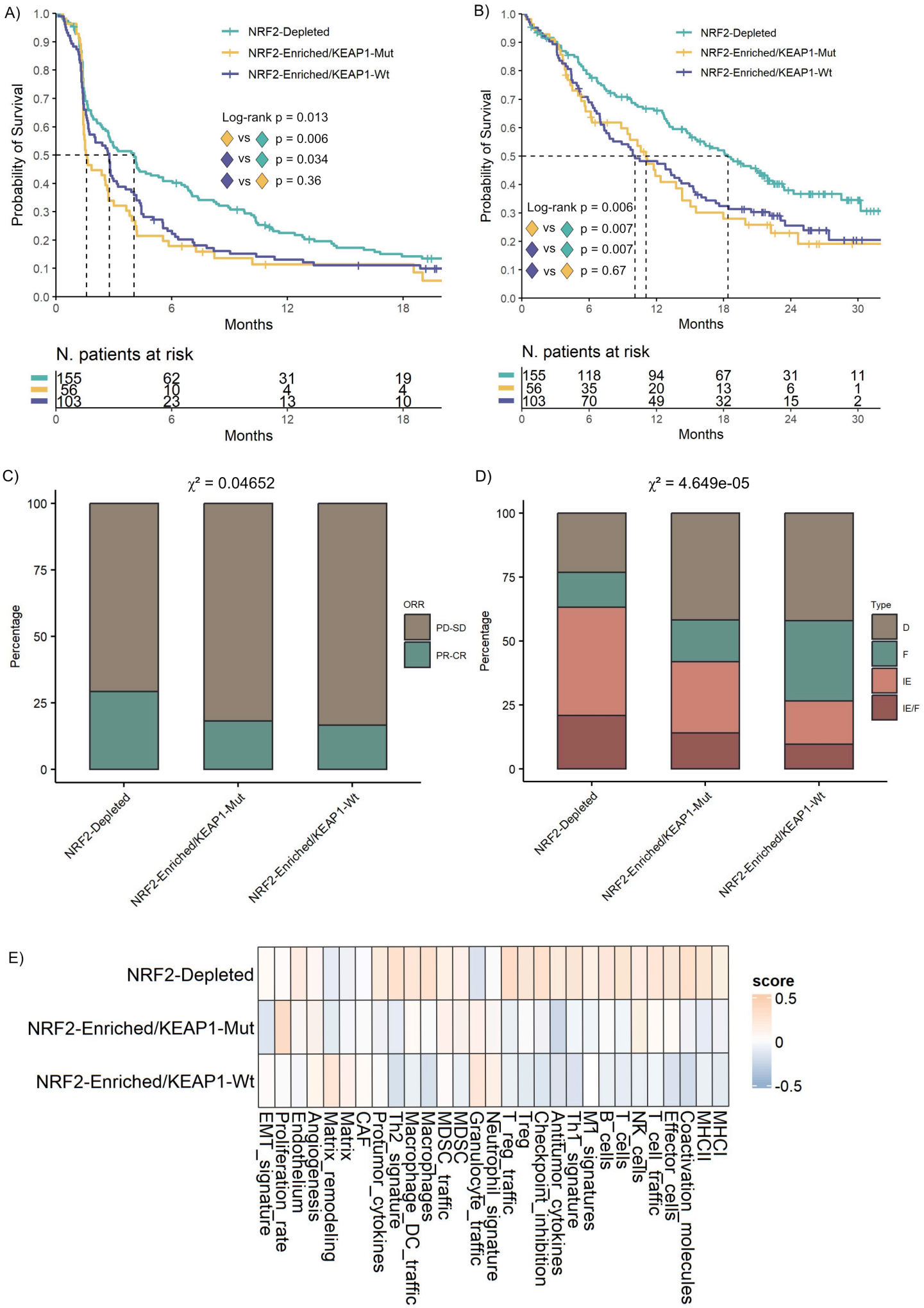

C)
PFS HR
Group
NRF2-Depleted (N=210) reference
NRF2-Enriched (N=229) 1.47 (1.18 - 1.8) <0.001 ***
Histology
Non-Squamous (N=316) reference
Squamous (N=123) 0.84 (0.66 - 1.1) 0.161
Sex
Female (N=152) reference
Male (N=287) 1.00 (0.81 - 1.2) 0.978

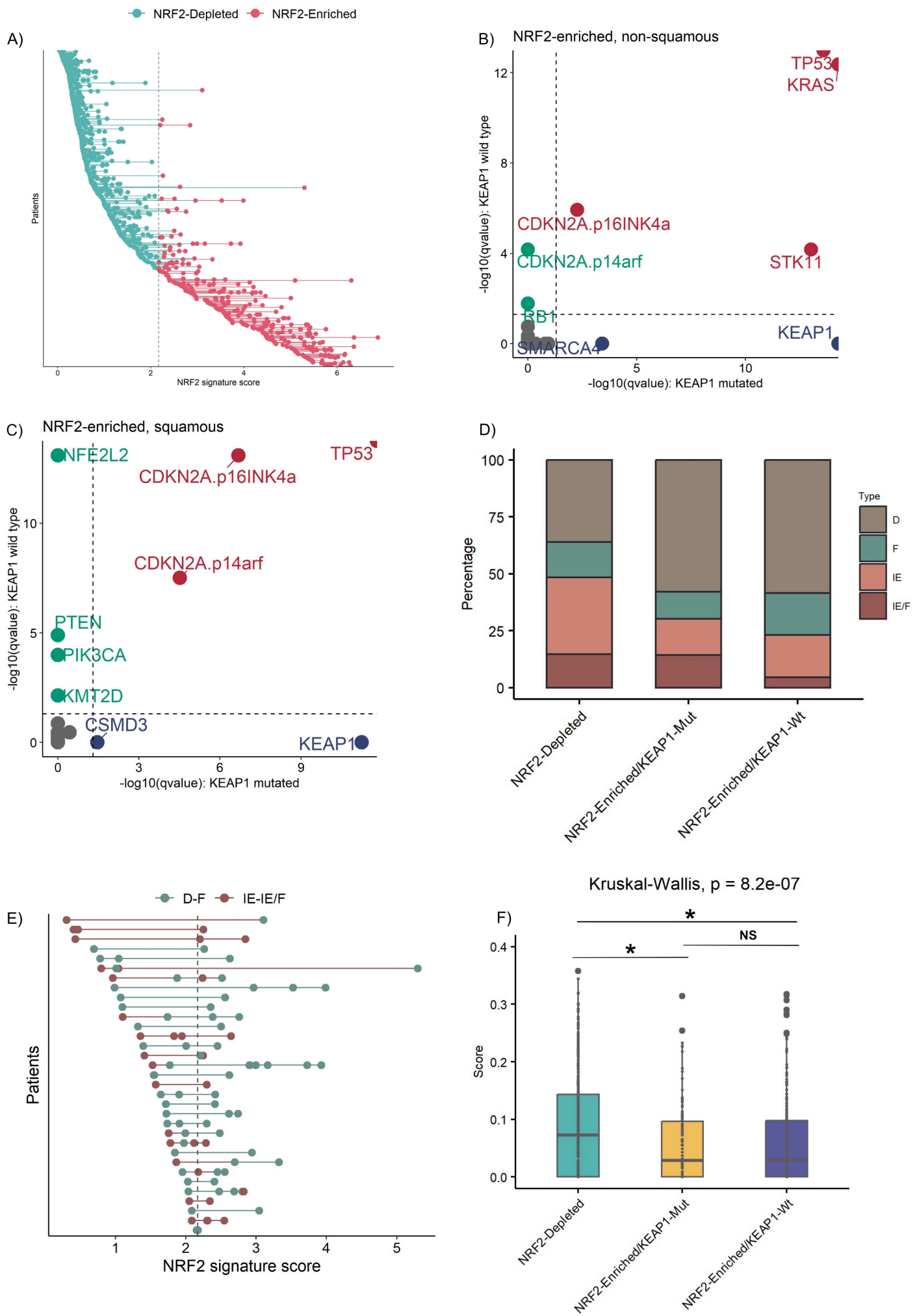

D)
OS HR
Group
NRF2-Depleted (N=210) reference
NRF2-Enriched (N=229) 1.5 (1.16 - 1.9) 0.002 **
Histology
Non-Squamous (N=316) reference
Squamous (N=123) 1.2 (0.92 - 1.6) 0.176
Sex
Female (N=152) reference
Male (N=287) 1.1 (0.83 - 1.3) 0.655


E)
χ² = 0.03491
Percentage
ORR
PD-SD
PR-CR
NRF2-Depleted
NRF2-Enriched


F)
χ² = 4.265e-06
Percentage
Type
D
F
IE
IE/F
NRF2-Depleted
NRF2-Enriched


G)
NRF2-Depleted
NRF2-Enriched
score
0.5
0
-0.5
EMT_signature
Proliferation_rate
Endothelium
Angiogenesis
Matrix_remodeling
Matrix
CAF
Protumor_cytokines
Th2_signature
Macrophage_DC_traffic
Macrophages
MDSC_traffic
MDSC
Granulocyte_traffic
Neutrophil_signature
T_reg_traffic
Treg
Checkpoint_inhibition
Antitumor_cytokines
Th1_signature
M1_signatures
B_cells
T_cells
NK_cells
T_cell_traffic
Effector_cells
Coactivation_molecules
MHCII
MHCI


A)
NRF2-Depleted
NRF2-Enriched/KEAP1-Mut
NRF2-Enriched/KEAP1-Wt
Log-rank p = 0.013
vs p = 0.006
vs p = 0.034
vs p = 0.36
Probability of Survival
Months
N. patients at risk
155 62 31 19
56 10 4 4
103 23 13 10
B)
Log-rank p = 0.006
vs p = 0.007
vs p = 0.007
vs p = 0.67
155 118 94 67 31 11
56 35 20 13 6 1
103 70 49 32 15 2
C)
χ² = 0.04652
Percentage
ORR
PD-SD
PR-CR
D)
χ² = 4.649e-05
Type
D
F
IE
IE/F
E)
score
0.5
0
-0.5
EMT_signature
Proliferation_rate
Endothelium
Angiogenesis
Matrix_remodeling
Matrix
CAF
Protumor_cytokines
Th2_signature
Macrophage_DC_traffic
Macrophages
MDSC_traffic
MDSC
Granulocyte_traffic
Neutrophil_signature
T_reg_traffic
Treg
Checkpoint_inhibition
Antitumor_cytokines
Th1_signature
M1_signatures
B_cells
T_cells
NK_cells
T_cell_traffic
Effector_cells
Coactivation_molecules
MHCII
MHCI


A)
NRF2-Depleted
NRF2-Enriched
Patients
NRF2 signature score
B)
NRF2-enriched, non-squamous
TP53
KRAS
CDKN2A.p16INK4a
CDKN2A.p14arf
STK11
RB1
KEAP1
SMARCA4
-log10(qvalue): KEAP1 wild type
-log10(qvalue): KEAP1 mutated
C)
NRF2-enriched, squamous
NFE2L2
CDKN2A.p16INK4a
TP53
CDKN2A.p14arf
PTEN
PIK3CA
KMT2D
CSMD3
KEAP1
-log10(qvalue): KEAP1 wild type
-log10(qvalue): KEAP1 mutated
D)
Percentage
Type
D
F
IE
IE/F
NRF2-Depleted
NRF2-Enriched/KEAP1-Mut
NRF2-Enriched/KEAP1-Wt
E)
D-F
IE-IE/F
Patients
NRF2 signature score
F)
Kruskal-Wallis, p = 8.2e-07
*
*
NS
Score
NRF2-Depleted
NRF2-Enriched/KEAP1-Mut
NRF2-Enriched/KEAP1-Wt