Article

Title: Transcriptional phenocopies of deleterious *KEAP1* mutations dictate 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], Giulia Schiavoni [1], Irene Terrenato [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, Sapienza University, 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

Mutational models denoting KEAP1-NRF2 pathway activation have emerged as determinants of survival outcomes in non-small cell lung cancer (NSCLC). Hypothesizing that genetically distinct tumors recapitulate the transcriptional footprint of *KEAP1* mutations (KEAPness), we identified a KEAP1-NRF2-related gene set shared by tumors with and without pathway mutations. KEAPness-dominant tumors were associated with poor survival outcomes and immune exclusion in two independent cohorts of immunotherapy-treated NSCLC (SU2C and OAK/POPLAR). Moreover, patients with KEAPness tumors had survival outcomes comparable to their *KEAP1*-mutant counterparts. In the TRACERx421, KEAPness exhibited limited transcriptional intratumoral heterogeneity and an immune-excluded microenvironment, as highlighted by orthogonal methods for T cell estimation. This phenotypic state widely occurred across genetically divergent tumors, exhibiting shared and private cancer genes under positive selection when compared to *KEAP1*-mutant tumors. Collectively, we discovered the pervasive nature of the KEAPness phenotypic driver across evolutionary divergent tumors. This model outperforms mutation-based classifiers in predicting survival outcomes.

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 tumors lacking PD-L1 expression [1]. Similarly, tumor mutational burden (TMB) has been intensively investigated as a 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 immune-disrupting 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]. Moreover, a connection between the magnitude of NRF2 transcription and *KEAP1*/*NFE2L2* mutation pathogenicity (damaging versus tolerated) was proposed [16]. However, the same considerations do not extend to squamous cell lung carcinoma (LUSC). In this context, where *NFE2L2* (the gene encoding for NRF2) 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 patterns, or that exhibit very high PD-L1/TMB levels.

We hypothesized that hallmarks of lethal *KEAP1* mutations are seeded in tumors with intact KEAP1-NRF2 pathway, and that this *KEAP1*-mutant-like state (KEAPness phenotype) shares the anti-immune properties of the canonical *KEAP1*-mutant counterpart, thus intercepting the large proportion of patients who do not obtain significant benefit from ICIs despite their tumors lack putative biomarkers of efficacy/resistance. On this basis, we conceived the following workflow: i) The TCGA was used for modelling KEAPness in NSCLC and to explore its association with the composition of the immune microenvironment at the pan-cancer level [6,7,17,18]; ii) The impact of the KEAPness state 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 prospective TRACERx421 multi-region sequencing study, which contains matched RNA-seq and whole-exome sequencing (WES) data from 347 early-stage NSCLC patients for a total of 947 tumor regions [23], was used to infer the evolutionary landscape of KEAPness tumors, the underlying oncogenic drivers, and KEAPness transcriptional heterogeneity. The workflow of the present study is illustrated in Figure 1.

Results

Identification of the KEAPness state in the TCGA NSCLC cohort and pan-cancer analysis

To provide conceptual ground to our hypothesis that *KEAP1*-loss transcriptional phenocopies frequently occur in NSCLC, we first exploited an approach based on differential gene expression analysis between tumors with and without KEAP1-NRF2 pathway mutations, followed by unsupervised hierarchical clustering and gene co-expression. Evidence that a pool of positively correlated pathway core target genes are broadly expressed in NSCLC, regardless of the presence of *KEAP1* mutations, enabled us to distillate the KEAPness gene set (Figure 2a-c, Supplementary Table 1).

To better characterize this transcriptional configuration, we investigated its relationship with histology, cancer-associated pathways inferred from perturbation experiments (PROGENy) [24], and activity of immune- and pathway-related transcription factors (DoRothEA) [25]. We noticed that LUSC had a significantly higher frequency of KEAPness when compared to LUAD (Supplementary Figure 1). Moreover, KEAPness tumors exhibited a distinct profile of pathway-level gene expression signatures (e.g., p53 and JAK-STAT) compared to KEAPness-free samples (Figure 2d; and Extended Data Figure 1). Differences in the related transcription factors were confirmed by estimating transcription factor-target interactions (Figure 2e). Next, we exploited immune subtyping from bulk RNA-Seq to more accurately examine the tumor immune microenvironment associated with the KEAPness state [26]. We

noted that NSCLC with KEAPness had an immune-excluded microenvironment, which was comparable to that of tumors with *KEAP1* mutations (Figure 2f; and Extended Data Figure 2).

To add a further level of resolution to these observations, we explored the distribution of the *KEAP1*-mutant-like phenotype, and its immunological correlates, across ~ 8,000 *KEAP1* wild-type tumors samples spanning 17 cancer types (Figure 3, and Supplementary Figure 2). Consistently with results from NSCLC, we observed that KEAPness was enriched in squamous tumors, such as head and neck squamous cell carcinoma, esophageal cancer and cervical cancer. In these tumors, KEAPness exhibited an immunological background and repertoire of pathway-level gene expression signatures which are reminiscent of NSCLC. To a similar extent, bladder cancer, colorectal cancer and gastric adenocarcinoma frequently carried the KEAPness configuration, which was accompanied by similar immunological and gene expression features. Next, KEAPness was extremely common in hepatocellular carcinoma, a finding that is consistent with the detoxification mechanisms mediated by the KEAP1-NRF2 pathway. Conversely, hormone-dependent tumors (e.g., breast and prostate cancers), gynecologic tumors (e.g., endometrial and ovarian cancers), and non-epithelial cancers (glioblastoma, skin melanoma, and sarcomas) were prevalently KEAPness-free.

Overall, the immunological background associated with KEAPness, denoting immune exclusion, closely resembles that of *KEAP1*-mutant NSCLC. Moreover, pan-cancer analyses revealed that this phenotypic state is a core feature of squamous tumors which also sporadically affects other histological types.

KEAPness enrichment predicts survival outcomes and immunotherapy efficacy in two independent clinical cohorts

Having defined that KEAPness is a common phenomenon in NSCLC, we next assessed whether a parsimonious binary model exclusively built on the KEAPness gene set predicted clinical outcomes

among advanced NSCLC patients treated with immunotherapy. 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 KEAPness cut point in the SU2C identification cohort (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 and OAK/POPLAR cohorts, and their association with the phenotype investigated, are reported in Supplementary Table 2-4.

In the SU2C cohort (n=153), patients with KEAPness-dominant tumors (i.e., those with a gene expression profile compatible with a KEAPness state when *KEAP1* mutations are not considered in the model) had shorter progression-free survival (PFS) and overall survival (OS) compared to those with KEAPness-free tumors (PFS log-rank p=0.042; OS log-rank p=0.008) (Figure 4a-b; Extended data figure 4a-b). Likewise, progressive disease (PD)/stable disease (SD) and immune exclusion were more frequently observed among KEAPness-dominant tumors (Figure 4c-d, and Supplementary Figure 3a). PD-L1 expression levels were lower among KEAPness-dominant NSCLC, even though this difference was not significant due to the limited size of the subgroups compared (n=71; Supplementary 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 classifier were further improved (Supplementary Figure 5).

Results were entirely confirmed in the independent OAK/POPLAR validation cohort (atezolizumab, n=439) in terms of survival outcomes, immunological correlates, and frequency of the investigated phenotype (PFS log-rank p=0.0014; OS log-rank p<0.0001) (Figure 4e-h; Extended data figure 4c-d,

and Supplementary Figure 3b). Of note, KEAPness-dominant tumors were associated with impaired clinical outcomes also in a separate cohort of chemotherapy-treated patients in the OAK/POPLAR trials (docetaxel, n=452; Supplementary Figure 6).

Further corroborating the consistency of our results, we confirmed that KEAPness-dominant tumors were enriched for the squamous histology (Supplementary Figure 7). However, the clinical implications of the KEAPness-dominant phenotype were independent of histology, even though survival analyses in squamous tumors were limited by the relatively low sample size (Supplementary Figure 8).

Collectively, these data indicate that a streamlined RNA-only classifier frames a large population of patients with tumors exhibiting *KEAP1*-mutant-like traits, and efficiently predicts immunotherapy efficacy and immune exclusion in advanced NSCLC patients. This classifier divides the population in two groups of comparable size, thus indicating that the model does not exclusively identify outlier patients.

Patients with KEAPness and *KEAP1*-mutant tumors have similar survival outcomes when treated with immunotherapy.

Having validated a model exclusively based on gene expression in the two independent cohorts, we used the pooled SU2C/OAK/POPLAR cohort with matched RNA-Seq and WES data to unambiguously demonstrate that KEAPness tumors recapitulate the canonical *KEAP1*-mutant counterparts. To this end, we compared the predictive ability of a combined model that included *KEAP1* mutations in the gene expression classifier to exactly frame the “pure” KEAPness configuration. Data on *KEAP1* mutations were available from 61 patients in the SU2C cohort and 251 patients in the OAK/POPLAR cohort.

In survival analyses, we confirmed the hypothesis that patients with KEAPness had shorter PFS and OS compared to those with KEAPness-free tumors (Figure 5a-b). Importantly, clinical outcomes of the KEAPness group were comparable to those of the *KEAP1*-mutant group (Figure 5a-b). A comparable pattern was observed when considering tumor radiological response to immunotherapy (Figure 5c). KEAPness was significantly more common that *KEAP1* mutations, being observed in approximately 30% of the whole population, thus formally confirming the hypothesis that a model based on gene expression outperforms mutational classifiers in predicting immunotherapy efficacy. Next, we validated in the clinical cohorts that NSCLC with KEAPness more frequently exhibited an immune-excluded microenvironment compared to KEAPness-free tumors, having an immunological phenotype reminiscent of that of *KEAP1*-mutant tumors (Figure 5d-e). The same considerations extend to p53 and JAK-STAT gene expression signatures (Extended Data Figure 5).

Overall, while a binary RNA-based model identified a large proportion of NSCLC that are poorly responsive to ICIs, we confirmed that KEAPness and *KEAP1*-mutant tumors share comparable survival outcomes and immune-related features.

Genetic selection and evolutionary trajectories in the TRACERx 421 cohort

We next investigated the genetic and evolutionary correlates of KEAPness leveraging multi-region WES and transcriptome sequencing data from the TRACERx 421 cohort (previously untreated, early-stage NSCLC) [23, 27].

First, 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 6a-b). In non-squamous NSCLC, KEAPness and *KEAP1*-mutant tumors shared positive selection for *TP53*, *KRAS* and *STK11* (Figure 6a). However, while *SMARCA4* was under positive selection exclusively in the *KEAP1*-mutant group, *CDKN2A* (p14(ARF)) was identified

as a private driver in non-squamous NSCLC with KEAPness (Figure 6a). In squamous tumors, *TP53* and *CDKN2A* were identified as shared drivers between *KEAP1*-mutant and KEAPness tumors (Figure 6b). Conversely, squamous tumors with KEAPness exhibited positive selection for PI3K pathway genes (*PIK3CA* and *PTEN*) and *NFE2L2* (Figure 6b). Most non-silent mutations in these cancer-associated genes were of clonal nature (Figure 6c). We also observed a similar number of clonal non-synonymous mutations and a similar frequency of genome doubling across *KEAP1*-mutant, KEAPness, and KEAPness-free tumors (Figure 6d-e), indicating that clinical findings are unrelated to these events.

We then assessed the relationship between evolutionary trajectories of driver alterations and the KEAPness phenotype using REVOLVER, a method that exploits transfer learning to infer repeated evolution by highlighting hidden evolutionary patterns (Figure 6f) [29]. We discovered that KEAPness pervasively occurred across multiple evolutionary trajectories. The lowest frequency was noted in *EGFR*-driven tumors, a finding that is consistent with the different etiology of *KEAP1*- and *EGFR*-mutant NSCLC (smoking exposure).

Thus, KEAPness exhibits both private and shared cancer genes under positive selection when compared to the *KEAP1*-mutant disease, and intercepts genetically divergent tumors with a similar clinical course and response to immunotherapy.

KEAPness transcriptional heterogeneity and immunological correlates in the TRACERx 421 cohort

We further exploited the TRACERx 421 to investigate the heterogeneity of the KEAPness state and expand the study of the associated immunological features.

First, we observed that KEAPness was a stable phenotype across multiple regions, given that only 31% of KEAPness tumors had at least one region classified as KEAPness-free (Figure 7a). This limited intratumoral heterogeneity denotes a phenotype which is consistent across time and space in tumor

evolution. However, non-squamous NSCLC had a higher fraction of KEAPness heterogeneous tumors when compared to squamous NSCLC, thus further highlighting the predominant role of KEAPness in the squamous histology (Extended Data Figure 6).

We then explored the immunological features associated with KEAPness. As already reported in the TCGA and in clinical cohorts, NSCLC with KEAPness were more commonly immune-excluded compared to KEAPness-free tumors, resembling the *KEAP1*-mutant counterparts (Figure 7b-c). Among KEAPness heterogeneous tumors, namely, KEAPness tumors with at least one KEAPness-free region, we observed a mixed immune-related pattern (Figure 7d). While some tumors were immunologically stable having the same immune subtype in all the tumor regions, in other tumors the acquisition of KEAPness was paralleled by changes in the tumor microenvironment (i.e., from immune-enriched to immune-excluded), suggesting tumor-microenvironment co-evolution (Figure 7d). 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 [30]. KEAPness tumors had lower T-cell infiltration compared to the KEAPness-free counterparts, being comparable to *KEAP1*-mutant tumors (Figure 7e).

Collectively, these analyses conveyed the message that KEAPness is characterized by limited intratumoral heterogeneity, and provided unbiased evidence on the relationship between this novel cancer phenotype and the immune microenvironment considering the reproducibility of the data across multiple independent cohorts and with different methods.

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 [31], and are associated with greater benefit from immunotherapy. However, primary and acquired resistance are common 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 [2, 32-36]. 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 [36]. 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 identifying transcriptional phenocopies of driver *KEAP1* mutations. By integrating survival analyses and computational studies, we reported that: i) Transcriptional phenocopies of *KEAP1* mutations are common in NSCLC and across a variety of cancer types, particularly squamous tumors; ii) A gene expression model denoting the transcriptional output of *KEAP1* mutations predicts inferior survival outcomes, lower response to immunotherapy, and immune exclusion, regardless of the presence of *KEAP1* mutations; iii) KEAPness phenodrivers have limited, but not negligible, transcriptional heterogeneity and associate with immune exclusion.

Furthermore, KEAPness occurs across genetically divergent tumors and, when compared to the *KEAP1*-mutant disease, exhibits some private genomic traits driving tumor evolution.

Our study has important strengths. First, KEAPness was identified in the TCGA NSCLC cohort, and then applied to the clinical cohorts. Second, we used the same classification for 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 the KEAPness configuration were comparable to those observed in the other cohorts, and further confirmed by a different method leveraging DNA sequencing (T cell ExTRECT) [30]. Third, we used a streamlined method to define KEAPness-dominant tumors, which could be easily exploited in the clinical setting. Indeed, while we dissected the clinical, biological and evolutionary patterns of KEAPness tumors, a parsimonious binary model for outcome prediction is largely preferred from a clinical perspective. 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 KEAPness-dominant cases deserves further mention. While current biomarkers mostly focus on outlier patients, the split achieved by our model suggests that it better delineates the actual population of immunotherapy responsive/resistant NSCLC.

From a clinical perspective, this study has several implications. Multiple first-line treatment options are available 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. Nevertheless, the lack of robust biomarkers is a critical hurdle in clinical practice. Our data suggest that patients with KEAPness 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, considering that the

addition of CTLA-4 inhibitors and chemotherapy can be beneficial in immune-excluded tumors with low/negative PD-L1 expression [37-39]. 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 implications for patients with early-stage NSCLC. PD-(L)1 blockade is now approved as adjuvant and neoadjuvant therapy [40-42]. However, the only marker of improved long-term outcomes is pathologic complete response in the case of pre-operative immunotherapy. Indeed, PD-L1 expression and TMB have been inconsistently associated with the risk of recurrence in these settings. Therefore, we envision opportunities for the use of additional biomarkers to identify early-stage NSCLC patients who will benefit from intensification of perioperative treatment.

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) 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, particularly in the SU2C study. Thus, our study was not powered to explore the impact of *KEAP1* in individual cohorts. 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 the consistency of our results. 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 KEAPness pervasively occurs across NSCLC with a different repertoire of driver mutations, and predicts survival outcomes and immune exclusion. 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.

Methods

Cohorts and patients

The TCGA cohort, containing data from 9,229 tumor samples, was used for the identification of NRF2 target genes in NSCLC and for the pan-cancer analyses (https://xenabrowser.net).

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 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 [20-22]. In the available

datasets, 439 patients were treated with atezolizumab, whereas 452 patients received docetaxel 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. 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 KEAPness were investigated in the TRACERx 421 cohort, containing matched RNA-seq and WES data from 347 non-metastatic NSCLC patients for a total of 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. Multivariable 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. When we investigated differences among three molecular subgroups, 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). The relationship between categorical variables was assessed with the Pearson Chi-square test of independence. The Pearson's correlation test was used to assess the connection among the twelve genes included in the KEAPness gene set. 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 KEAPness signature score

The KEAPness gene set score was defined by the mean expression value of the 12 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, KEAPness was defined using a z-score ≥ 0.5, calculated with the mean absolute deviation. In order to identify a KEAPness 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 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 (≥ 2.1678 versus <2.1678) was adopted for all the analyses in the SU2C, OAK/POPLAR, pooled SU2C/OAK/POPLAR, and TRACERx 421. Uncommon transcriptionally silent *KEAP1* mutation (i.e. those without enrichment for the KEAPness gene set) were classified as KEAPness-free.

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 activity of selected transcription factors was 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). 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. Evolutionary clustering was performed with REVOLVER, a method that exploits transfer learning to detect repeated evolution [29]. T cell infiltration was estimated from WES data using the T cell ExTRECT method [30].

Data availability

WES and RNA-Seq data related to the TCGA studies are freely available at https://xenabrowser.net. 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. **M.J.-H.** has consulted for, and is a member of, the Achilles Therapeutics Scientific Advisory Board and Steering Committee; has received speaker honoraria from Pfizer, Astex Pharmaceuticals, Oslo Cancer Cluster and Bristol Myers Squibb and is listed as a co-inventor on a European patent application relating to methods to detect lung cancer (PCT/US2017/028013). This patent has been licensed to commercial entities and, under terms of employment, M.J.-H. is due a share of any revenue generated from such license(s).

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. Caravagna, G., et al. Detecting repeated cancer evolution from multi-region tumor sequencing data. Nat Methods. 15, 707-714 (2018).

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

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

32. 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).

33. 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).

34. 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).

35. 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)

36. 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).

37. 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).

38. 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).

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

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

41. 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).

42. 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 KEAP1-NRF2-associated genes shared by *KEAP1-NFE2L2*-mutant and wild-type cases. The relationship between the KEAPness configuration and the composition of the immune microenvironment was explored at the pan-cancer level (TCGA). 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 KEAPness on survival outcomes and immune subtyping in advanced NSCLC patients treated with immune checkpoint inhibitors. The TRACERx 421 multi-region sequencing study was exploited to investigate the distribution of KEAPness across evolutionary clusters, the underlying genetic drivers, and KEAPness transcriptional heterogeneity. This cohort was also used to validate the immunological features associated with KEAPness.

Figure 2. Identification of KEAPness 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 $\geq$1.2 and $\leq$ -1.2, adjusted p-value $\leq$ 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 KEAPness genes. **d,** Box plots illustrating the differences in the p53, JAK-STAT, and NFkB PROGENy

signatures across KEAPness (KEAPness+), *KEAP1*-mutant (mKEAP1) and KEAPness-free (KEAPness-) tumors. **e**, Box plots illustrating DoRothEA-predicted activity of selected transcription factors in the same subsets. **f,** Stacked bar chart displaying the distribution of immune subtypes. 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.

Figure 3. Frequency of KEAPness at the pan-cancer level (TCGA). Distribution of KEAPness across 17 *KEAP1* wild-type cancer types, along with the associated immune subtyping and selected pathway signatures. Abbreviations: HNSC: head and neck squamous cell carcinoma, ESCA: esophageal carcinoma, CESC: cervical squamous cell carcinoma, GBM: Glioblastoma, THCA: thyroid carcinoma, BRCA: breast cancer, SKCM: skin cutaneous melanoma, UCEC: uterine corpus endometrial carcinoma, SARC: sarcoma, PRAD: prostate adenocarcinoma, SOC: serous ovarian carcinoma, HCC: hepatocellular carcinoma, STAD: stomach adenocarcinoma, KIRK: kidney renal clear cell carcinoma, PAAD: pancreatic adenocarcinoma, COADRED: colorectal cancer, BLCA: bladder cancer.

Figure 4. Survival analyses and immune subtyping in the SU2C identification cohort and OAK/POPLAR validation cohort. a, b, Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS) comparing KEAPness-dominant and KEAPness-free NSCLC in the SU2C cohort. **c,** Stacked bar chart illustrating the differences in the overall response rate (ORR) between patients with KEAPness-dominant and KEAPness-free NSCLC in the SU2C study. Abbreviations: CR: complete response, PR: partial response, SD: stable disease, PD: progressive disease. **d,** Stacked bar chart showing the distribution of immune subtypes in the SU2C cohort. Abbreviations: 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-h,** Survival analyses (e, f), ORR (g), and immune subtypes (h) in the OAK/POPLAR validation cohort.

Figure 5. Survival analyses in the pooled SU2C/OAK/POPLAR cohort to explore the clinical significance of KEAPness in relation to *KEAP1* mutations. a, b, Kaplan-Meier survival curves of progression-free survival (PFS) and overall survival (OS) comparing KEAPness (KEAPness+), *KEAP1*-mutant (mKEAP1), and KEAPness-free (KEAPness-) NSCLC. **c,** Stacked bar chart displaying the differences in terms of overall response rate (ORR) across the subgroups compared. **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. Abbreviations: 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.

Figure 6. Genetic selection and evolutionary trajectories in the TRACERx 421 cohort. a, b, Shared and private COSMIC cancer genes under positive selection in KEAPness and *KEAP1*-mutant non-squamous (a) and squamous (b) tumors (red: shared genes; blue: private genes in KEAPness tumors; yellow: private genes in *KEAP1*-mutant tumors). **c,** Oncoprint illustrating the mutational frequency and clonal status of genes under positive selection in KEAPness and *KEAP1*-mutant NSCLC. Clonal mutations are indicated in water green. **d**, **e,** Number of clonal and subclonal non-synonymous mutations in KEAPness (KEAPness+), *KEAP1*-mutant (mKEAP1) and KEAPness-free (KEAP1ness-) tumors (d) and presence of clonal and subclonal genome doubling (cGD and sGD). **f**, Oncoprint for repeated evolution (REVOLVER) showing the distribution of the KEAPness phenotype across genetically divergent clusters (bottom), and the driver trajectories detected in each cluster (left

side of the oncoprint). The trajectory type is also reported, e.g., initiation: germline (GL)→*TP53*; progression: *KRAS*→*STK11*.

Figure 7. KEAPness intratumoral heterogeneity and immunological correlates in the TRACERx 421 cohort. a, Dumbbell plot showing KEAPness at the patient-level. Each dot represents a tumor region. The dotted line represents the KEAPness cutoff identified in clinical cohorts. Tumors with KEAPness transcriptional heterogeneity are those with both water green and blue dots. **b,** Stacked bar chart displaying the sample-level distribution of the four immune subtypes across KEAPness (KEAPness+), *KEAP1*-mutant (mKEAP1), and KEAPness-free (KEAPness-) 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. **c,** Heatmap illustrating the 29 TIME signatures exploited for generating four immune subtypes (BostonGene) across the same subgroups. **d,** Dumbbell plot showing KEAPness heterogeneous tumors and the related region-level immune subtyping (D-F: desert/fibrotic, IE-IE/F immune-enriched, either fibrotic or not). **e,** Box plot illustrating T cell infiltration estimated by T cell ExTRECT.

Extended data figures

Extended Data Figure 1. KEAPness and pathway-level signatures (PROGENy) in NSCLC (TCGA). Box plots illustrating the relationship between PROGENy signatures and KEAPness (androgen, EGFR, estrogen, hypoxia, MAPK, PI3K, TGFb, TNFa, TRAIL, VEGF, and WNT). JAK-STAT, p53 and NFkB are reported in figure 2.

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

Extended Data Figure 3. Optimal cutoff of the KEAPness signature. a, b ROC-AUC curve and bootstrap resampling to identify the optimal KEAPness cutoff 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. Multivariable Cox regression models in the SU2C and OAK/POPLAR cohorts. a-d, Forest plots illustrating multivariable Cox regression models for PFS and OS in the SU2C (a, b) and OAK/POPLAR (c, d) cohorts.

Extended Data Figure 5. KEAPness and pathway-level signatures (PROGENy) in the pooled SU2C/OAK/POPLAR cohort. Box plots illustrating the relationship between 14 PROGENy signatures and KEAPness, *KEAP1*-mutant, and KEAPness-free NSCLC.

Extended Data Figure 6. KEAPness transcriptional heterogeneity in relation to histology in the TRACERx 421 cohort. a, b, Dumbbell plot showing KEAPness heterogeneity at the individual patient level in non-squamous (a) and squamous (b) tumors. **c,** Stacked bar chart illustrating the frequency of KEAPness heterogeneity according to histology.

Supplementary figures and tables

Supplementary Figure 1. KEAPness in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) in the TCGA. a, Stacked bar chart illustrating the distribution of KEAPness in relation to histology (KEAPness: KEAPness+; KEAPness-free: KEAPness-; *KEAP1*-mutant: mKEAP1).

Supplementary Figure 2. KEAPness and cancer-associated pathway signatures at the pan-cancer level. Box plots illustrating the relationship between KEAPness and PROGENy pathway signatures across the different cancer types (tumors that are prevalently KEAPness-free were excluded).

Supplementary Figure 3. Immune-related processes and the KEAPness-dominant phenotype in the SU2C and OAK/POPLAR cohorts. a, b, Heatmaps illustrating the 29 TIME signatures used for immune subtyping (BostonGene) in relation to the KEAPness state in the SU2C (a) and OAK/POPLAR (b) cohorts.

Supplementary Figure 4. 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 KEAPness-dominant and KEAPness-free NSCLC.

Supplementary 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).

Supplementary Figure 6: Impact of the KEAPness-dominant phenotype 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 KEAPness-dominant and KEAPness-free NSCLC.

Supplementary Figure 7. KEAPness enrichment in relation to histology in the SU2C identification cohort and OAK/POPLAR validation cohort. a, b, Stacked bar charts showing the

difference in the distribution of the KEAPness-dominant phenotype between non-squamous and squamous tumors in the SU2C (a) and OAK/POPLAR (b) cohorts.

Supplementary Figure 8: Impact of the KEAPness-dominant phenotype according to 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 KEAPness-dominant and KEAPness-free non-squamous NSCLC. **c, d,** Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS) comparing the same subgroups in squamous lung cancer. **e, f,** Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS) comparing KEAPness-dominant squamous and non-squamous tumors. **g, h,** Kaplan-Meier survival curves for progression-free survival (PFS) and overall survival (OS) comparing KEAPness-free squamous and non-squamous tumors.

Supplementary Table 1. List of up-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.

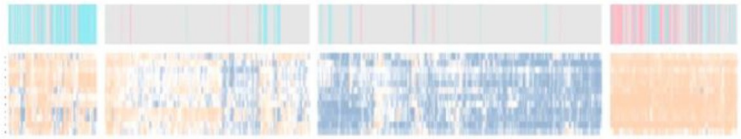
Supplementary Table 3. Association between the KEAPness-dominant phenotype and baseline characteristics of NSCLC patients included in the SU2C identification cohort.

Supplementary Table 4. Clinical features of the patients included in the OAK/POPLAR validation cohort and association with the KEAPness-dominant phenotype.

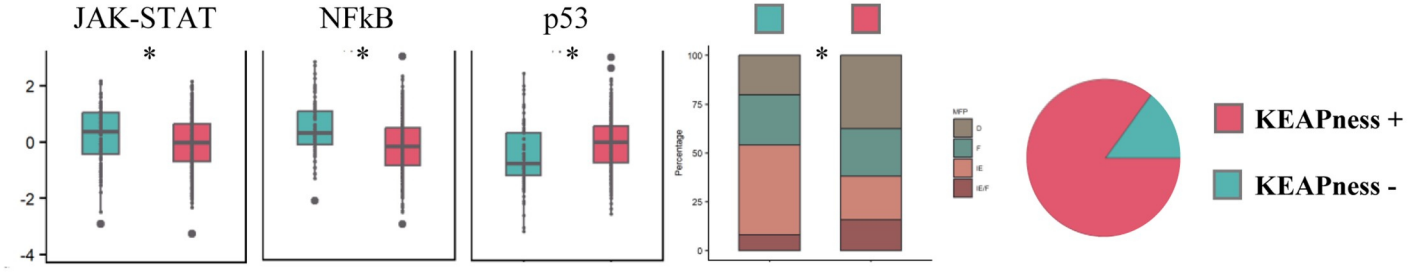
TCGA NSCLC cohort

(n=943)

KEAPness Identification



Pan-cancer analysis of the KEAPness configuration



Immunotherapy-treated cohorts

Lung cancer evolution

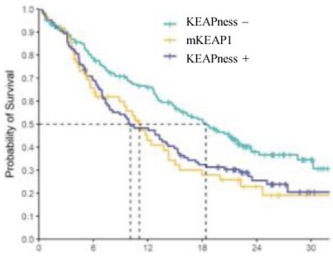
SU2C (n=153)
Identification cohort

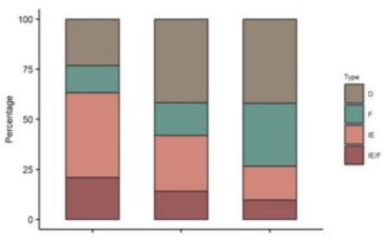
OAK/POPLAR (n=439)
Validation cohort

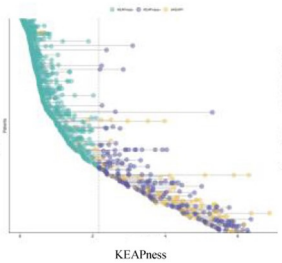
TRACERx (n=421)

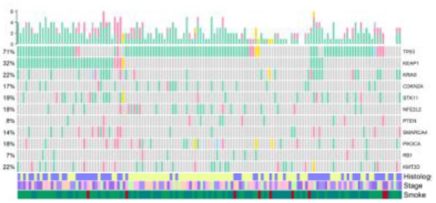
Survival Outcomes and Immune Subtyping

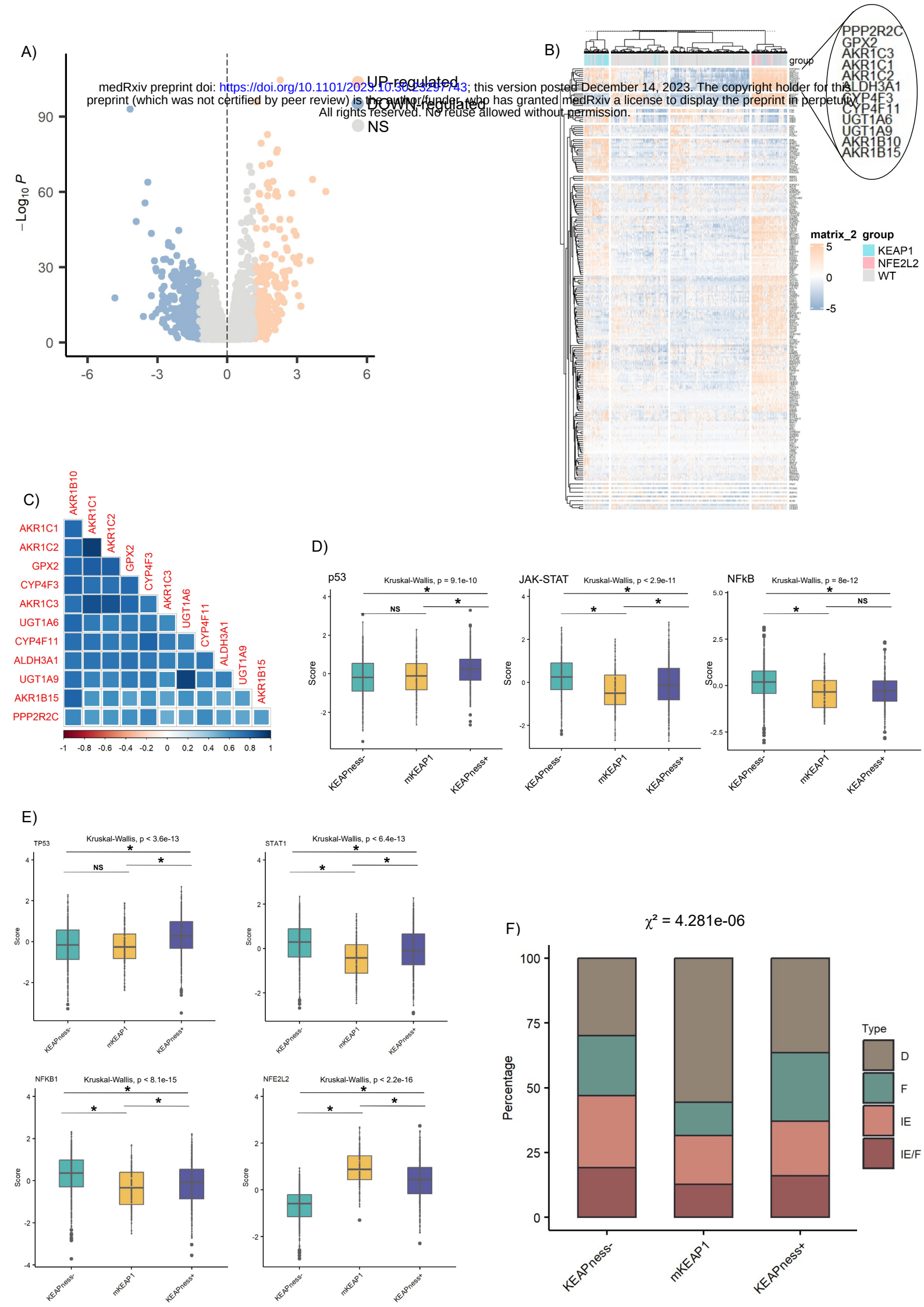
KEAPness heterogeneity and clonal drivers

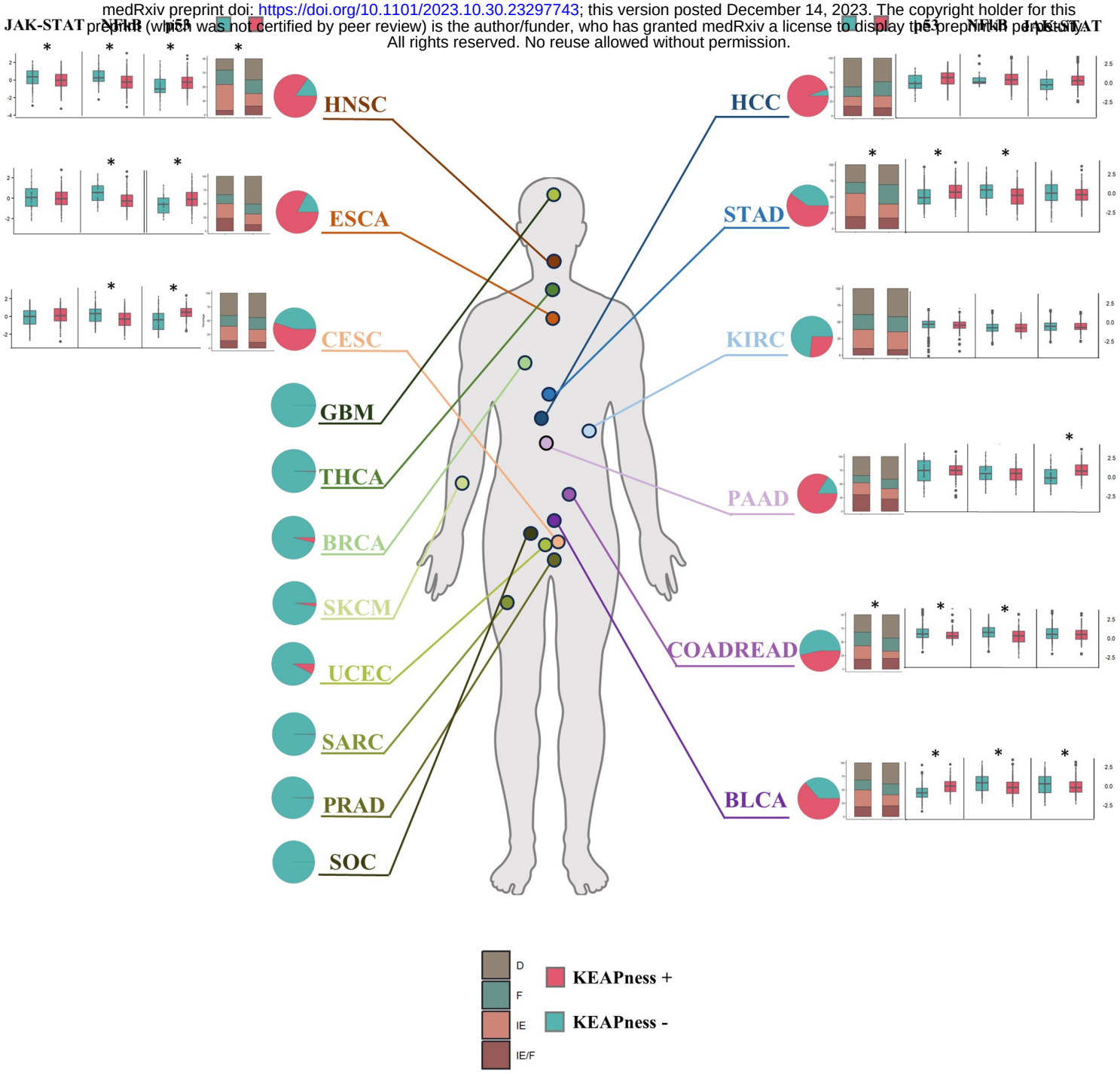






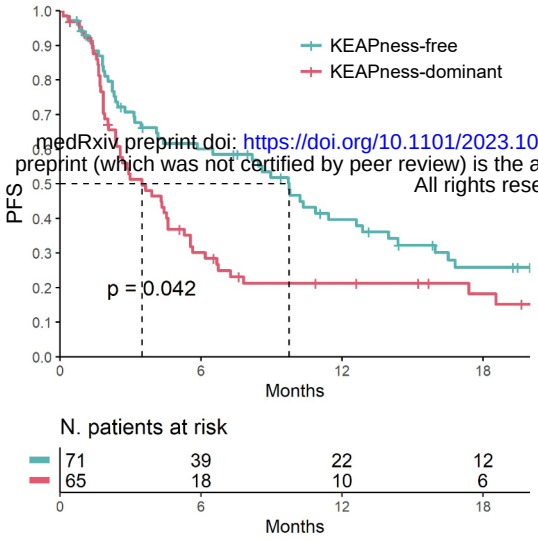




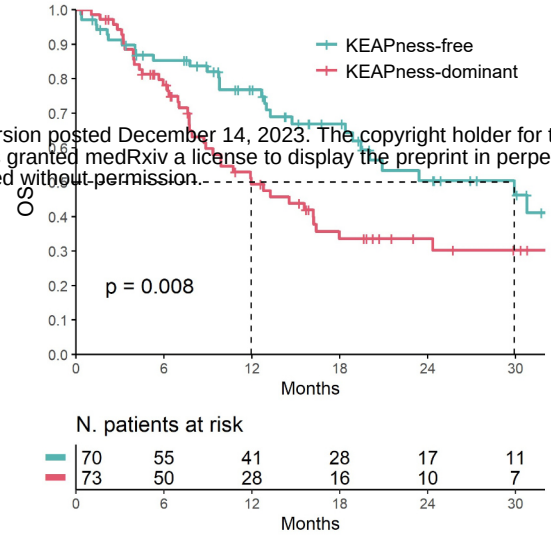


SU2C

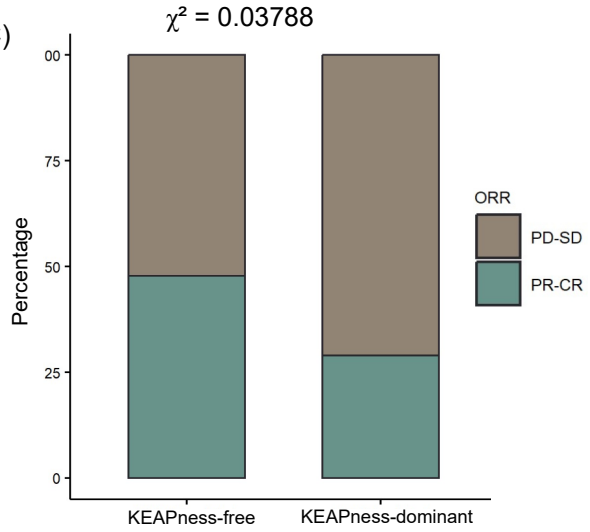
A)



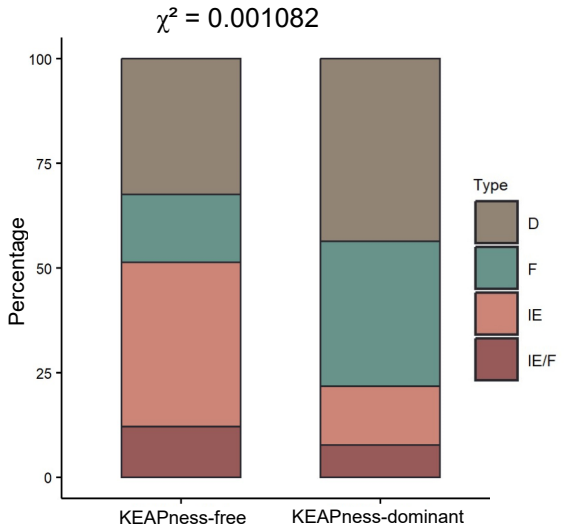
B)



C)

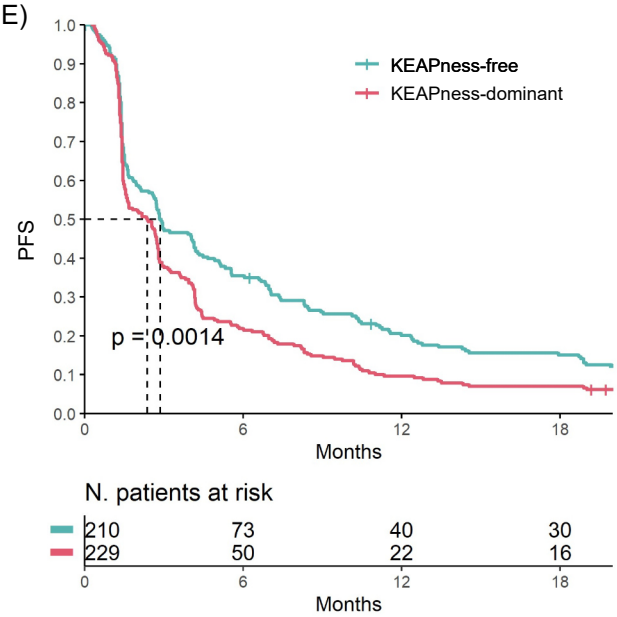


D)

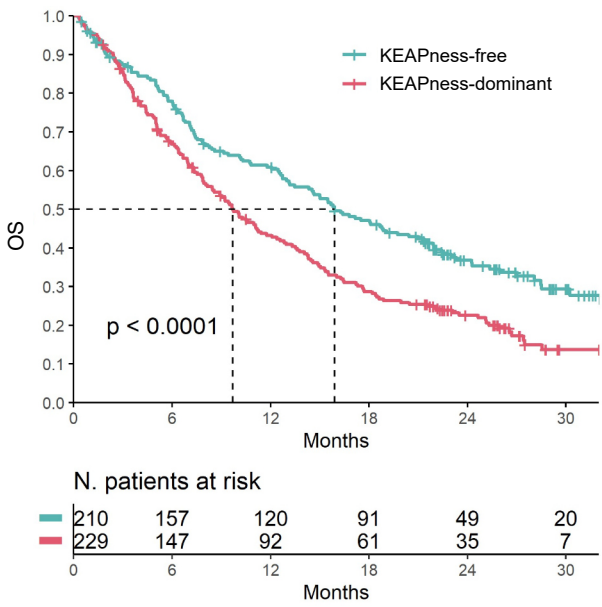


OAK/POPLAR

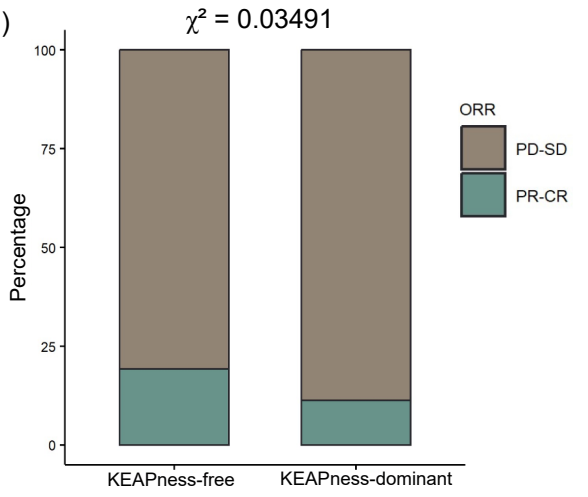
E)



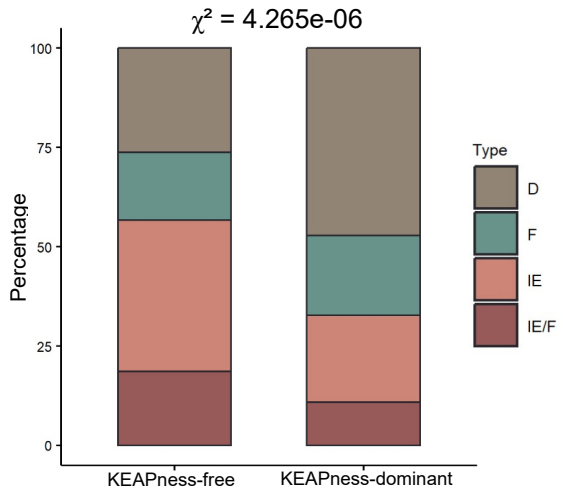
F)

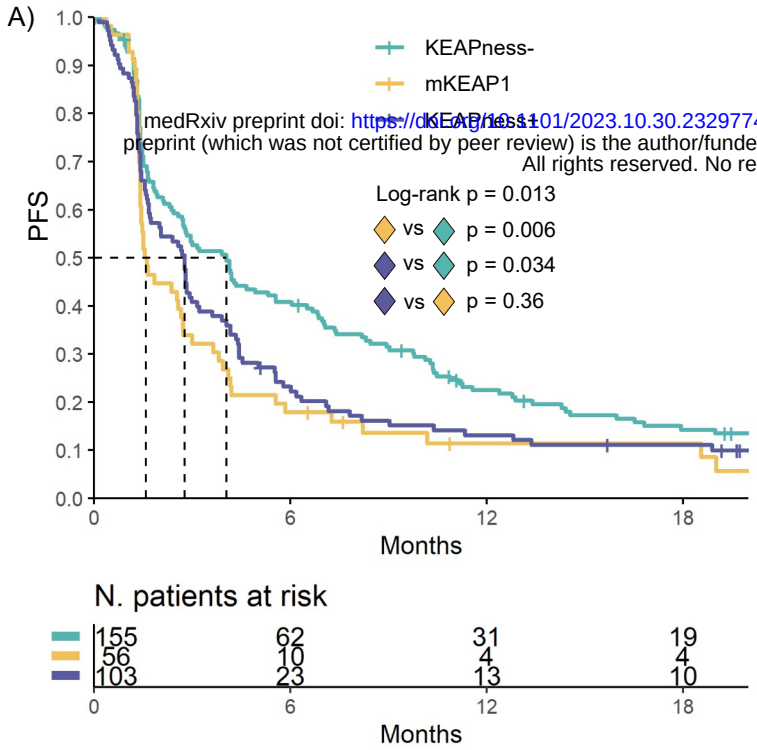


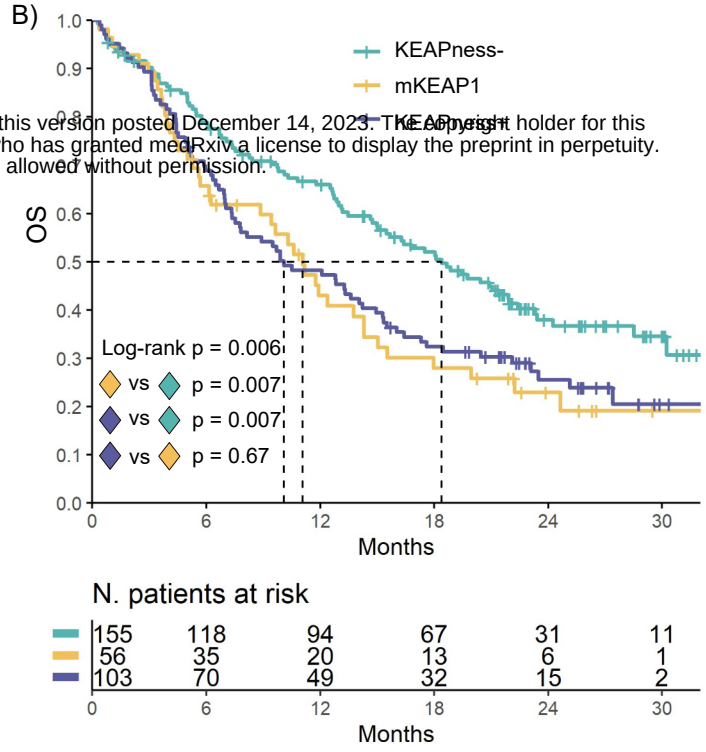
G)

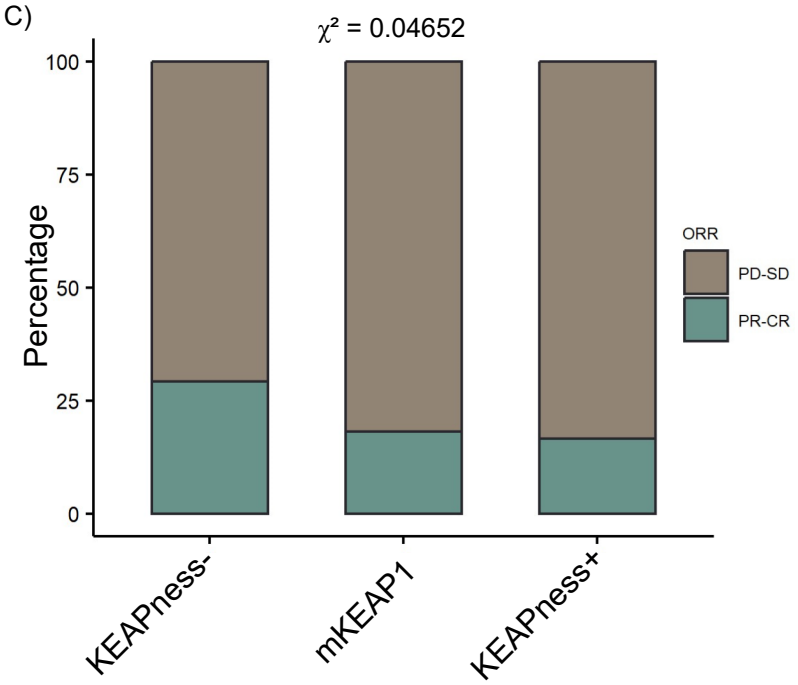


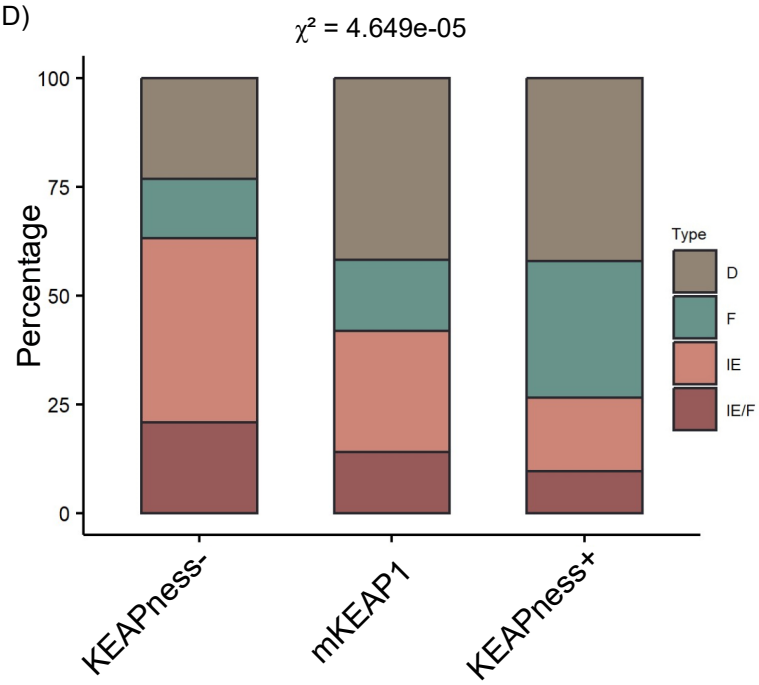
H)

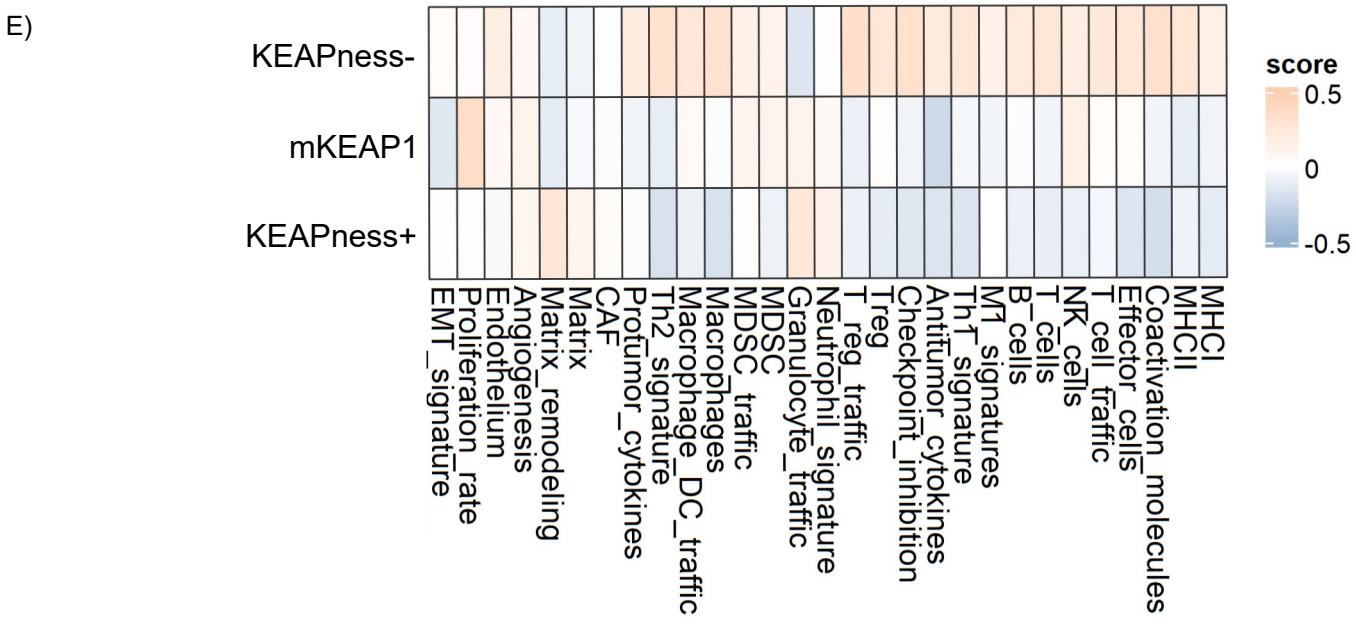


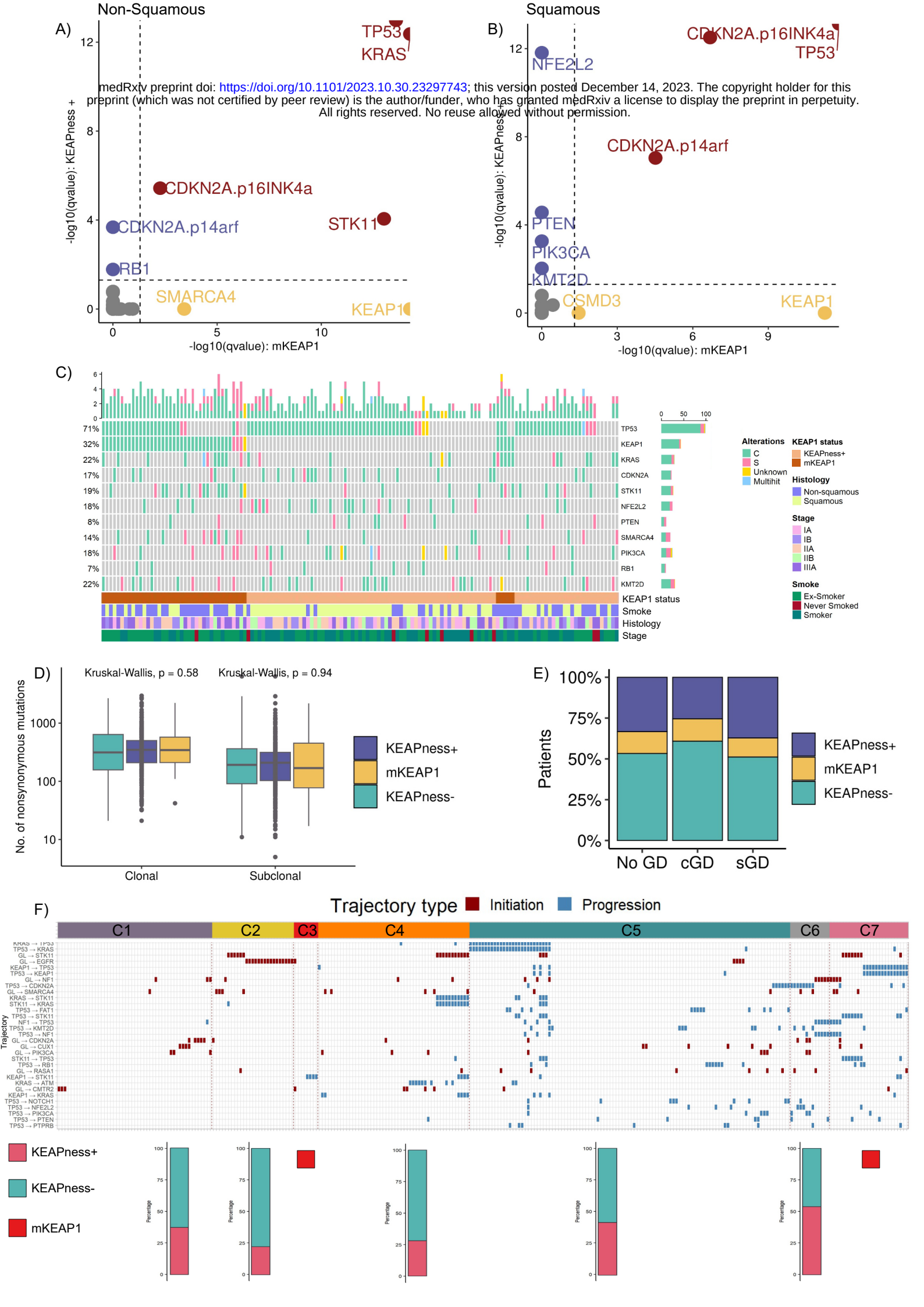


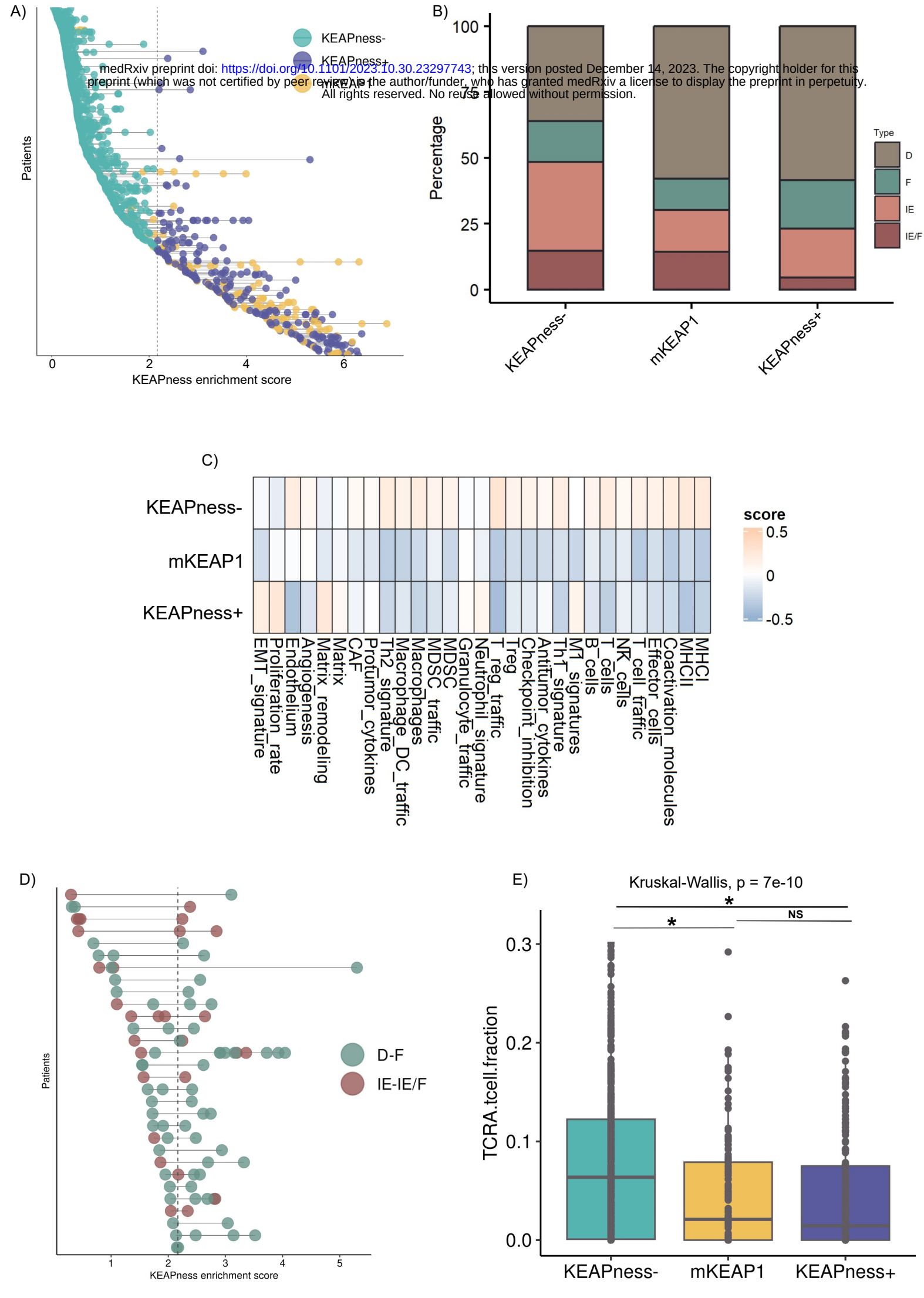












Supplementary Table 1. List of up-regulated genes identified by differential gene expression when comparing *KEAP1-NFE2L2*-mutant and wild-type NSCLC in the TCGA study.

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
PCSK2	1356,416762	4,23402618	0,251090399	16,86255703	8,48379E-64	7,52997E-61
SLC38A11	159,0383515	3,665162525	0,208908236	17,54436586	6,56819E-69	1,0202E-65
CPLX2	1096,61797	3,564178267	0,28526666	12,49419848	8,03015E-36	1,87092E-33
TAC1	21,75561002	3,166359665	0,38090154	8,312803516	9,34672E-17	3,28086E-15
FXYD4	62,56571959	3,099560792	0,2422391	12,79546032	1,73819E-37	4,62832E-35
GUCY2E	18,70284701	3,077846796	0,251175232	12,25378301	1,60338E-34	3,47505E-32
NR0B1	387,380977	3,051677504	0,322836966	9,452689211	3,30235E-21	2,07248E-19
AKR1C4	95,12998911	2,974558805	0,240604431	12,36285964	4,15122E-35	9,21126E-33
AMBP	221,6777239	2,958862705	0,207779591	14,24039146	5,14381E-46	1,84376E-43
SOST	678,4218981	2,897454181	0,273066812	10,61078848	2,65499E-26	2,57741E-24
ABCC2	800,5916557	2,86705172	0,197543602	14,5135134	9,94913E-48	3,70884E-45
CYP4F3	1452,273275	2,847988629	0,224484555	12,68679098	6,99954E-37	1,78718E-34
CES1	17088,44522	2,766366779	0,16491976	16,77401649	3,78067E-63	3,06382E-60
AKR1C1	28312,55313	2,689225537	0,176570292	15,23033977	2,22433E-52	1,06306E-49
GKN2	302,1802682	2,666082948	0,227687475	11,70939662	1,14089E-31	1,95092E-29
DIRAS2	120,4465937	2,580160099	0,19508349	13,22592749	6,21656E-40	1,75561E-37
AKR1C2	26071,85121	2,540698636	0,189590049	13,40101259	5,96438E-41	1,73703E-38
AKR1B10	11898,53897	2,512423062	0,234306486	10,72280627	7,95524E-27	8,28367E-25
FGFBP2	532,763464	2,478585159	0,221039912	11,21329239	3,50893E-29	4,51055E-27
SLC14A2	179,5779423	2,474069877	0,242225983	10,21389138	1,71832E-24	1,42346E-22
CYP4F11	2907,917259	2,472585884	0,21222837	11,65059076	2,27874E-31	3,79226E-29
TREML3	85,79167112	2,463361703	0,175247018	14,05651137	7,02656E-45	2,42533E-42
CYP4F2	41,6226854	2,432426206	0,277406813	8,768444361	1,81152E-18	8,09709E-17
UGT1A6	1933,902694	2,389933452	0,206306037	11,58440872	4,94366E-31	7,94352E-29
PGC	19651,73296	2,356586813	0,256973171	9,170555824	4,70586E-20	2,55722E-18
UGT1A9	104,2570689	2,33459464	0,315802652	7,39257452	1,44013E-13	2,92084E-12
HMP19	11,23363856	2,327419976	0,18439454	12,6219571	1,598E-36	3,97135E-34
SPINK4	10,86980771	2,323825689	0,250911256	9,261544205	2,01497E-20	1,14276E-18
CBR1	10656,55049	2,275890652	0,10234099	22,23830991	1,4637E-109	2,7281E-105

CHGA	97,82524354	2,263386035	0,216112392	10,47318952	1,14712E-25	1,03792E-23
KIAA0319	382,375528	2,261255972	0,148225891	15,25547233	1,51396E-52	7,42595E-50
C4orf17	4,313552979	2,255977268	0,290272012	7,771942094	7,72918E-15	1,94945E-13
AKR1C3	14861,64677	2,245280188	0,153669364	14,611111134	2,38594E-48	9,46203E-46
UGT1A3	11,80029561	2,201248927	0,231765196	9,497754484	2,14465E-21	1,37368E-19
CGA	61,06237234	2,195711641	0,343090285	6,399807095	1,55573E-10	1,75955E-09
PAH	126,7038543	2,194320462	0,233646619	9,391620884	5,90839E-21	3,5989E-19
DDC	316,6744422	2,186239237	0,213754959	10,227782552	1,4889E-24	1,24446E-22
AKR1B15	477,479742	2,169946127	0,228002365	9,517208845	1,77893E-21	1,1513E-19
CABYR	885,6208325	2,168187901	0,130705299	16,588370225	8,45846E-62	5,83916E-59
ALDH3A1	11316,67634	2,160275004	0,206213425	10,475918344	1,11451E-25	1,01333E-23
NQO1	14799,43984	2,147471188	0,112613383	19,069413744	4,53359E-81	1,40836E-77
STXBP5L	104,622659	2,140789979	0,246127267	8,697898475	3,38087E-18	1,4587E-16
INHA	300,5895854	2,12953422	0,19949536	10,674605224	1,33825E-26	1,33389E-24
ADCY8	131,6857495	2,121181359	0,299345677	7,086059762	1,37984E-12	2,36604E-11
SRXN1	10713,13188	2,112387745	0,111564686	18,934197088	5,96177E-80	1,58745E-76
OSGIN1	2012,497353	2,105728394	0,124869452	16,863439088	8,35811E-64	7,52997E-61
PHOX2B	5,597906111	2,103983161	0,265904529	7,912551051	2,52167E-15	6,94261E-14
FGA	8305,464507	2,103821184	0,263181769	7,99379529	1,30847E-15	3,79884E-14
HGD	698,1969866	2,096991433	0,202721806	10,344182887	4,44694E-25	3,85518E-23
TM4SF20	6,473628244	2,093029625	0,243611795	8,591659639	8,5722E-18	3,50389E-16
LOC100190940	65,91349341	2,078034646	0,345893562	6,007728605	1,8814E-09	1,70313E-08
NECAB2	84,03529913	2,070923723	0,185259775	11,178485555	5,19693E-29	6,63462E-27
TRIM16L	1899,766838	2,065614455	0,123813126	16,683323722	1,73303E-62	1,29207E-59
INSL4	116,0367525	2,058868812	0,429651784	4,791947551	1,6517E-06	8,15525E-06
F2RL2	405,7505644	2,051624345	0,158548839	12,940014922	2,67578E-38	7,33438E-36
PRKAG3	6,833073851	2,048418831	0,249565208	8,207950322	2,24996E-16	7,47541E-15
GPX2	10216,83391	2,018939826	0,188739221	10,696980833	1,05149E-26	1,06514E-24
ZFP42	96,87952714	1,985745627	0,287414711	6,908990911	4,88113E-12	7,569E-11
CBR3	1051,021617	1,96829058	0,112821483	17,446061966	3,688E-68	5,28773E-65
SLC7A11	3286,669687	1,947023821	0,131657213	14,7885844	1,73566E-49	7,35248E-47

				9		
LOC200726	3,429217811	1,934385536	0,332448675	5,818599023	5,93429E-09	4,89638E-08
CDK5R2	49,8581717	1,925005602	0,16981806	11,33569423	8,73368E-30	1,22396E-27
GABRA5	47,37968705	1,878365525	0,298621437	6,290122848	3,17215E-10	3,38635E-09
SYT4	17,38518723	1,865437735	0,26668851	6,994818553	2,65602E-12	4,31234E-11
RAB3B	498,3580402	1,865048801	0,209564212	8,899653178	5,60215E-19	2,63294E-17
SHISA9	144,589193	1,85039949	0,249372114	7,420234212	1,16913E-13	2,41056E-12
HPD	32,06115538	1,846142093	0,152143939	12,13418102	6,96029E-34	1,45767E-31
TXNRD1	20250,2495	1,822930635	0,099180218	18,3799821	1,90031E-75	4,42749E-72
SCG2	363,1922852	1,815897757	0,150843471	12,03829208	2,23536E-33	4,48009E-31
UCHL1	5265,452818	1,80925236	0,15136878	11,95261245	6,29159E-33	1,23441E-30
DCC	197,5972083	1,807446025	0,199265367	9,070547756	1,1842E-19	6,04722E-18
CPS1	7180,436979	1,802805462	0,249229558	7,233513872	4,70654E-13	8,83437E-12
PTGR1	3967,525801	1,793873094	0,105043727	17,0773938	2,1867E-65	2,26433E-62
UGT1A7	553,7945329	1,787342477	0,406459417	4,397345475	1,09583E-05	4,55918E-05
WNT11	383,3682105	1,779635082	0,157879608	11,27210225	1,80215E-29	2,41657E-27
CLDN8	392,6241942	1,774411739	0,199285957	8,903847329	5,39436E-19	2,55841E-17
ADAM23	1963,902774	1,760488224	0,214390613	8,211591908	2,18275E-16	7,26505E-15
S100P	7808,489884	1,757157301	0,187997642	9,346698611	9,04262E-21	5,33371E-19
GAST	34,8578073	1,75649539	0,304309449	5,772069826	7,83037E-09	6,32366E-08
GCLM	3225,974505	1,753048824	0,101011396	17,35496086	1,80924E-67	2,24816E-64
CCNA1	299,0009391	1,751742301	0,191675178	9,139119213	6,2963E-20	3,35305E-18
ADCY10	71,76604543	1,747783779	0,212576365	8,221910181	2,00287E-16	6,70225E-15
TINAG	30,03832089	1,746342297	0,324433266	5,382747332	7,33575E-08	4,78918E-07
PANX2	1110,106483	1,736605849	0,104481898	16,62111692	4,9011E-62	3,51352E-59
JAKMIP3	186,7963014	1,733617022	0,147360007	11,76450152	5,94766E-32	1,03606E-29
G6PD	10620,06109	1,730048323	0,087355737	19,80463315	2,71533E-87	1,26528E-83
SPANXC	3,562214981	1,725971011	0,456016228	3,784889451	0,000153777	0,000489539
ME1	3200,103384	1,715737436	0,099376928	17,26494747	8,63824E-67	1,0063E-63
CA8	227,0300824	1,703842966	0,152559959	11,16834969	5,82531E-29	7,33635E-27
AKR1CL1	5,456556656	1,699128812	0,20724551	8,198627874	2,43146E-16	8,04974E-15
TSPAN7	2523,824354	1,694957056	0,124315331	13,6343365	2,50256E-42	7,93549E-40

				3		
VWA5B2	73,15869693	1,693794842	0,160925637	10,52532632	6,60322E-26	6,09295E-24
WFDC12	21,40559052	1,688907034	0,209140295	8,075474076	6,7215E-16	2,06736E-14
ADH7	3101,964369	1,681710829	0,278722628	6,033635803	1,60311E-09	1,4785E-08
NELL1	487,4299977	1,676001823	0,253112906	6,621558143	3,55432E-11	4,54383E-10
SPANXA2	0,774344044	1,667365564	0,734301666	2,270681984	0,023166236	0,04330079
UGT1A1	25,97784498	1,66630147	0,235159741	7,085827971	1,38215E-12	2,36783E-11
PCSK1	380,3280983	1,662955211	0,193592425	8,589980779	8,69839E-18	3,53994E-16
NPAS3	101,8350971	1,661502518	0,156934944	10,587205558	3,41639E-26	3,29938E-24
MAP2	1780,657027	1,637209934	0,111775436	14,6473142	1,40147E-48	5,67869E-46
AGPAT9	627,2863486	1,633781127	0,106868002	15,287841999	9,21535E-53	4,90757E-50
KCNE4	867,1564402	1,626170118	0,128530164	12,6520504	1,08993E-36	2,74529E-34
GSTM3	3845,464622	1,584713033	0,135605333	11,686214655	1,49921E-31	2,51745E-29
HAO1	1,4758602	1,582369091	0,312706843	5,060231742	4,18747E-07	2,34456E-06
KYNU	1936,829977	1,581571455	0,130532093	12,116341811	8,65362E-34	1,79216E-31
TREML4	11,73395686	1,581408502	0,15899494	9,946281952	2,61781E-23	1,99157E-21
ORM2	276,1124758	1,564983063	0,200225625	7,816097772	5,44861E-15	1,41051E-13
SLC47A2	108,4081651	1,560620116	0,193261335	8,075180235	6,7377E-16	2,06893E-14
C1orf61	91,81044894	1,555375822	0,197733739	7,866011293	3,66128E-15	9,79089E-14
BHMT2	448,8605283	1,544343029	0,182022105	8,484370792	2,16889E-17	8,42208E-16
RSPO3	250,2578423	1,542934837	0,138867991	11,110802559	1,11156E-28	1,36305E-26
C14orf162	23,92254211	1,535558082	0,227338643	6,754496567	1,43332E-11	2,0087E-10
LIN7A	78,70767676	1,531773038	0,141369112	10,835273822	2,34264E-27	2,55347E-25
CT45A3	11,51129912	1,522909479	0,475470534	3,20295238	0,001360265	0,003492764
RFX6	5,984553597	1,518258659	0,265124377	5,726590202	1,02469E-08	8,04857E-08
HHIPL2	283,3184285	1,514873968	0,200583485	7,55233645	4,27518E-14	9,49763E-13
GCLC	9091,727683	1,50896733	0,118926888	12,688193211	6,87536E-37	1,77986E-34
ADD2	452,0139446	1,505120952	0,161260598	9,333469989	1,0246E-20	5,9494E-19
NOTUM	377,3172022	1,503196617	0,163790277	9,177569282	4,40932E-20	2,41721E-18
UGT1A4	2,838762364	1,501194976	0,239882276	6,258048739	3,89824E-10	4,08887E-09
NKX2-2	8,673130458	1,500466362	0,295535306	5,077113733	3,83212E-07	2,15987E-06
SLC16A12	118,2040678	1,489492007	0,160473684	9,281845908	1,66567E-20	9,52347E-19

FGG	13463,03102	1,489263743	0,245334694	6,070334841	1,27644E-09	1,19796E-08
PGD	18393,85347	1,488359044	0,081471764	18,26840327	1,4771E-74	3,05907E-71
CALCA	1341,741103	1,48577667	0,332967846	4,462222664	8,11139E-06	3,46523E-05
MDGA1	727,2532881	1,482131059	0,124419541	11,912365555	1,02041E-32	1,94075E-30
TMC2	10,70564707	1,48205253	0,137977902	10,741231055	6,51695E-27	6,94111E-25
PSG3	4,793700985	1,476419494	0,362470544	4,07321234	4,63691E-05	0,000166752
GLTPD2	56,27532699	1,476198815	0,184996059	7,979623055	1,46781E-15	4,209E-14
EPHA5	13,24248821	1,470115148	0,196725926	7,47291005	7,84405E-14	1,67283E-12
NAA11	8,314100276	1,459417173	0,471344966	3,096282504	0,001959636	0,004846822
GSTM2P1	427,709008	1,456273503	0,251223617	5,796722149	6,76236E-09	5,52338E-08
GPR6	0,537520399	1,454418951	0,316796986	4,591012588	4,41101E-06	1,99458E-05
PKD1L2	99,0012667	1,45259563	0,149939758	9,687861612	3,39564E-22	2,3015E-20
ABCB6	2616,837522	1,451398813	0,074874053	19,384536661	1,0424E-83	3,88584E-80
HRG	14,14693784	1,450942097	0,207660315	6,98709379	2,80639E-12	4,51322E-11
NKAIN2	248,8055289	1,449348815	0,208216505	6,960777735	3,38399E-12	5,37259E-11
ALDH1A1	10425,48581	1,447187091	0,129653255	11,161980441	6,25817E-29	7,815E-27
NOS1	246,5891902	1,44226725	0,225314249	6,401136439	1,54225E-10	1,74676E-09
EPHX1	15729,12575	1,439570315	0,091451286	15,741389551	7,86964E-56	4,58382E-53
OR7A5	27,14159837	1,423750347	0,270257887	5,268117666	1,3783E-07	8,53298E-07
TCL1B	3,015149871	1,423416439	0,23685451	6,009665752	1,85906E-09	1,68618E-08
HCN4	48,32275774	1,418355146	0,171810294	8,255356009	1,5145E-16	5,15124E-15
CYP4F8	13,1180445	1,416872419	0,295171038	4,80017426	1,58528E-06	7,86478E-06
UGT1A5	1,448819932	1,412312925	0,231791274	6,093037493	1,10788E-09	1,05142E-08
PAX7	342,1166149	1,409038011	0,288175921	4,889506411	1,01089E-06	5,21914E-06
KPRP	34,43445155	1,406120291	0,339933886	4,136452263	3,52717E-05	0,000130267
HRASLS2	82,89225667	1,401866069	0,158240299	8,859096459	8,06656E-19	3,71241E-17
CDX2	53,88432915	1,400502064	0,2470298	5,669364851	1,43328E-08	1,08996E-07
PLA2G10	441,525306	1,400040693	0,180554542	7,754115061	8,89616E-15	2,21975E-13
UGDH	7267,038582	1,391833248	0,083224198	16,723900096	8,77806E-63	6,81726E-60
SCN9A	910,0504622	1,390849609	0,127181627	10,935931855	7,76046E-28	8,92884E-26
KCNV1	25,84395841	1,384341297	0,205281517	6,743623658	1,54484E-11	2,14883E-10
TMED6	125,037754	1,383761033	0,157551072	8,782936322	1,59261E-18	7,13575E-17

LRP1B	77,24989491	1,383024851	0,217922444	6,346408503	2,204E-10	2,42076E-09
ATP4A	29,38456857	1,379002755	0,270683685	5,094517438	3,49631E-07	1,98198E-06
TRIM16	3083,439197	1,372186642	0,081800571	16,774780071	3,73234E-63	3,06382E-60
DNAJB3	15,08618931	1,371379158	0,240223284	5,708768663	1,13796E-08	8,85617E-08
C11orf92	1142,793878	1,36803605	0,142525065	9,598564626	8,10654E-22	5,35808E-20
MAGEB6	6,606857388	1,366956582	0,289853063	4,716032902	2,40488E-06	1,14876E-05
C9orf4	12,63285025	1,357923833	0,251010838	5,409821523	6,30876E-08	4,17426E-07
SLC16A4	1172,319176	1,355085431	0,137943223	9,823501331	8,91907E-23	6,27331E-21
GSR	5755,195755	1,354977271	0,084491682	16,036812688	7,06905E-58	4,392E-55
RGS6	123,9176094	1,354640178	0,180714125	7,496039263	6,57751E-14	1,41896E-12
ORM1	545,8037423	1,354112686	0,191584119	7,067979802	1,57205E-12	2,6735E-11
C12orf39	13,01511182	1,353948491	0,238169864	5,684801885	1,30964E-08	1,00662E-07
NEFM	99,48779602	1,350826642	0,23949147	5,640395637	1,6966E-08	1,27358E-07
FZD10	695,3785448	1,346447842	0,172869476	7,788811976	6,76422E-15	1,73423E-13
FBXW10	21,8444946	1,346240654	0,135748496	9,917168131	3,50561E-23	2,63472E-21
RIMBP2	146,6891107	1,333421881	0,193405534	6,894435008	5,40793E-12	8,32357E-11
RAB6B	1726,694696	1,330459845	0,121574788	10,943550602	7,13497E-28	8,26017E-26
DPP10	121,5555226	1,328535407	0,196397402	6,764526367	1,33746E-11	1,88855E-10
ODC1	17122,35915	1,32730889	0,114613027	11,580785558	5,15713E-31	8,2157E-29
SLC7A2	3892,425631	1,319580392	0,155962455	8,460884972	2,65356E-17	1,01145E-15
OLFM1	1745,116312	1,313645364	0,155788142	8,432255135	3,39067E-17	1,26905E-15
POPDC3	189,6352652	1,309465839	0,18565299	7,053297858	1,74726E-12	2,93134E-11
SLC2A12	963,9293338	1,300922569	0,121337782	10,721496222	8,06874E-27	8,35518E-25
GABRA2	13,00578222	1,298866324	0,239909431	5,413986102	6,1637E-08	4,08844E-07
ADH1C	1396,063105	1,297646675	0,218292663	5,944527218	2,77256E-09	2,43507E-08
RFX4	6,326105633	1,291494067	0,20944297	6,166328088	6,9894E-10	6,97779E-09
C11orf53	79,92271894	1,287481298	0,255801183	5,033132699	4,82529E-07	2,66012E-06
JAKMIP2	297,9429229	1,274894027	0,153226358	8,320331089	8,77179E-17	3,08486E-15
WNT5A	3757,31699	1,274183348	0,117593841	10,835459888	2,33788E-27	2,55347E-25
CT45A1	67,37702067	1,270953928	0,380950418	3,336271247	0,000849103	0,002286627

AGXT2L1	33,20092091	1,263994009	0,34477454	3,666146607	0,000246233	0,00075066
ESX1	1,785739171	1,263265484	0,375669038	3,362708548	0,000771818	0,00210044
RUNDC3A	150,3194774	1,262952655	0,153016787	8,25368691	1,53581E-16	5,20493E-15
FOXI1	19,6056838	1,261702991	0,233394414	5,405883423	6,44897E-08	4,25345E-07
TALDO1	13804,82297	1,256794528	0,059013233	21,29682551	1,2148E-100	1,13213E-96
PTPRD	651,7051101	1,256694471	0,131723174	9,540420495	1,42254E-21	9,2709E-20
PIR	964,4954038	1,254424239	0,085295642	14,706780005	5,8317E-49	2,41549E-46
CBX2	1412,378949	1,253051955	0,104301361	12,013764119	3,00833E-33	5,96514E-31
OR51E2	6,387629622	1,252104269	0,187320215	6,684298706	2,32033E-11	3,08698E-10
PAPPA2	36,51924905	1,250664705	0,196842705	6,353624874	2,10299E-10	2,31694E-09
CHODL	254,5137129	1,248623562	0,151223408	8,256814084	1,49612E-16	5,09803E-15
IGFBP1	25,45884842	1,242081761	0,214870008	5,780619517	7,4426E-09	6,05247E-08
SCGN	7,734579024	1,237650221	0,306505319	4,037940429	5,39225E-05	0,000190968
tAKR	5,06580479	1,234531518	0,224266723	5,504746753	3,69699E-08	2,58762E-07
PPP2R2C	1274,016857	1,234360983	0,175039512	7,051899141	1,76492E-12	2,95565E-11
KLK6	577,1803729	1,233180728	0,231493531	5,327063445	9,98133E-08	6,35172E-07
B4GALNT1	464,9377741	1,231158532	0,143047814	8,606622466	7,52448E-18	3,10973E-16
DGKG	391,9549991	1,226699137	0,115116154	10,656185933	1,63152E-26	1,60899E-24
TAAR1	1,074607111	1,22207663	0,332025528	3,680670696	0,000232621	0,00071454
GPR50	64,24368545	1,221354783	0,308152039	3,963481107	7,38647E-05	0,000253922
LOC441177	5,141220775	1,218775419	0,285554878	4,268095247	1,97149E-05	7,74592E-05
GSTM1	1467,391512	1,218609366	0,242564179	5,02386366	5,06422E-07	2,77848E-06
SSX1	9,632214833	1,216899242	0,358774006	3,391826674	0,000694283	0,001910638
NDRG4	1490,210659	1,214534293	0,139514749	8,705418632	3,16405E-18	1,38438E-16
BCAN	148,1330742	1,214339456	0,160484581	7,566704849	3,8281E-14	8,60341E-13
NRCAM	1595,251326	1,210800752	0,157040083	7,710138238	1,25682E-14	3,04397E-13
NEIL3	244,5137733	1,208903416	0,102011316	11,850679552	2,13453E-32	3,93916E-30
CSAG3	35,09957316	1,208043964	0,482480048	2,503821595	0,012285997	0,024652675
FREM2	1272,331086	1,207651081	0,149496061	8,078146482	6,57586E-16	2,02591E-14
LCE2D	0,602474327	1,207515566	0,440213204	2,743024418	0,006087616	0,013294326
KCNA4	9,481136892	1,206576184	0,179714417	6,713853042	1,89551E-11	2,57136E-10

DMRT2	416,1366602	1,205800825	0,199635637	6,040007913	1,54107E-09	1,42763E-08
HEPACAM2	81,35629859	1,204151773	0,234261554	5,140202267	2,74443E-07	1,58556E-06
CCL26	73,67016642	1,203293811	0,157008973	7,66385378	1,80435E-14	4,24638E-13
MOV10L1	42,58115082	1,201666432	0,132147604	9,093365275	9,602E-20	4,99921E-18
MAP1B	4293,997763	1,197973102	0,116537965	10,27968099	8,70151E-25	7,37216E-23
RASD1	3077,422596	1,196885951	0,145684181	8,215620539	2,1107E-16	7,0378E-15
C1orf180	1,365459818	1,194168506	0,202992387	5,882824106	4,03324E-09	3,43582E-08
GNAT3	1,25239399	1,187011651	0,362550596	3,274057924	0,001060149	0,002793344
XKR7	3,851688056	1,18656965	0,225617594	5,259207078	1,44678E-07	8,88644E-07
CALB2	445,1777893	1,186132625	0,19249366	6,161930859	7,18632E-10	7,15523E-09
CA5A	2,404065754	1,185460266	0,262630482	4,513795409	6,36776E-06	2,78678E-05
ABCC1	10521,10616	1,181966105	0,094548257	12,50119406	7,35383E-36	1,73504E-33
IL1RAPL2	23,140971	1,179391855	0,177360197	6,649698597	2,93694E-11	3,82274E-10
TGM6	2,954797303	1,176395885	0,326276045	3,605523313	0,000311525	0,000929934
LCE3E	20,85541532	1,171478791	0,355161251	3,29844201	0,00097223	0,002586183
SYBU	1078,023594	1,162808175	0,127073411	9,150680407	5,65761E-20	3,03023E-18
GDA	528,0561458	1,160560177	0,209258904	5,546049193	2,92197E-08	2,09069E-07
TESC	2342,114322	1,155913804	0,187801517	6,154975866	7,50889E-10	7,44855E-09
MOBP	22,54364431	1,154004713	0,166242601	6,941690656	3,87435E-12	6,08887E-11
BTBD11	736,6755287	1,153288057	0,137655704	8,378062233	5,38062E-17	1,95496E-15
COL25A1	194,8121177	1,151895457	0,231583246	4,97400168	6,55847E-07	3,5067E-06
KRT83	25,61050994	1,141905478	0,192693239	5,926027719	3,1035E-09	2,69781E-08
CNTN5	86,25598267	1,138825784	0,186500307	6,106294418	1,01971E-09	9,7719E-09
GALNT13	230,8862994	1,138161197	0,164607728	6,914384943	4,69898E-12	7,31087E-11
STC2	1814,405117	1,137875522	0,113041659	10,06598389	7,81024E-24	6,16844E-22
NRSN1	3,346980626	1,136595865	0,285656492	3,978890364	6,92377E-05	0,000239384
GUCA2B	2,965684613	1,131922393	0,346208274	3,269483945	0,001077439	0,002833292
EID3	116,3012794	1,12764092	0,094542442	11,92735133	8,52427E-33	1,63798E-30
GSTA2	171,7669708	1,127219067	0,196425802	5,738650706	9,54338E-09	7,55795E-08
PTPRN	126,5766313	1,125236145	0,167100861	6,733874022	1,65204E-11	2,27923E-10
MIAT	1881,321748	1,12451199	0,1306929	8,604231709	7,68296E-18	3,1682E-16
GLI2	642,3031137	1,123380286	0,120722392	9,30548396	1,33383E-20	7,64965E-19

				4		
LRP8	1289,158936	1,11961405	0,0947117	11,8212856	3,03012E-32	5,48333E-30
CEL	915,4935798	1,117941575	0,196468482	5,690182779	1,26903E-08	9,79037E-08

Supplementary Table 2. Clinical characteristics of patients included in the SU2C cohort (n=153).

Clinical characteristic	***n (%)***
Histology	
Non-Squamous	117 (76.5)
Squamous	32 (20.9)
NA	4 (2.6)
Stage at diagnosis	
I	14 (9.2)
II	10 (6.5)
III	35 (22.9)
IV	82 (53.6)
NA	12 (7.8)
Sex	
Male	66 (43.1)
Female	85 (55.6)
NA	2 (1.3)
Line of treatment	
1	48 (31.4)
2	64 (41.8)
3	23 (15.0)
≥ 4	14 (9.2)
NA	4 (2.6)
Smoking	
Never	23 (15.0)
Former	105 (68.7)
Current	21 (13.7)
NA	4 (2.6)
Age *(years)*	
≥ 65	72 (47.1)
< 65	77 (50.3)
NA	4 (2.6)
Prior TKIs	
yes	12 (7.8)
no	137 (89.6)
NA	4 (2.6)
ICI treatment type	
Monotherapy	135 (88.2)
Combination	18 (11.8)

Abbreviations: NA, Not Available. TKIs, Tyrosine Kinase Inhibitors. ICI, immune checkpoint inhibitor.

Supplementary Table 3. Association between clinical characteristics and the KEAPness phenotype in the SU2C cohort (n=153).

Clinical characteristic	**Group (*n*)**		***χ²* test (*p*)**
	KEAPness-free	***KEAPness-dominant***	
Histology			
Non-Squamous	63	54	0.079
Squamous	11	21	
Stage at diagnosis			
Other	28	31	0.5
IV	44	38	
Sex			
Female	41	42	1.0
Male	33	33	
Line of treatment			
1-2	53	59	0.4
≥ 3	21	16	
Smoking			
Never/Former	66	62	0.36
Current	8	13	
Prior TKIs			
yes	12	0	< 0.01
no	62	75	
ICI treatment type			
Monotherapy	65	70	0.73
Combination	10	8	

Abbreviations: TKIs, Tyrosine Kinase Inhibitors. ICI, immune checkpoint inhibitor.

Supplementary Table 4. Clinical characteristics of patients and their association with the KEAPness phenotype in the OAK/POPLAR cohort (n=891, docetaxel and atezolizumab arms).

Clinical characteristic	***n (%)***		
Histology			
Non-Squamous	629 (70.6)		
Squamous	262 (29.4)		
Sex			
Female	562 (63.1)		
Male	329 (36.9)		
Therapy			
Atezolizumab	439 (49.3)		
Docetaxel	452 (50.7)		
	Group (*n*)		***χ^2 test (p)***
	KEAPness-free	***KEAPness-dominant***	
Histology			
Non-Squamous	368	261	< 0.01
Squamous	50	212	
Sex			
Female	184	145	< 0.01
Male	234	328	