



Published in final edited form as:

*JCO Precis Oncol.* 2019 ; 3: .

## Harnessing Clinical Sequencing Data for Survival Stratification of Patients with Metastatic Lung Adenocarcinomas

Ronglai Shen<sup>1,\*</sup>, Axel Martin<sup>1</sup>, Ai Ni<sup>1</sup>, Matthew Hellmann<sup>2</sup>, Kathryn C. Arbour<sup>2</sup>, Emmet Jordan<sup>3</sup>, Arshi Arora<sup>1</sup>, Ryan Ptashkin<sup>4</sup>, Ahmet Zehir<sup>4</sup>, Mark G. Kris<sup>2</sup>, Charles M. Rudin<sup>2</sup>, Michael F. Berger<sup>4,5</sup>, David B. Solit<sup>3,5,6</sup>, Venkatraman E. Seshan<sup>1</sup>, Maria Arcila<sup>4</sup>, Marc Ladanyi<sup>4,6</sup>, and Gregory J. Riely<sup>2,\*</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics

<sup>2</sup>Thoracic Oncology Service, Division of Solid Tumor Oncology, Department of Medicine

<sup>3</sup>Department of Medicine

<sup>4</sup>Department of Pathology

<sup>5</sup>Marie-Jose and Henry R. Kravis Center for Molecular Oncology

<sup>6</sup>Human Oncology and Pathogenesis Program

### Abstract

**Purpose**—Broad panel sequencing of tumors facilitates routine care of people with cancer as well as clinical trial matching for novel genome-directed therapies. We sought to extend the use of broad panel sequencing results to survival stratification and clinical outcome prediction.

**Patients and Methods**—Using sequencing results from a cohort of 1,054 patients with advanced lung adenocarcinomas, we developed OncoCast, a machine learning tool for survival risk stratification and biomarker identification.

**Results**—With OncoCast, we stratified this patient cohort into four risk groups based on tumor genomic profile. Patients whose tumors harbored a high-risk profile had a median survival of 7.3 months (95% CI 5.5-10.9), compared to a low risk group with a median survival of 32.8 months (95% CI 26.3-38.5), with a hazard ratio of 4.6 ( $P < 2e-16$ ), far superior to any individual gene predictor or standard clinical characteristics. We found that co-mutations of both *STK11* and *KEAP1* are a strong determinant of unfavorable prognosis with currently available therapies. In patients with targetable oncogenes including *EGFR/ALK/ROS1* and received targeted therapies, the tumor genetic background further differentiated survival with mutations in *TP53* and *ARID1A* contributing to a higher risk score for shorter survival.

**Conclusion**—Mutational profile derived from broad-panel sequencing presents an effective genomic stratification for patient survival in advanced lung adenocarcinoma. OncoCast is available as a public resource that facilitates the incorporation of mutational data to predict individual patient prognosis and compare risk characteristics of patient populations.

**Corresponding author:** Ronglai Shen (shenr@mskcc.org) and Gregory J. Riely (rielyg@mskcc.org).

\*Co-corresponding author Memorial Sloan Kettering Cancer Center, New York, New York.

## Keywords

Metastatic lung adenocarcinoma; prognostic model; mutational profiling

---

## Introduction

With the growth of genomic testing-driven precision medicine programs as well as recent FDA clearance of next-generation sequencing platforms for clinical use, there is a rapid growth in the availability of broad sequencing data for patients with cancer. Broad panel sequencing facilitates clinical trial matching of novel genome-directed therapies. Zehir et al.<sup>1</sup> delineated the molecular landscape of 10,000 metastatic cancer in a pretreated real-world cohort sequenced by the MSK-IMPACT platform, a hybridization capture-based NGS panel which can detect mutations and copy number alterations in 341 or more cancer-associated genes<sup>2</sup>. This study showed that 37% of patients harbored at least one therapeutically actionable alteration and 11% were matched to genome-directed clinical trials.

In patients with lung adenocarcinoma, tumor genotyping is now an essential step in routine clinical decision making. To determine treatment, patients with lung adenocarcinomas are currently categorized based on the presence of mutated driver oncogenes (e.g. *EGFR*, *ALK*, *ROS1*, and *BRAF*). A multi-institutional study characterized genetic aberrations across 10 genes in 733 tumor samples and identified an oncogenic driver in 64% of the patients<sup>3</sup>. Using a broad-panel based sequencing, Jordan et al.<sup>4</sup> reported from a single institution experience of 860 patients with metastatic lung adenocarcinoma that over 37% of patients received a matched therapy, with the use of matched therapy strongly influenced by the level of pre-existent clinical evidence that the mutation identified predicts the drug response.

While the focus of tumor genotyping has been on the identification of mutations that identify therapeutic targets, even within specific molecular subsets of patients with metastatic cancer there is considerable unexplained variability in clinical outcomes. Similarly, association between high mutational load and clinical benefit has been observed in patients treated with PD-1/PD-L1 inhibitors<sup>5</sup>. However, additional markers are needed to predict durable benefit and long-term survival among these patients. Prior attempts to evaluate the effects of co-mutations have had a relatively limited scope. Some studies have explored the effects of single co-occurring alterations on the outcome in patients with *EGFR*-mutant and *KRAS*-mutant lung adenocarcinoma and observed that the presence or absence of pairs of co-occurring events could be used to identify those patients with a poor prognosis most in need of novel therapeutic approaches<sup>3,6-8</sup>. However, a systematic approach is needed to further improve our understanding of survival and treatment outcome of patients.

Here, we developed OncoCast, a computational tool for survival stratification, and applied it to a large clinical series of metastatic lung adenocarcinomas to improve the understanding of heterogeneity in clinical outcome for these patients and the mutation and co-mutational patterns that underlie such heterogeneity. Analysis of this large clinical cohort provides real-world evidence for understanding survival outcome for patients undergoing current standard care in advanced lung adenocarcinomas. Such real-world evidence can supplement the information from randomized clinical trials (RCT) in that the results are more generalizable

to patients treated outside of RCT and the larger sample sizes of real-world data sets allow subset analysis that clinical trials are not powered for. The open-source computational pipeline we have developed can facilitate the application of statistical and machine learning approaches for clinico-genomics analysis of precision medicine data sets.

## Results

### Cohort characteristics

Consecutive patients with metastatic or recurrent lung adenocarcinomas for which MSK-IMPACT data were available were included. Electronic medical record was used to identify patient clinical factors as well as survival outcomes as previously described<sup>4</sup>. Overall survival (OS) was defined as the time from date of diagnosis of advanced disease (stage IV or recurrent cancer) until date of death or last follow-up. In this cohort, the majority (68%) of the tumors in our cohort were biopsied and sequenced within 30 days of diagnosis of metastatic disease. However, a fraction (21%) of the tumors were sampled and sequenced more than six months from the met recurrence date, with 16% more than a year and 8% more than 2 years, representing older samples used for sequencing analysis. Those patients with older samples taken at initial diagnosis of advanced lung cancer were “immortal” from their initial sampling time to the time of referral for MSK-IMPACT sequencing. This interval can be long for a small fraction of patients (8% with a delayed interval more than 2 years as mentioned earlier), introducing survival bias. Left-truncation was used to adjust for this bias. Details of left-truncation analysis are described in Supplementary Material. The most frequently mutated genes were *TP53* (55.1%), *KRAS* (30%), *EGFR* (29.4%), *STK11* (17.7%), and *KEAP1* (17.7%) (Supplementary Figure 1a). Among the frequently co-mutated gene pairs, *STK11* and *KEAP1* were co-mutated in 10% of the tumors, *KRAS* and *STK11* were co-mutated in 9% of the tumors (Supplementary Figure 1b).

### Prognostic relevance and clonality of cancer genes

To define the prognostic significance of MSK-IMPACT panel genes sequenced, we developed OncoCast, a machine learning tool for survival risk stratification and biomarker identification by implementing a lasso-penalized proportional hazards regression for deriving prediction rules for overall survival and feature selection (Supplementary Methods). OncoCast uses an ensemble learning strategy by repeatedly splitting the cohort into a training and test set multiple times that generate an ensemble of classifiers with varying selection of genes and gene combinations (Supplementary Figure 2).

The median number of prognostic genes selected was 19 (range 4-67), with a total of 169 cancer genes selected at least once by the ensemble learner. The relative prognostic importance of each gene was measured by how often it was selected (fraction of the models containing the gene) and its average regression coefficients that determined the corresponding weights for individual genes in the scoring rule. Figure 1a shows frequency of selection and average coefficient value for individual genes. *STK11* and *KEAP1* were observed to be highly prognostically relevant. They were followed in importance by *TP53*, *KRAS*, *SMARCA4*, and *EGFR*. An additional 21 genes were selected in at least 20% of the models (Figure 2a), including *ALK* fusion (58%), *ROS1* fusion (22%), and *ERBB2*

mutation (31%), further highlighting the power of our approach to aggregate effects from rare events, which otherwise would not be included in standard association analyses.

Several recent studies examining tumor heterogeneity have shown that clonal or “truncal” drug targets and/or tumor neo-antigens are more predictive of response to systemic targeted and immunotherapies than subclonal events<sup>9-11</sup>. We thus used the high depth of coverage afforded by our sequencing data (mean coverage of 758×) to determine the clonality of the gene alterations found to be associated with prognosis in the analysis outlined above. Clonality analysis revealed that all of the most prognostic alterations were predominantly clonal with average cancer cell fraction above 85% (Figure 1b).

### **OncoCast: an integrated prognostic scoring system based on NGS tumor profiling**

OncoCast aggregates prognostic effects across the sequenced cancer genes to derive a genomic risk score for each patient (scaled from 0 to 10), with a higher score indicating a greater likelihood of shorter survival. The distribution of the risk scores revealed a wide spread within the cohort (Figure 2a). To evaluate prognostic performance, we calculated the C-index that measures the concordance between the risk score and survival (Figure 2b). A 3-fold cross-validation was used for unbiased assessment. Clinical demographic factors (including age, sex, smoking) were weakly concordant with survival, with median C-indices ranging from 0.53 to 0.57. By contrast, the OncoCast risk score as determined by tumor genomic profiling demonstrated a significantly better concordance with a median C-index above 0.65. In advanced/metastatic lung cancers, studies have only shown weak to moderate strength of survival association for clinical factors<sup>12,13</sup>. The genomic classifier we developed here presents a significant improvement. In addition, the OncoCast classifier based on clinical-grade sequencing as part of routine care highlights the immediate application and strong practical utility. The prediction strength is comparable to established gene expression based genomic classifiers in early stage lung cancer from large multi-site study<sup>14</sup>. Inclusion of copy number alteration (CNA) data in the analysis revealed no discernible improvement in C-index upon incorporation of the CNA data (Figure 2b).

We categorized the patient cohort into low (0-25th percentile), low-intermediate (25-75th percentile), high-intermediate (75th-90th percentile) and high (90th percentile and above) risk groups informed by the multiple modes of the risk score distribution (Figure 3a). For the low risk group, median overall survival was 32.8 months (95% CI: 26.3-38.5). By contrast, for the high-risk group, the median survival was 7.3 months (95% CI: 5.5-10.9). There was a more than 6-unit difference in average risk score between the high and low risk groups. The OncoCast classification substantially outperformed all of the individual genes as a predictor of overall survival (Figure 3b). The hazard ratio for the OncoCast risk score classified high versus low risk group was 4.6 (95%CI 3.2-6.5), far superior to clinical factors or any individual gene. OncoCast risk score remained a highly significant predictor after adjusting for clinical variables and treatment types as potential confounding factors in a multivariate Cox regression model (Supplementary Table 1).

To confirm the validity of the OncoCast survival stratification we applied it to a separate data set of patients with mostly early stage NSCLC, obtained from the TCGA lung adenocarcinoma analysis. We saw a similar stratification of the risk groups (Supplementary

Figure 3). The high-risk group shows significantly worse overall survival with a hazard ratio of 1.9 (P=0.03) and highly enriched for concurrent *STK11* and *KEAP1* mutations.

### Prognostic mutation and co-mutation patterns

Overlaying the OncoCast risk score with the mutational landscape in an OncoPrint plot highlights that the tumors with the highest risk profile were enriched for co-mutation of *STK11* and *KEAP1* (Figure 4). *STK11* is a tumor suppressor gene that encodes for the serine/threonine kinase LKB1 that functions as a negative regulator of mammalian target of rapamycin (mTOR) signaling. Consistent with its functional role as a tumor suppressor, the majority of *STK11* mutations were truncating and frame-shift indels. To examine the status of the other allele in such cases, we performed allele-specific copy number analysis using the FACETS algorithm<sup>15</sup>. Strikingly, over 90% of *STK11* mutant tumors had evidence of bi-allelic inactivation through LOH of chromosome 19p (Supplementary Figure 4). *KEAP1* also resides in the 19p region with 89% of *KEAP1* mutant tumors demonstrating LOH. The majority of *KEAP1* mutations were missense and the mutant copy was frequently duplicated (present as copy-neutral LOH and uniparental gains) as reflected in the average total copy number.

We also explored whether other tumor characteristics such as tumor mutational signature or intra-tumor heterogeneity were prognostic and associated with OncoCast risk score. Previously defined smoking-associated and APOBEC signatures were the most prevalent mutational signatures in this cohort. The identification of a smoking signature was highly concordant with patient reported smoking status (Figure 4). An APOBEC signature was enriched in low risk groups and in patients who self-reported as never smokers. The median mutation burden in the overall dataset was 9.95 mutations/Mb (IQR 4.97-19.07). Tumor mutation burden (TMB) was associated with risk groups with a median of 16.58, 16.05, 9.95, and 5.76 mutations/Mb in the high, high-intermediate, low-intermediate, and low-risk groups, respectively (Kruskal-Wallis test P < 0.0001) (Supplementary Figure 5). However, mutation burden did not provide additional prognostic value beyond OncoCast risk score in a multivariable Cox model (Supplementary Table 2).

Clonal diversity was calculated for each tumor by summarizing the cancer cell fraction (CCF) for all somatic mutations using the Shannon index. A diversity index of zero represents homogeneity in which case all of the mutations detected in the tumor are clonal (CCF=100%). The median diversity index in the cohort was 1.19 (range 0.00-56.32). A trend of increased clonal heterogeneity was observed in high risk groups (Supplementary Figure 5). Again, clonal heterogeneity did not provide additional prognostic value on top of the OncoCast risk score.

### Genotypes associated with risk groups

To better understand the association between targetable cancer driver mutations and OncoCast risk score, we explored the distribution of genotypes within four major risk categories (Figure 5). The low risk group was highly enriched for *EGFR* mutants and tumors harboring oncogenic fusions in *ALK*, *RET* and *ROS1*. However, 30% of the low risk group tumors lacked targetable alterations in these 4 genes. The defining characteristic for the

patients in the low risk group without *EGFR*, *ALK*, *RET*, *ROS1* alterations was an absence of any of the poor prognostic gene alterations. In the high risk groups, common co-mutation patterns were observed. Strikingly, over 95% of the patients in the high risk group had tumors that harbor co-mutations of *KEAP1* and *STK11*. The top three major genotype categories in the high risk group were *KRAS-STK11-KEAP1*, *TP53-STK11-KEAP1*, and *KRAS-STK11-KEAP1-SMARCA4*. Furthermore, when compared to the TCGA resected lung adenocarcinoma cohort, *STK11* and *KEAP1* co-mutation was two-fold more common in the metastatic lung cohort (10.2% vs. 5.2%,  $P < 0.001$ ). This suggests that *STK11* and *KEAP1* co-mutation defines a cohort of resected lung adenocarcinoma patients at higher risk of disease progression and cancer specific mortality.

### Survival stratification in specific treatment subsets

We then sought to demonstrate the ability of the Oncocast technique to explore tumor genomic predictors outcomes in the subset of patients with mutated driver oncogenes treated with kinase inhibitors. The model, OncoCast-TR, was derived using the genomic profiles and survival outcomes from the start of therapy for the 387 patients who received targeted therapies (*EGFR*, *ALK*, *ROS*). Survival time was calculated from the start of treatment to death or last follow-up. Late entry was accounted for by using left-truncation analysis incorporating time to tumor sampling and sequencing. The risk score from the OncoCast-TR model clearly indicated two groups with distinct survival differences (Figure 6a,b). *TP53* is strongly associated with worse survival outcome (Figure 6c), with 96% of patients in the TR2 group harboring concurrent *TP53* mutations and 0% concurrent *TP53* mutation in the TR1 group. *ARID1A* also showed association with poor survival with 11/14 (78%) *ARID1A* mutations in TR2.

### Interactive tool for visualizing and exploring mutation pattern and survival.

To facilitate clinical translation and research use of the data, we created an interactive web application (Supplementary Figure 6) that allows for visualization of mutation patterns and individualized prediction of overall survival to be generated based on a user-defined mutational profile and clinical characteristics. There are two main functional modules: GeneView and PatientView. In GeneView, the user can input the genes of interest and interactively explore the prognostic effects in a dynamic volcano plot and with genotype pie-charts. In PatientView, an OncoCast risk score will be calculated from a user supplied mutational profile. Along with a patient's clinical profile, the application outputs the predicted survival probabilities along with a 95% confidence interval. While the tool was built using the largest metastatic lung adenocarcinoma cohort to date, it was designed to be dynamically updated as new data are incorporated into the model.

## Discussion

Tumors from patients with lung adenocarcinoma have a high frequency of actionable oncogenic drivers<sup>16,17</sup>. The introduction of targeted therapy has transformed the clinical care of patients with lung adenocarcinoma by incorporating tumor genotyping into therapeutic decision making. However, the prognostic value of the additional information from broad genomic profiling has not been explored. We developed OncoCast, a statistical

learning tool, for integrating broad genomic profiling data and clinical outcomes for survival stratification and the identification of associated biomarkers. Using an ensemble learning approach we demonstrated the prognostic utility of data from a large panel next generation sequencing assay interrogating 341 cancer-associated genes in 1,054 patients with metastatic lung adenocarcinoma treated with currently available therapies. We show that it is a practical approach to molecular stratification in metastatic lung adenocarcinoma and provides the ability to identify biomarkers that predict survival outcome.

Overlaying the OncoCast risk score with the tumor genomic landscape revealed novel biological and clinical insights. The major prognostic genes associated with poor survival included *STK11*, *KEAP1*, *TP53*, *KRAS*, and *SMARCA4*. Remarkably, co-mutations of both *STK11* and *KEAP1* defined an exceptionally high risk profile with a short median overall survival of 7.3 months, and a more than 6 unit increase in risk score compared to a low risk group, corresponding to a hazard ratio of 4.6. The low risk group had no mutations in these major poor prognostic genes. Some of the favorable prognostic genes may reflect the availability of effective targeted therapies in the case of *EGFR*, *ALK*, and *ROS1*. However, in patients with these targetable oncogenes, there is heterogeneity in the additional mutations their tumors harbor. Our model is novel in that it outputs a continuous risk score based on the specific genetic background of the patient's tumor, including additional genetic alterations beyond the driver oncogene. This provides finer granularity for understanding the heterogeneity in clinical outcome. For example, our model revealed that mutations in *TP53* and *ARID1A* define a high risk subgroup with shorter survival in the TKI-treated patient cohort.

Conversely, some of the unfavorable prognostic genes may reflect negative associations with treatment responses (e.g., the recent observation that *STK11* inactivation is associated with poor responses to immunotherapy<sup>18</sup>). Our results further highlight the significant risk imparted by concurrent mutations in genes such as *KEAP1*, *STK11*, and *SMARCA4*. *KEAP1* is a negative regulator of *NRF2*. Mutations in the *KEAP1* are associated with chemo-resistance and poor survival. Some studies have suggested that targeting *NRF2* may enhance chemotherapy sensitization<sup>19,20</sup>. mTOR is a kinase downstream of LKB1 (*STK11*) and mTOR inhibitors have been proposed as a potential therapeutic approach in *STK11*-mutant tumors<sup>21</sup>. *SMARCA4* was also identified as a major driver of poor prognosis factor in metastatic lung adenocarcinoma in our study. It is a core factor in SWI/SNF chromatin remodeling complexes that regulate genomic instability and DNA repair. *SMARCA4* was mutated in 9.4% of lung adenocarcinomas in the MSK-IMPACT cohort, with 42% of the mutations being truncating and residing in regions of LOH. It has recently been shown that *SMARCA4*-inactivating mutations increase sensitivity to Aurora kinase A inhibitor in non-small cell lung cancers<sup>22</sup>.

In a multivariate analysis (Supplementary Table 1) including treatment covariates, the OncoCast readout remained highly significant suggesting that the genomic profile provides more than simple readout of genotype-matched therapies. In addition, this model can be dynamically updated over time. With increasing sample sizes, we will be poised to identify outcome associations with rare alterations and increasingly better explain the heterogeneity in clinical outcomes for patients with advanced lung adenocarcinoma.

The OncoCast risk score defined here provides a valuable tool for more accurately describing prognosis of patients enrolled into clinical trials or included in real world data sets. A key component of any analysis of clinical research data is a clear description of the patient population explored. While conventional clinical factors such as age, sex, and performance status have long been used, our data indicate that we can significantly improve the description of patient populations by incorporating the genomic risk scores in our understanding to allow better comparisons of groups of patients enrolled in clinical trials or included in real world data sets.

With the growth of genomic testing-driven precision medicine programs as well as recent FDA clearance of next-generation sequencing platforms for clinical use, there will be a rapid growth in the availability of broad sequencing data for patients with a variety of cancers and growing integration with clinical data through institutional databases and electronic health records. Oncocast, the computational tool generated here will facilitate the incorporation of mutational data as a stratification factor in both prospective clinical trials and retrospective/real-world data collections to more precisely describe patient populations, allowing better generalization of the results from such research efforts.

## Materials and Methods

The OncoCast method is described in Supplementary Methods, and R package is available at <https://github.com/shenmskcc/OncoCast>. An interactive web interface was developed using R Shiny with two main functions: GeneView and PatientView. In GeneView, users can interactively explore gene importance and co-mutation pattern by risk groups. In PatientView, users can type in a patient's mutational profile and specify the clinical characteristics. The genomic risk score along with the predicted probability of survival at different time marks will be calculated and viewable in a dynamic plot. The Shiny app is available at <https://github.com/shenmskcc/LungIMPACT>.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgement

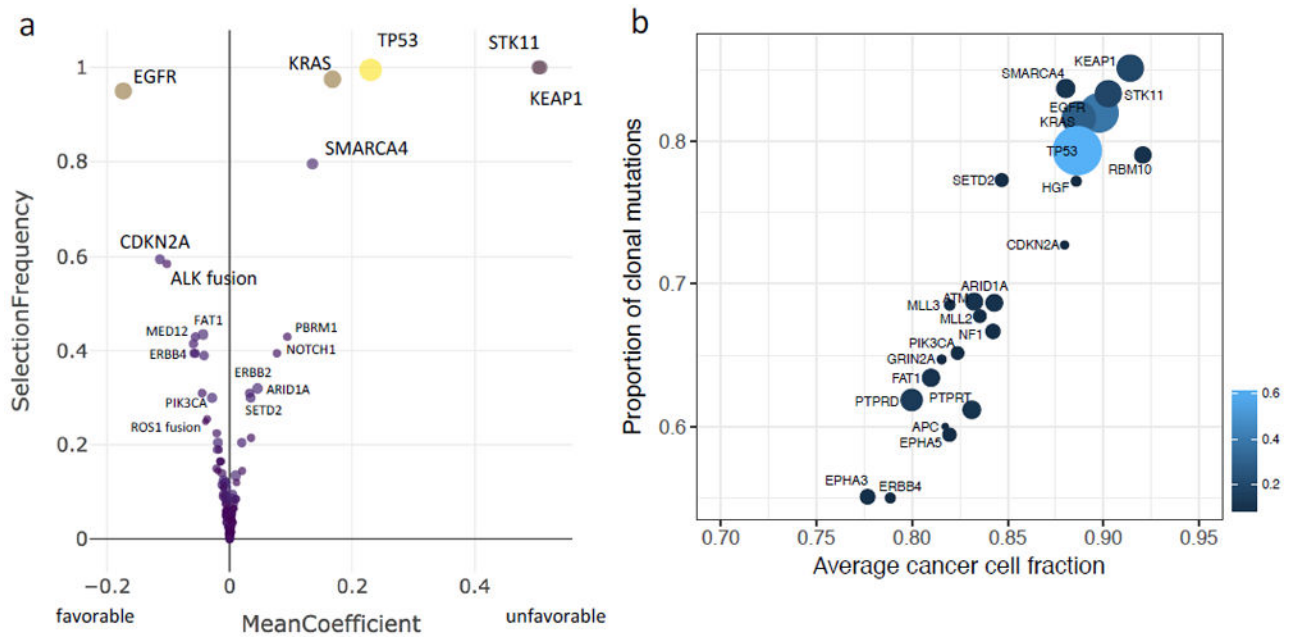
This research was supported in part by a lung P01 CA129243 and cancer center support grant P30 CA008748.

## References

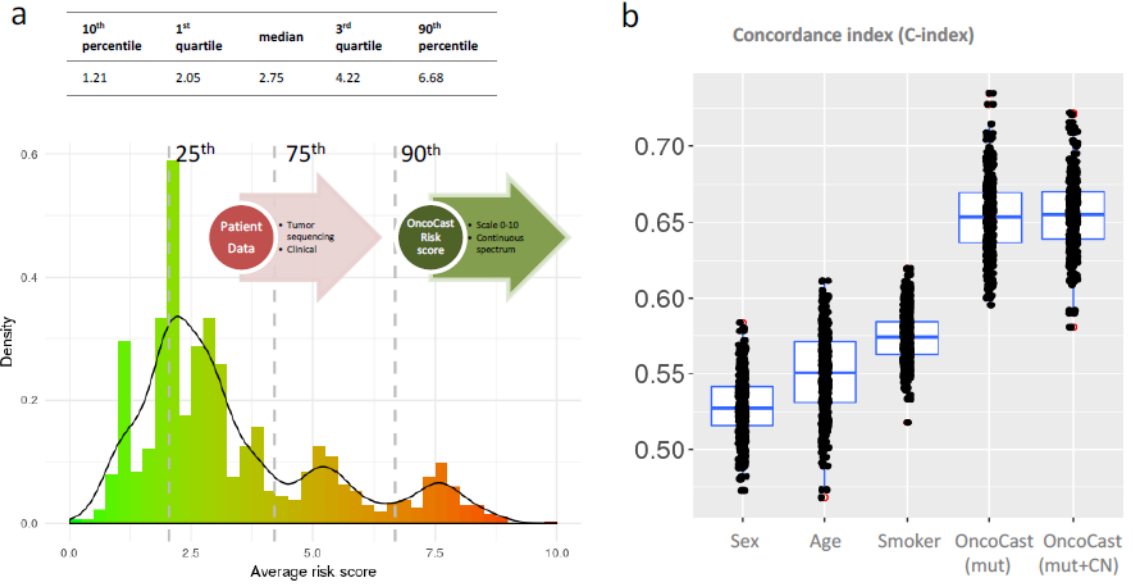
1. Zehir A, Benayed R, Shah RH, et al.: Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nature Medicine* 23:703–713, 2017
2. Cheng DT, Mitchell TN, Zehir A, et al.: Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *The Journal of molecular diagnostics* 17:251–264, 2015 [PubMed: 25801821]
3. Kris MG, Johnson BE, Berry LD, et al.: Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *Jama* 311:1998–2006, 2014 [PubMed: 24846037]



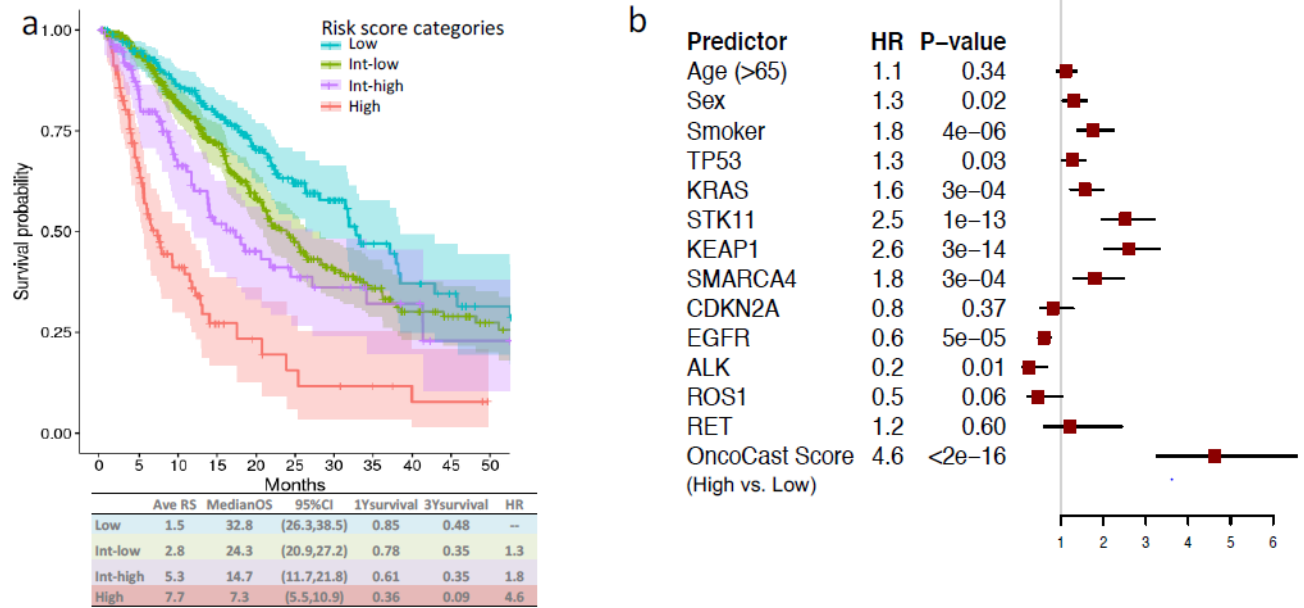
4. Jordan EJ, Kim HR, Arcila ME, et al.: Prospective comprehensive molecular characterization of lung adenocarcinomas for efficient patient matching to approved and emerging therapies. *Cancer Discovery* 7:596–609, 2017 [PubMed: 28336552]
5. Hellmann MD, Nathanson T, Rizvi H, et al.: Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. *Cancer Cell* 33:843–852 e4, 2018 [PubMed: 29657128]
6. Arbour KC, Jordan E, Kim HR, et al.: Effects of Co-occurring Genomic Alterations on Outcomes in Patients with KRAS-Mutant Non--Small Cell Lung Cancer. *Clinical Cancer Research*, 2017
7. Yu H, Suzawa K, Jordan EJ, et al.: Concurrent alterations in EGFR-mutant lung cancers associated with resistance to EGFR kinase inhibitors and characterization of MTOR as a mediator of resistance. *Clinical Cancer Research* [Epub ahead of print], 2018
8. Yu HA, Sima CS, Shen R, et al.: Prognostic impact of KRAS mutation subtypes in 677 patients with metastatic lung adenocarcinomas. *Journal of Thoracic Oncology* 10:431–437, 2015 [PubMed: 25415430]
9. de Bruin EC, McGranahan N, Mitter R, et al.: Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 346:251–256, 2014 [PubMed: 25301630]
10. Rizvi H, Sanchez-Vega F, La K, et al.: Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand (PD-L)-Ligand 1 Blockade in Patients With Non--Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. *Journal of Clinical Oncology*:JCO-2017, 2018
11. Rizvi NA, Hellmann MD, Snyder A, et al.: Mutational landscape determines sensitivity to PD-1 blockade in non--small cell lung cancer. *Science* 348:124–128, 2015 [PubMed: 25765070]
12. Azzoli CG, Temin S, Giaccone G: 2011 Focused Update of 2009 American Society of Clinical Oncology Clinical Practice Guideline Update on Chemotherapy for Stage IV Non-Small-Cell Lung Cancer. *J Oncol Pract* 8:63–6, 2012 [PubMed: 22548014]
13. Gerber DE, Dahlberg SE, Sandler AB, et al.: Baseline tumour measurements predict survival in advanced non-small cell lung cancer. *Br J Cancer* 109:1476–81, 2013 [PubMed: 23942074]
14. Director's Challenge Consortium for the Molecular Classification of Lung A, Shedden K, Taylor JM, et al.: Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 14:822–7, 2008 [PubMed: 18641660]
15. Shen R, Seshan VE: FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic acids research* 44:e131–e131, 2016 [PubMed: 27270079]
16. Clinical Lung Cancer Genome P, Network Genomic M: A genomics-based classification of human lung tumors. *Sci Transl Med* 5:209ra153, 2013
17. Pao W, Girard N: New driver mutations in non-small-cell lung cancer. *Lancet Oncol* 12:175–80, 2011 [PubMed: 21277552]
18. Skoulidis F, Goldberg ME, Greenawalt DM, et al.: STK11/LKB1 Mutations and PD-1 Inhibitor Resistance in KRAS-Mutant Lung Adenocarcinoma. *Cancer Discov* 8:822–835, 2018 [PubMed: 29773717]
19. Chian S, Thapa R, Chi Z, et al.: Luteolin inhibits the Nrf2 signaling pathway and tumor growth in vivo. *Biochemical and biophysical research communications* 447:602–608, 2014 [PubMed: 24747074]
20. Ren D, Villeneuve NF, Jiang T, et al.: Brusatol enhances the efficacy of chemotherapy by inhibiting the Nrf2-mediated defense mechanism. *Proceedings of the National Academy of Sciences* 108:1433–1438, 2011
21. Klümper H-J, Queiroz KCS, Spek CA, et al.: mTOR inhibitor treatment of pancreatic cancer in a patient with Peutz-Jeghers syndrome. *Journal of clinical oncology* 29:e150–e153, 2010 [PubMed: 21189378]
22. Tagal V, Wei S, Zhang W, et al.: SMARCA4-inactivating mutations increase sensitivity to Aurora kinase A inhibitor VX-680 in non-small cell lung cancers. *Nature communications* 8:14098–14098, 2017



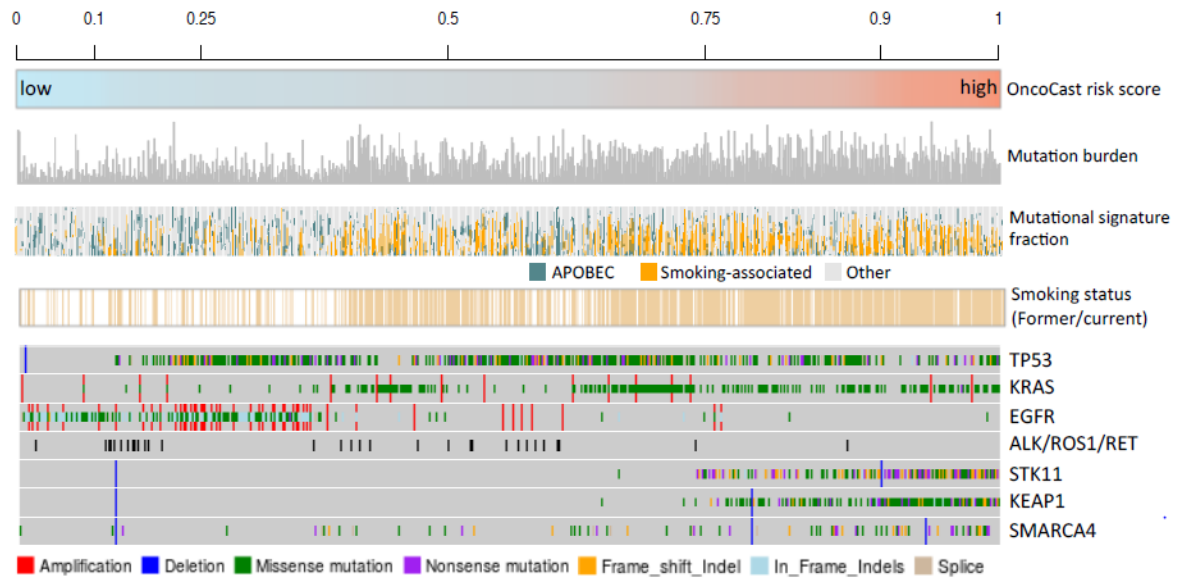
**Figure 1.** Prognostic relevance and clonality of cancer genes. a, Plot of frequency of selection in each model developed and regression coefficients for individual genes with signs indicating “favorable” versus “unfavorable” association with overall survival. b, Clonality analysis of the cancer gene alterations. Circle size is proportional to mutation frequency.



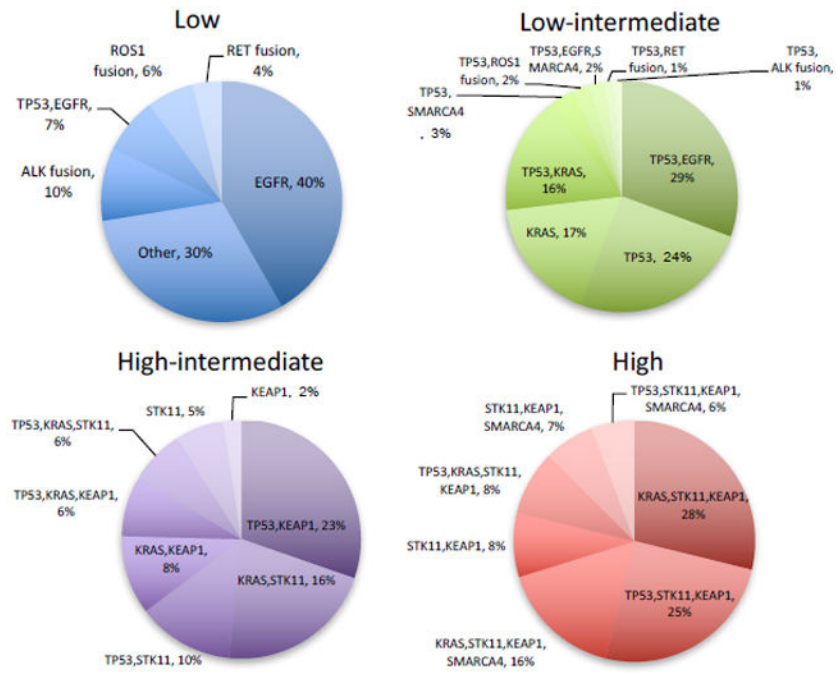
**Figure 2.** An integrated prognostic scoring system for metastatic lung adenocarcinomas. a, Histogram of the prognostic risk score computed using the OncoCast model in n=1,054 metastatic lung adenocarcinomas. The risk score is scaled from 0 to 10 with higher score indicates higher likelihood of shorter survival. Dashed lines indicate the percentile cutoffs used to stratify patients into four risk subgroups (Loa, Int-low, Int-high, High). b, Boxplots of concordance index for predicting overall survival using clinical demographic factors including age, sex, and smoking, and the OncoCast risk score.



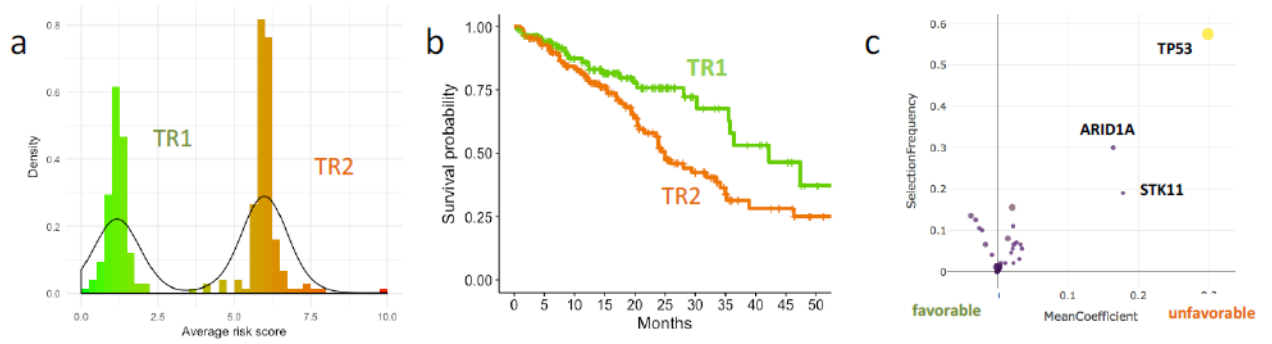
**Figure 3.** Outcome prediction based on risk score a, Kaplan-Meier plot of survival curves for the four risk subgroups. Colored areas represent 95% CI. The average risk score (Ave RS), median OS, 1yr and 3yr survival probabilities are reported for each group. b, Forest plot of hazard ratios for age, sex, smoking stats, and individual driver gene alterations compared to the OncoCast integrated scoring approach.



**Figure 4.** Overlaying OncoCast risk score with tumor mutation burden, mutational signature, smoking status, and mutation status for commonly mutated genes.



**Figure 5.** Genomic risk stratification. b, Distribution of common mutation and co-mutation patterns by each risk group.



**Figure 6.**

a, Risk score distribution from the OncoCast-TR model stratifies patients into two subsets. b, Overall survival probability for the two TR patient subsets. c, Gene importance plot for the OncoCast-TR model.