**SupplementARY figures and tables**

**Validation of the 12-gene predictive signature for Adjuvant Chemotherapy Response in Lung Cancer**

Authors**:** Yang Xie1,2,3\*#, Wei Lu4\*, Shidan Wang1\*, Ximing Tang4, Hao Tang1, Yunyun Zhou5, Cesar Moran4, Carmen Behrens6, Jack Roth7, Qinghua Zhou8, David Johnson9, Stephen Swisher7, John Heymach6, Vassiliki Papadimitrakopoulou6, Guanghua Xiao1,2,3, John Minna3,9,10, Ignacio Wistuba4#

1Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A.

2Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A.

3Simmons Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A.

4Department of Translational Molecular Pathology, University of Texas MD Anderson Cancer Center, Houston, Texas, U.S.A.

5Department of Data Science, University of Mississippi Medical Center, MS, U.S.A.

6Department of Thoracic/Head and Neck Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas, U.S.A.

7Department of Thoracic and Cardiovascular Surgery, University of Texas MD Anderson Cancer Center, Houston, Texas, U.S.A.

8Lung Cancer Institute, West China Hospital, Sichuan University, China.

9Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A.

10Hamon Center for Therapeutic Oncology Research, and Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A.

**#Corresponding Authors:**

Yang Xie, M.D., Ph.D., yang.xie@utsouthwestern.edu;

Ignacio Wistuba, M.D., iiwistuba@mdanderson.org.

**\*** Equally-contributing first authors.

**Supplementary Table S1**. Summary of Tables and Figures.

|  |
| --- |
| Summary of results |
| Patient cohorts | Table 1. Patient characteristics for the three cohorts.Suppl. Table S2. Propensity score matching methodsSuppl. Table S3. Propensity score matching results. |
|  Assay development | Fig. 1. Assay development and validation flowchart. |
| Suppl. Fig. S1. Determining the cutoff value in risk scores.Suppl. Table S4. Quality control measures for assay development |
| Prognostic performances | Supp. Fig. S2. Recurrence-free survival curves of high- and low-risk groups for 258 NSCLC patients (pv=0.0114). |
| Fig. 2 (A). Recurrence-free survival curves of high- and low-risk groups for 166 ADC patients (pv=1.25E-5). |
| Fig. 2 (B). Recurrence-free survival curves of high- and low-risk groups for 147 stage I ADC patients (pv=0.000161). |
| Supp. Fig. 2 (A). Time-dependent ROC curve for 166 ADC patients (AUC=0.75). |
| Supp. Fig. 2 (B). Time-dependent ROC curve for 147 stage I ADC patients (AUC=0.74). |
| Table 2. Multivariate analysis for recurrence-free survival prognostic performance in ADC patients. |
| Fig. 2(D). Recurrence-free survival curves in predicted high- and low-risk groups for 90 SCC patients (pv=0.841). |
| Supp. Table S5. Multivariate analysis for recurrence-free survival prognostic performance in stage I ADC patients. |
| Predictive performances | Fig. 3 (A & B) Recurrence-free survival curves in predicted ACT benefit and non-benefit groups. |
| Fig. 3 (C & D) Multivariate survival analysis in predicted ACT benefit and non-benefit groups. |
| Table 3. Multivariate analysis for interaction between predicted groups and ACT. |

**Supplementary Table S2**. Propensity score-matched coefficients for each variable using logistic regression.

|  |  |  |
| --- | --- | --- |
|  | Odds ratio | p value |
| Histology  Other vs. ADC | 1.8E-7 | 0.98 |
|  SCC vs. ADC | 0.76 | 0.46 |
| Smoking (Yes vs. No) | 0.79 | 0.66 |
| Age (Year) | 0.93 | 3.66E-04 |
| Gender (M vs. F) | 0.99 | 0.98 |
| Size (>=4cm vs. < 4cm) | 3.22 | 2.50E-03 |
| Stage (II vs. I) | 9.03 | 5.47E-09 |

**Supplementary Table S3. Propensity score matching results.** In the matched patient cohort, there are 138 patients without ACT treatment and 69 patients with ACT. The matching ratio is 2:1, and the variables (Histology, Smoking status, Gender, Tumor stage, Tumor Size and Age at the diagnosis) listed in Supplementary Table S2 were used to estimate the propensity scores.

|  |  |  |
| --- | --- | --- |
|  | Without ACT | ACT |
| All | 258 | 69 |
| Matched | 138 | 69 |
| Unmatched | 120 | 0 |

**Supplementary Table 4. T**he numbers of patients with and without ACT by stage after propensity score matching in ADC and SCC patient groups.

|  |  |  |
| --- | --- | --- |
| ADC | Without ACT | With ACT |
| Stage 1 | 77 | 16 |
| Stage 2 | 19 | 26 |

|  |  |  |
| --- | --- | --- |
| SCC | Without ACT | With ACT |
| Stage 1 | 23 | 5 |
| Stage 2 | 19 | 22 |

**Supplementary Table S5.** The dynamic range and correlation between expression levels measured in 30 pairs of FFPE and frozen samples of each gene measured by the nCounter gene expression assay.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gene Name | Avg Count | Min Count | Max Count | Pearson correlation coefficient |
| ATP8A1 | 695.02 | 24.12 | 4510.31 | 0.62 |
| AURKA | 219.8 | 33.17 | 666.85 | 0.51 |
| C1orf116 | 1279.98 | 45.42 | 7383.24 | 0.65 |
| COL4A3 | 63.56 | 1.12 | 457.18 | 0.52 |
| DOCK9 | 1094.98 | 237.71 | 3107.84 | 0.51 |
| HOPX | 8123.28 | 210.43 | 52256.67 | 0.73 |
| HSD17B6 | 213.1 | 8.32 | 2250.06 | 0.62 |
| IFT57 | 1459.41 | 261.62 | 8044.46 | 0.78 |
| MBIP | 1570.23 | 244.83 | 7832.9 | 0.77 |
| NKX2-1 | 1010.64 | 3.09 | 9669.5 | 0.76 |
| RRM2 | 1240.19 | 146.34 | 4103.03 | 0.68 |
| TTC37 | 995.05 | 261.62 | 1862.19 | 0.80 |

**Supplementary Table S6.** Multivariate analysis for validating the prognostic performance of the 12-gene signature measured from FFPE samples after adjusting for other patient characteristics in the 147 stage I ADC patients without ACT.

|  |  |  |
| --- | --- | --- |
|  | HR | p-value |
| Smoking status (Yes vs. No) | 2.06 (0.47, 8.99) | 0.34 |
| Age (Year) | 1.03 (0.98, 1.08) | 0.21 |
| Gender (M vs. F) | 3.08 (1.30, 7.28) | 0.011 |
| Size (>=4cm vs. < 4cm) | 1.92 (0.41, 9.03) | 0.41 |
| Group (High risk vs. Low risk) | 3.23 (1.26, 8.30) | 0.015 |

**Supplementary Figure S1**. Determining the cutoff value of pre-defined risk scores for predicted high-risk and low-risk groups using the assay development cohort. (A) Density plot of the predicted risk score in the assay development cohort (solid black line). The predicted risk scores were fitted with a mixture of two normal distributions using a model-based clustering algorithm. One normal distribution corresponds to the low-risk group (dashed green line), and the other corresponds to the high-risk group (dashed green line). In the figure, each dot represents a patient, and the axis is the predicted risk score for each patient in the assay development cohort. The green dots are the patients in the predicted low-risk group and the red dots are patients in the predicted high-risk group. The risk groups are defined using a model-based clustering method. (B) The predicted risk score of a patient vs. the probability that the patient is in the predicted high-risk group. The functional relationship was determined by the model-based clustering result. The patients whose probabilities of being in the high-risk group were greater than 0.5 are defined as the predicted high-risk group (presented as red dots), and the predicted low-risk group (green dots) otherwise. The corresponding risk score cutoff value is 1.035.



**Supplementary Figure S2**. (C) Time-dependent ROC curves in (A) ADC patients and (B) stage I ADC patients.



**Supplementary Figure S3**. The ROC curve for the predictive analysis. The ROC curve was estimated based on the R package tsm (Huang et. al. Biometrics 2012). The AUC for the predictive performance is 0.806 and the 95% confidence interval is (0.694, 0.865).

