**Supplemental tables legends**

**Supplemental Table 1.** List of TaqMan assays designed using Primer Express Software V3.0 (Thermo Scientific).

**Supplemental Table 2.** Ridge penalized Cox regression model coefficients obtained from the training set. Factors used to scale the risk score in a 0-1 range and the cut-off (median) used to categorize patients into high- and low-risk classes are reported.

**Supplemental Table 3.** Cox proportional hazards models in the stage I training and validation sets. Low- and high-risk classes were defined according to different cut-offs (median, 25th, 33rd, 66th and 75th percentiles derived from the training set). P-values were computed using Wald test.

**Supplemental Table 4.** Patient and tumor characteristics of the subgroup of 156 patients (Analyzed) of the Advanced Cancer Set, compared to the remaining cohort of 129 patients with stage II and III adenocarcinomas not included in the analysis. Wilcoxon, Fisher’s exact tests and log-rank test were used to compare distributions between included and not included patients.

**Supplemental Table 5.** List of TCGA patients used for the *in silico* analysis with clinical-pathological information and cluster annotation.

**Supplemental Table 6. A.** List of the most informative transcripts (column “SELECTED”, flagged as “YES”) and those significantly regulated according to the SAM analysis (column “q<0.05”, flagged “YES”). The d(i) is the observed relative difference for two-class problem. The geometric means (Geom mean) of intensities in class 1 and class 2, and Fold-change differences in the C1 samples versus the “Other” (C2, C3 and C4) are also shown. **B.** List of the most informative promoter regions (column “SELECTED”, flagged as “YES”) and of those significantly regulated according to the SAM analysis (column “q<0.05”, flagged “YES”). The d(i) is the observed relative difference for two-class problem. The geometric means (Geom mean) of intensities in class 1 and class 2, and Fold-change differences in the C1 samples versus the “Other” (C2, C3 and C4) are also shown. **C.** List of significantly regulated proteins according to the SAM analysis (q<0.05) among the 190 proteins in total. The d(i) is the observed relative difference for two-class problem. The geometric means (Geom mean) of intensities in class 1 and class 2, and Fold-change differences in the C1 samples versus the “Other” (C2, C3 and C4) are also shown.

**Supplemental Table 7.** Causes of death in Training and Validation Sets (stage I). HR for the 10-gene risk model is also reported considering all causes of death or deaths of tumor.

**Supplemental Table 8.** Mutation rate in C1-stage I, according to smoking status. P-value was computed using the exact Poisson test.

**Supplemental Table 9.** List of upstream modulators predicted to be activated/inhibited by IPA analysis. “Exp Log Ratio” indicates expression regulation for modulators that were originally included in the list of 2349 genes (C1-stage I *vs.* others-stage I; q<0.05). The type of modulator (Modulator Type, i.e., transcription regulator, ligand-dependent nuclear receptor, kinase, or enzyme), the activation z-score, and the modulators’ “target genes” (included in the list of 2349 genes) are also indicated.

**Supplemental Table 10.** Biofunctions analysis of the C1-stage I gene expression signature (N=2349. The enriched biofunctions were selected based on their significance p-values (P<0.05; Benjiamini-Hochberg correction).

**Supplemental Table 11.** List of upstream modulators predicted to be activated/inhibited by IPA analysis of the DCC dataset. “Exp Log Ratio” indicates expression regulation for modulators that were originally included in the list of 853 genes (C1-stage I *vs.* others-stage I; q<0.05). The type of modulator (Modulator Type, i.e., transcription regulator, ligand-dependent nuclear receptor, kinase, or enzyme), the activation z-score, and the modulators’ “target genes” (included in the list of 853 genes) are also indicated.

**Supplemental Table 12A.** TCGA patients with mutated KEAP1; the cBioPortal for Cancer Genomics - Mutation Assessor tool (42) (21) was also used to annotate the fraction of patients with KEAP1 mutation with medium/high predicted functional (inactivating) impact (Mutation assessor tool)(43). Total N, number of patients as in TCGA dataset. Mut (% over N), percentage of patients with KEAP1 mutations. cBIO N’, number of TCGA patitents present in cBioPortal database. cBIO – Mut Functional, percentage of patients with KEAP1 inactivating mutations predicted by cBioPortal. **B.** TCGA patients with KEAP1 copy-number alterations. GISTIC (44) scores, as reported in TCGA, were used to call for deletion (DEL; GISTIC= -2). **C.** TCGA patients with KEAP1 promoter methylation. The KEAP1 promoter was considered methylated when beta value was >0.3 as described in (45)(see also “Methods, Methylation analysis”).

**Supplemental Table 13.** Alterations (mutations, CNV and methylation) of NRF2/KEAP1 genes in “C1” patients. GISTIC (44) scores, as reported in TCGA, were used to call for amplification or deletion (AMP, GISTIC=2; DEL, GISTIC= -2). Promoter methylation was called if beta> 0.3 or unmethylation if beta ≤ 0.3, as described in (45)(see also “Methods, Methylation analysis”).

**Supplemental Table 14.** Patient and tumor characteristics of a subgroup of 50 patients (25 at low-risk and 25 at high-risk according to the 10-genes signature) from the stage I IEO cohort. Median, first and third quartiles were also reported for normalized Cq (Cqn) of NRF2 regulated genes (AHR, GCLC, NQO1, PRDX1). Wilcoxon and Fisher’s exact tests were used to compare distributions between low- and high-risk patients.

**Supplemental Figure Legends**

**Supplemental Figure 1. Optimization of the 10-gene signature screening protocol in FFPE samples and effect of long-term storage of FFPE samples on RNA quality.**

We profiled by RT-qPCR matched FFPE and FF samples from the same individuals (N=3) using our optimized protocol for FFPE 10-gene signature screening. **A.** Differences in the raw Cq of FF and FFPE matched samples without pre-amplification (10-gene signature plus the three reference genes). **B.** Differences in the raw Cq of FFPE samples with (FFPE preAMP) or without pre-amplification (FFPE). P-values were calculated using Student’s t-test. **C.** Correlation plot of normalized Cq (Cqn; using *TBP*, *HPRT1* and *GUSB*, as reference genes) of the 10-gene signature in matched FF and FFPE samples (without pre-amplification). **D.** Correlation plot of normalized Cq (Cqn; using *TBP*, *HPRT1* and *GUSB*, as reference genes) of the 10-gene signature in matched FF and FFPE samples (with pre-amplification). R, R-squared of bivariate linear fit. **E.** Themedians of Raw Cq are reported for the reference genes *GUSB, HPRT1, TBP* and *ESD* by year of sample collection. **F.** Themedians of Raw Cq are reported for the 10 genes in the signature by year of sample collection. Black lines connect the values at the boundaries of the time period considered. P-values for linear trend are reported.

**Supplemental Figure 2. Additional analyses on IEO and TCGA cohorts.**

**A.** Kaplan-Meier survival curves of all IEO patients with stage I disease stratified by set (training, validation). **B.** Kaplan-Meier survival curves of all IEO patients with stage I disease in the training set stratified using the 10-gene model. **C.** Patient distribution (blue, alive; red, dead) according to the 10-gene signature risk score. The current cut-off obtained from the training set used to categorize patients into high- and low-risk classes is reported (Supplemental Table 2-3). **D.** Gene expression regulation of the 10 genes in the four clusters found by hierarchical clustering analysis of the cohort of 468 lung cancer patients in the TCGA data set (see Figure 2A). Yellow bars indicate the 9 genes originally described to be upregulated in poor prognosis stage I lung adenocarcinoma patients (10). Blue bars indicate the *SCGB3A1* gene originally found to be downregulated in poor prognosis patients stage I lung adenocarcinoma patients (10). P-values were calculated by ANOVA analysis. **E.** Gene expression regulation of the 10 genes in stage I lung cancer patients of the TGCA cohort (N=247) in the 4 different clusters described in **D**. P-values were calculated by ANOVA analysis. **F**. Mutation frequency and enrichment in “C1” and “Other” clusters of the 18 most mutated genes in TCGA lung adenocarcinoma dataset (13).

**Supplemental Figure 3. Gene expression analysis of other lung cancer datasets and prognostic signatures.**

**A.** Hierarchical cluster analysis of the 10-gene signature in the DCC cohort of lung adenocarcinoma patients screened by Affymetrix. In yellow, increased expression; in blue, decreased expression. The main identified clusters are indicated by a color code (red, cluster C1; green, cluster C2; orange, cluster C3). Kaplan-Meier survival curves of all patients (left panel), or limited to stage I (central panel), or to stage II-IV lung cancer patients (right panel), using the ‘Cluster IDs’ as the grouping parameters. P-values were computed by using the log-rank test. **B.** Hierarchical cluster analysis of TCGA lung adenocarcinoma patients using another 11-gene signature (40). In yellow, increased expression; in blue, decreased expression. The main identified clusters are indicated by a color code (red, cluster C1; green, cluster C2; blue, cluster C3). Kaplan-Meier survival curves of all patients (left panel), or limited to stage I lung cancer patients (right panel), using the ‘Cluster IDs’ as the grouping parameters. P-values were computed by using the log-rank test. **C.** Hierarchical cluster analysis of TCGA lung adenocarcinoma patients relative to the CCP score (12, 41). In yellow, increased expression; in blue, decreased expression. The main identified clusters are indicated by a color code (red, cluster C1; blue, cluster C2). Kaplan-Meier survival curves for all patients (left panel), or limited to stage I lung cancer patients (right panel), using the ‘Cluster IDs’ as the grouping parameters. P-values were computed by using the log-rank test.

**Supplemental Figure 4. Molecular characteristics of different groups of TCGA lung adenocarcinoma patients.**

**A.** Number of differentially expressed and methylated genes, and differentially expressed proteins in Stage I-C1 patients compared to Stage I-Other or to Stage II-IV. **B.** Number of differentially expressed and methylated genes, and differentially expressed proteins in Stage II-IV patients compared to Stage I-Other ot to Stage I-C4. **C.** Number of patients included in the analysis.

**Supplemental Figure 5. IPA analysis of the NFE2L2 modulator.**

Gene network of NFE2L2 regulated genes present in the “C1-stage I” gene expression signature (2349 genes; q<0.05). In the bottom part of the diagram, the IPA-predicted activated upstream modulator, NFE2L2, is indicated in orange. Lines connect the modulator to direct targets, and colors indicate consistency with the predicted activity with the expression change observed in C1-stage I *vs.* other-stage I patients (i.e., Target Expression). Orange, consistent predicted activity on targets; yellow, inconsistent predicted activity on targets. Numbers indicate expression change (log2) of targets in C1-stage I patients *vs.* other-stage I patients.

**Supplemental Figure 6. KEAP1 and NFR2 target genes analyses.**

**A.** Gene network of KEAP1-regulated genes present in the “C1-stage I” gene expression signature identified in the DCC dataset (853 genes; q<0.05). The IPA-predicted inhibited upstream modulator, KEAP1, is shown in blue. Lines connect the modulator to direct targets, and orange color indicate consistency between the predicted activity and the expression change observed in DCC-C1-stage I (N=124) versus DCC-other-stage I (N=152) patients. Numbers indicate expression change (log2) of targets. **B.** Gene expression of NRF2 target genes in wt (WT) and mutated (Mut) KEAP1 patients of TCGA cohort. P-values for Wilcoxon test are reported. **C.** Normalized Cq (Cqn; using TBP, HPRT1 and GUSB, as reference genes) of the NRF2 target genes reported in Figure 3D, in a group of 25 high- and 25 low-risk (10-genes, continuous risk score) IEO lung adenocarcinoma stage I patients. P-values for Wilcoxon test are reported.

**supplemental TABLES**

**Supplemental Table 1.TaqMan Assays**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene**  **Symbol** | **Ref-Seq** | **TaqMan**  **Assay Type** | **TaqMan**  **Assay** | **Forward**  **Primer Sequence** | **Reverse**  **Primer Sequence** | **FAM**  **probe sequence** | **Amplicon**  **Length** | **Position**  **(Custom)** |
| **E2F1** | NM\_005225 | **Custom** |  | 5’GGGTCCCTGAGCTGTTCTTCT | 5’TCTGTCTCCCTCCCTCACTTTC | 5’CCCATACTGAAGGAACT | 87 | chr20:32263789-32263875 |
| **E2F4** | NM\_001950 | Made to order | Hs00608100\_g1 |  |  |  | 61 |  |
| **HOXB7** | NM\_004502 | **Custom** |  | 5’AAAACCTACCACTCGCGTGTTC | 5’GGACGGGAAGCAAGAAGCA | 5’CAAGCGCCTGGCTG | 58 | chr17:46,684,667-46,684,724 |
| **HSPG2** | NM\_005529 | Inventoried | Hs01078536\_m1 |  |  |  | 81 |  |
| **MCM6** | NM\_005915 | Made to order | Hs00962409\_m1 |  |  |  | 91 |  |
| **NUDCD1** | NM\_032869 | **Custom** |  | 5’GGCAACCACACTCCAGCAA | 5’GGACATAGCCTAAAGCATTGAAAGT | 5’TGGGAGCACATCGC | 72 | chr8:110,257,542-110,257,613 |
| **RRM2** | NM\_001034 | **Custom** |  | 5’TGAAAGGCTTTGTCTTGCATTG | 5’CAGCTAATGAGAGACAGAATCCTAAAAC | 5’AGGTACAGGCGGAAGT | 77 |  |
| **SCGB3A1** | NM\_052863 | Inventoried | Hs00369360\_g1 |  |  |  | 58 |  |
| **SERPINB5** | NM\_002639 | **Custom** |  | 5’GCTGTCCCATCTGGTCATTTG | 5’TCCCTGTGCAAGTCAGCTAGAA | 5’TTGGCACTAGACTGGTGG | 69 | chr18:61171844-61171912 |
| **SF3B1** | NM\_012433 | Made to order | Hs00961636\_m1 |  |  |  | 77 |  |
| **GUSB** | NM\_000181 | Inventoried | Hs99999908\_m1 |  |  |  | 81 |  |
| **HPRT1** | NM\_000194 | Inventoried | Hs02800695\_m1 |  |  |  | 82 |  |
| **TBP** | NM\_003194 | **Custom** |  | 5’ACCGCGCAGCGTGACT | 5’AGCGCTGCCCAGATAGCA | 5’TGAGTTGCTCATACCGTG | 54 | chr6:170,881,549-170,881,602 |
| **ESD** | NM\_001984 | Inventiored | Hs00382667\_m1 |  |  |  | 82 |  |

**Supplemental Table 2. 10-gene model**

|  |  |
| --- | --- |
| **Gene Symbol** | **Value** |
| **E2F1** | -0.0145293 |
| **E2F4** | -0.0742783 |
| **HOXB7** | -0.0481070 |
| **HSPG2** | -0.0447603 |
| **MCM6** | 0.0144221 |
| **NUDC1** | 0.0344735 |
| **RRM2** | -0.0913007 |
| **SCGB3A1** | -0.0465072 |
| **SERPINB5** | -0.0476968 |
| **SF3B1** | 0.0218761 |
|  |  |
| **Scale factors** |  |
| **minimum** | -8.3263546 |
| **maximum** | -6.0477828 |
|  |  |
| **Cut-off value** | 0.3709966 |

**Supplemental Table 3.** **Cut-offs**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **10-gene risk high (*vs.* low)** | | | | | | | |
| **Cut-off** | **Training Set** | | | | **Validation Set** | | | |
| **N** | **N deaths** | **Univariate** | | **N** | **N deaths** | **Univariate** | |
| **HR (95% CI)** | **Pa** | **HR (95% CI)** | **Pa** |
| Median:  0.3709965762 | 95 | 15 | 3.93 (1.31-11.85) | 0.02 | 83 | 12 | 4.04 (1.14-14.31) | 0.03 |
| 25th percentile:  0.2529352141 | 142 | 18 | 6.39 (0.85-47.88) | 0.07 | 131 | 13 | 1.61 (0.36-7.12) | 0.53 |
| 33rd percentile:  0.2906502102 | 126 | 15 | 1.94 (0.65-5.85) | 0.24 | 118 | 13 | 2.59 (0.59-11.48) | 0.21 |
| 66th percentile:  0.4243740328 | 64 | 9 | 1.80 (0.73-4.42) | 0.20 | 53 | 7 | 1.91 (0.69-5.27) | 0.21 |
| 75th percentile:  0.4540547643 | 48 | 9 | 2.81 (1.14-6.90) | 0.02 | 41 | 5 | 1.46 (0.50-4.29) | 0.49 |

**a***Wald test p-value*

**Supplemental Table 4.** **Clinical and pathological characteristics of Stage II-III cohort**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Stage II-III cohort** | | | |
| **All** | **Analyzed** | **Not analyzed** | **P** |
| **N** | 285 | 156 | 129 |  |
| **Age at surgery [years]** |  |  |  |  |
| median (Q1;Q3) | 65 (57;71) | 65 (60;70) | 64 (55;72) | 0.41 a |
| min-max | 34-83 | 39-81 | 34-83 |
| **Gender** (female) | 88 (30.9%) | 43 (27.6%) | 45 (34.9%) | 0.20 b |
| **Smoking history** |  |  |  |  |
| Current/ Former | 233 (81.7%) | 128 (82.1%) | 105 (81.4%) | 0.29 b |
| Never | 33 (11.6%) | 15 (9.6%) | 18 (14.0%) |
| Unknown | 19 (6.7%) | 13 (8.3%) | 6 (4.7%) |
| **Stage** |  |  |  |  |
| II | 157 (55.1%) | 105 (67.3%) | 52 (40.3%) | <0.0001 b |
| IIA | 92 (32.3%) | 64 (41.0%) | 28 (21.7%) |
| IIB | 65 (22.8%) | 41 (26.3%) | 24 (18.6%) |
| III | 128 (44.9%) | 51 (32.7%) | 77 (59.7%) |
| IIIA | 116 (40.7%) | 47 (30.1%) | 69 (53.5%) |
| IIIB | 12 (4.2%) | 4 (2.6%) | 8 (6.2%) |
| **Follow-up** |  |  |  | 0.76 c |
| **Deaths**  (within 3 years) | 86 (30.2%) | 51 (32.7%) | 35 (27.1%) |
| **Survivors**  (follow-up) |  |  |  |
| <1 yr | 28 (9.8%) | 10 (6.4%) | 18 (14.0%) |
| 1-2 yrs | 15 (5.3%) | 5 (3.2%) | 10 (7.8%) |
| 2-3 yrs | 18 (6.3%) | 5 (3.2%) | 13 (10.1%) |
| >3 yrs | 138 (48.4%) | 85 (54.5%) | 53 (41.1%) |
| Total pt-yrs | 623 | 357 | 266 |
| **Survival %**  (95% CI) |  |  |  |
| 1 yr | 89.7 (86.1-93.3) | 89.3 (84.3-94.2) | 90.4 (85.2-95.6) |
| 2 yrs | 78.4 (73.3-83.4) | 78.8 (72.1-85.4) | 77.8 (70.0-85.6) |
| 3 yrs | 65.1 (59.0-71.1) | 64.0 (56.1-72.0) | 66.6 (57.3-75.8) |

*a Wilcoxon’s test p-value.b Fisher’s exact test p-value.c Log-rank test for survival p-value*

**Supplemental Table 5. TCGA Clinical Info: see .xls file**

**Supplemental Table 6A. TCGA Gene Expression: see .xls file**

**Supplemental Table 6B. TCGA Methylation: see .xls file**

**Supplemental Table 6C. TCGA Protein Expression: see .xls file**

**Supplemental Table 7.** **Causes of death in Training and Validation Sets**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Training+Validation Set** | | | **Training Set** | | | **Validation Set** | | |
| **(stage I)** | | | **(stage I)** | | | **(stage I)** | | |
|  | **Total** | **Low-risk** | **High-risk** | **Total** | **Low-risk** | **High-risk** | **Total** | **Low-risk** | **High-risk** |
| **N** | **N** | **N** | **N** | **N** | **N** | **N** | **N** | **N** |
| **Deaths**  (within 3 years) | 34 | 7 | 27 | 19 | 4 | 15 | 15 | 3 | 12 |
| **Cause of Death** |  |  |  |  |  |  |  |  |  |
| Tumor | 21 | 4 | 17 | 13 | 2 | 11 | 8 | 2 | 6 |
| Other | 2 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| Unknown | 11 | 2 | 9 | 5 | 2 | 3 | 6 | 0 | 6 |
| **HR (95% CI)** |  |  |  |  |  |  |  |  |  |
| All causes | 3.98 (1.73-9.14) P=0.001a | | | 3.93 (1.31-11.85) P=0.02a | | | 4.04 (1.14-14.31) P=0.03a | | |
| Tumor | 4.39 (1.48-13.06) P=0.01a | | | 5.80 (1.29-26.16) P=0.02a | | | 3.04 (0.61-15.08) P=0.17a | | |

*a Wald test p-value*

**Supplemental Table 8.** **Stage I – C1 smoking status**

|  |  |  |  |
| --- | --- | --- | --- |
| **Stage I - C1 with available mutation data** | **N** | **Mutation**  **Rate** | **P a** |
| ALL | 65 | 2.3 |  |
| Smoker (current/former) | 54 | 2.4 | 0.66 |
| Never-smoker | 11 | 2.1 |

*a Exact Poisson test p-value*

**Supplemental Table 9. IPA Analysis “Upstream Modulator” TCGA: see xls file**

**Supplemental Table 10. IPA Analysis “Biofunctions” TCGA: see xls file**

**Supplemental Table 11. IPA Analysis “Upstream Modulator” DCC dataset: see xls file**

**Supplemental Table 12A. KEAP1 Mutation in TCGA patients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Total**  **N** | **Mut**  **(% over N)** | **cBIO**  **N’** | **cBIO - Mut functional**  **(% over N’)** |
| C1 | 139 | 37 (27%) | 121 | 14 (12%) |
| other | 295 | 38 (13%) | 273 | 12 (4.4%) |
| II-IV-C1 | 74 | 21 (28%) | 63 | 7 (11%) |
| II-IV-other | 122 | 15 (12%) | 113 | 5 (4.4%) |
| I-C1 | 65 | 16 (25%) | 58 | 7 (12%) |
| I-other | 173 | 23 (13%) | 160 | 7 (4.4%) |

**Supplemental Table 12B. KEAP1 copy-number variants in TCGA patients**

|  |  |  |
| --- | --- | --- |
| **Class** | **Total**  **N** | **DEL**  **N (% over N)** |
| C1 | 144 | 2 (1.4%) |
| other | 309 | 4 (1.3%) |
| II-IV-C1 | 77 | 2 (2.6%) |
| II-IV-other | 130 | 3 (2.3%) |
| I-C1 | 67 | 0 (0%) |
| I-other | 179 | 1 (0.6%) |

**Supplemental Table 12C. KEAP1 promoter methylation in TCGA patients**

|  |  |  |
| --- | --- | --- |
| **Class** | **Total**  **N** | **Methyl**  **N (% over N)** |
| C1 | 121 | 120 (99%) |
| other | 267 | 267 (100%) |
| II-IV-C1 | 63 | 62 (99%) |
| II-IV-other | 113 | 113 (100%) |
| I-C1 | 58 | 58 (100%) |
| I-other | 154 | 154 (100%) |

**Supplemental Table 13. Mutation, Copy Number Variation and Methylation in C1-patients (TCGA) of KEAP1 and NRF2 genes.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type** | **Total N** | **NFR2**  **N (% over N)** | **KEAP1**  **N (% over N)** | **NRF2 and KEAP1**  **N (% over N)** |
| **Mutation** | 141 | 5 (3.5%) | 38 (27%) | 0 |
| **CNV a** |  |  |  |  |
| AMP | 147 | 5 (3.4%) | 0 | **\*** |
| DEL | 147 | 0 | 2 (1.4%) |
| **Methylation b** |  |  |  |  |
| methyl | 121 | 0 | 120 (99%) | 0 |
| unmethyl | 121 | 121 (100%) | 1 (1%) | 1 |

*aamplification (AMP) if GISTIC = 2; deletion (DEL) if GISTIC = -2. bmethylation if beta value > 0.3; unmethylation if beta value ≤ 0.3; \* N=1 patient with NRF2 amplification and KEAP1 deletion (GISTIC)*

**Supplemental Table 14. Patient and tumor characteristics of a subgroup of 50 patients from the stage I IEO cohort.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | |  | **Stage I cohort** | | |
| **Low risk** | **High risk** | **P** |
| **N** | |  | 25 | 25 |  |
| **10-gene risk score** | |  |  |  |  |
| median (Q1;Q3) | |  | 0.24 (0.21-0.26) | 0.51 (0.45-0.52) | <0.0001 a |
| min-max | |  | 0.15-0.29 | 0.43-0.68 |
| **Age at surgery [years]** | |  |  |  |  |
| median (Q1;Q3) | |  | 65 (59;73) | 60 (54;67) | 0.12 a |
| min-max | |  | 50-84 | 42-74 |
| **Gender (female)** | |  | 6 (24.0%) | 9 (36.0%) | 0.54 b |
| **Smoking history** | |  |  |  |  |
| Current/ Former | |  | 20 (80.0%) | 20 (80.0%) | 1.00 b |
| Never | |  | 1 (4.0%) | 2 (8.0%) |
| Unknown | |  | 4 (16.0%) | 3 (12.0%) |
| **Stage I** | |  |  |  |  |
| IA | |  | 13 (52.0%) | 14 (56.0%) | 1.00 b |
| IB | |  | 12 (48.0%) | 11 (44.0%) |
| **AHR** | Cqnc  median (Q1;Q3) | | 21.9 (21.3;22.6) | 21.6 (21.2;22.3) | 0.44 a |
| **GCLC** | 25.2 (24.5;25.9) | 24.3 (23.4;25.1) | 0.0478 a |
| **NQO1** | 24.4 (22.4;24.9) | 22.6 (21.8;24.3) | 0.0283 a |
| **PRDX1** | 20.6 (20.2;21.1) | 20.4(19.8;20.8) | 0.07 a |

**a***Wilcoxon’s test p-value.bFisher’s exact test p-value. cCqn=normalized Cq*