**Supplementary Appendix**

**Part I: Computer tomography image acquisition, interpretation, and feature extraction**

1. **Pre-therapy images**

Pre-therapy computer tomography (CT) was acquired two weeks before the commencing of EGFR TKI therapy and chemotherapy (for comparison cohort only). Conventional modern CT equipment and techniques were used in this study. Four independent hospitals were joined in this multicenter retrospective study. The details of CT acquisition are described as following:

**Training set N = 117:** Both non-enhanced and contrast-enhanced chest CT images were acquired onPhilips Brilliance 40 and Siemens Defintion AS. The acquisitionparameters of Philips Brilliance 40 are as follows: tube voltage =120 kV, tube current = 200 mA, rotation time = 0.75s, detectorcollimation = 32 × 1.25 mm, field of view (FOV) = 300 × 300 mm, pixel matrix = 512 × 512, Filter sharp (C) for CT reconstruction, while the Siemens Defination AS is with the following acquisition parameters: tube voltage = 120 kV, tube current = 130 mA, rotation time = 0.5s, detector collimation = 64 × 0.625 mm, FOV = 300 × 300 mm, image matrix = 512 × 512, kernel B31f medium sharp + for CT reconstruction.

**Independent validation set N = 101:** Both chest non-enhanced and contrast-enhanced CT were performed on every patient using one of the two multi-detector row CT (MDCT) systems (GE Lightspeed Ultra 8, GE Healthcare, Hino, Japan or 64-slice LightSpeed VCT, GE Medical systems, Milwaukee, Wisconsin), with the following acquisition parameters: 120 kV; 160 mAs; 0.5- or 0.4-second rotation time; detector collimation: 8 × 2.5 mm or 64 × 0.625mm; field of view, 350 × 350 mm; matrix, 512 × 512. After routine non-enhanced CT, contrast-enhanced CT was performed after 25 s delay following intravenous administration of 85 mL of iodinated contrast material (Ultravist 370, Bayer Schering Pharma, Berlin, Germany) at a rate of 2.5–3.0 mL/s with a pump injector (Ulrich CT Plus 150, Ulrich Medical, Ulm, Germany). All CT images were reconstructed with the standard kernel. Moreover, all the CT images were retrieved from the picture archiving and communication system (PACS) (Carestream, Canada).

**Independent validation set N = 96:** CT scans were obtained with SIEMENS SOMATOM Definition Flash scanners (Munich, Germany). The following parameters were used to obtain HRCT images: collimator with 64 × 0.6 mm, section thickness of 1 mm, reorganization interval of 0.66 mm, and tube voltage of 120 kV. Tube current is calculated according to an individual weight, height, and body mass index. The tube current was 220 mAs for body mass index ≤ 25 kg/m2 and 330 mAs for body mass index > 25 kg/m2.

**Chemotherapy set N = 56:** This cohort was consisted of two parts. There are 37 patients enrolled from the Department of Radiology, Shanghai Pulmonary Hospital, Shanghai, China, and 19 patients enrolled from the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. Contrast-enhanced CT images from Shanghai were acquired according to the standard above. Patients enrolled from the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College underwent contrast-enhanced abdominal and pelvic CT using one of the two 64-detector row spiral CT systems (Discovery 750 HD, GE Medical systems, USA or Volume Ultra, GE Medical systems, USA). CT scan was performed after 60s delay following intravenous injection of 100ml Iopromide (Uitravist-300; Bayer Schering Pharma, Berlin Germany) at a rate of 3ml/s for enhancement. The scanning parameters were as follows: 120 kv; 160 mAs; 0.6-second rotation time; matrix, 512×512. Portal venous-phase CT images were reconstructed with slice thickness of 1.25mm.

In addition, distant metastases of each enrolled stage IV EGFR-mutant NSCLC patient were diagnosed by the clinician in each participating institution, respectively. Emission Computed Tomography was used to detect bone metastases, and liver and brain metastases were mainly diagnosed by MRI and PET/CT images. For patients who may have multiple distant metastases, the sites of suspected metastases were examined accordingly. Distant metastases of the enrolled patients from each participating institution were recorded normatively for further statistics.

1. **CT image interpretation**

CT scans were interpreted qualitatively and quantitatively by radiologists at each institution; standardized reporting forms were used to record lymph node status, and common sites of distant metastasis (i.e., bones, liver, and brain). The primary tumors were manually segmented by one radiologist, and the results were reviewed in a comprehensive assessment. These reviewers provided with educational materials on image interpretation that specifically described how to recognize the boundary of fuzzy tumors. However, formal demonstration of expertise was not mandated.

As the stage IV NSCLC patients are often accompanied with atelectasis, necrosis, pleural effusion, adhesion of mediastinum or pulmonary wall, and multiple metastases. These situations cause difficulty to manual segmentation. So how to locate the boundary of primary tumor plays a great influence on feature extraction. Therefore, we use contrast enhanced CT images to locate the primary tumor. Besides, a guideline should be provided to the radiologists when they meet the tumors difficult to segment. We analyzed all the images and the corresponding clinical diagnosis collected from each hospital. For cases which are difficult to identify, especially for the tumors combined with pleural effusion or other noncancerous tissues, the radiologists should control his/her manual delineation only including the unambiguous tumor region on CT images. The manual segmentation should be strictly controlled, and eliminating noncancerous region and fuzzy boundary. According to this principle the noncancerous tissues are strictly excluded when manual segmentation. There were 1032 phenotypic features of NSCLC tumor from regions of interest (ROI) calculated in this study.

1. **Phenotypic features decoding**

The phenotypic feature set used in this paper comprised of 1032 descriptors, which contained 440 features from Aerts1 and 592 features from Song2. All the features were grouped by: 3D, texture, Gabor, and wavelet features that covered one-, two- and three-dimensional features. Formulae and additional information can be found in these two papers. Finally, 12 differently expressed phenotypic descriptors were obtained from the feature set, as described in the Supplementary Part II: “2. Construction of a twelve-feature-based signature by Lasso Cox regression”.

1. **Assessment of reproducibility of feature extraction by different observers**

Two radiologists with more than 10 years of experience in chest CT interpretation were chiefly responsible for the accuracy evaluation of manual tumor segmentation. The reproducibility of phenotypic features noted per reviewer was evaluated as described previously. A set of 50 stage IV *EGFR*-mutant NSCLC patients were randomly selected for the analysis. The two observers were double-blinded to the segmentation.

We use Kruskal-Wallis H test in the reproducibility assessment experiment to assess the differences of phenotypic features generated between the two radiologists. The inter-class correlation coefficient (ICC) was used to determine the inter-observer agreement of phenotypic features extracted from the ROI segmented manually by each radiologist, and an ICC greater than 0.75 was considered a mark of excellent reliability. The ICC is not the correlation between a predictor variable and dependent variable, but it reflects the extent to which members of the same group or class tend to act alike3. It is the proportion of the total variability in the measured factor that is due to the variability between individuals. In this study, the ICC ranged from 0.872 to 0.935 for the radiologists.

**Part II: X-tile, Lasso Cox, and Nomogram**

1. **X-tile**

X-tile4 was used for determining the optimum cut-off for differentiating rapid- and slow-progression subgroups of EGFR TKI therapy in stage IV *EGFR*-mutant NSCLC patients using a single and intuitive method. The optimum cut-off score for every signature acquired using X-tile plots was based on the patients’ progression-free survival (PFS). The X-tile program can automatically select the optimum cut-point according to the highest χ² value (minimum p value) defined by Kaplan-Meier survival analysis and log-rank test. X-tile software (version 3.6.1, Yale University School of Medicine, New Haven, USA) was used to analysis the twelve signatures we used in this paper, as described in the Supplementary Table S1 and Supplementary Figure S1.

1. **Construction of a feature-based signature by Lasso Cox regression**

LASSO (least absolute shrinkage and selection operator) method was proposed by Tibshirani and colleagues in the context of regression analysis. As a recently proposed shrinkage method, it has been widely used in regression analyses for large models or high-dimensional potential prognostic factors5. The L1 penalty was used to shrink some regression coefficients to exactly zero. LASSO is a useful tool when the sample size-to-variables ratio is too low. The method has been extended and broadly applied to the Cox proportional hazard regression model for survival analysis with high-dimensional data because it can perform automatic feature selection in a manner that results in signatures with generally good prognostic performance. The method carries a strong prognostic value and reduces correlations among discovered variables to prevent overfitting; which is represented by the penalty parameter (λ). The larger the value of λ, the fewer the number of predictors selected. We used LASSO Cox regression model to select the most useful prognostic markers of all the features by the training set, as described in the main text Section of “Methods: Phenotypic Feature Selection, Signature Building, and Validation”. A multi-feature-based signature was constructed to risk stratification for stage IV *EGFR*-mutant NSCLC patients in the training set. We used the R software version 3.2.3 and the “glmnet” package (R Foundation for Statistical Computing, Vienna, Austria) to perform LASSO Cox regression model analysis.

The Cox proportional hazards regression analysis is the most popular approach to model covariate information for survival times, is not suitable for high-dimensional matrix data when the sample size to variables ratio less than 10:16,7. Instead, LASSO was introduced to eliminate this limitation8,9. The value of the partial likelihood deviance tends to be smaller with fewer variables (Supplementary Figure S2). We chose λ via the 1-SE (standard error) criteria, i.e, the optimal value of λ indicates the smallest value of partial likelihood deviance. Herein, we have plotted the partial likelihood deviance versus log(λ), where λ is the tuning parameter. A λ value of 0.67 with a log(λ) = -0.4 was chosen in this study. The optimal tuning parameter was accompanied by twelve non-zero coefficients. Twelve features and the coefficients (as described below) were selected in LASSO Cox regression model. According to the cut-off of the signature by X-tile, patients lower than the cut-off were classified into slow-progression group, and patients higher than the cut-off were classified into rapid-progression group.

***Signature*** = (2.231 × 10^(-8) × value of “Contrast of Co-occurrence on LL in the 0° direction”) + (7.590 × 10^(-4) × value of “Maximum-Probability of Co-occurrence on LL in the 0° direction”) + (3.034 × 10^(-5) × value of “Maximum-Probability of Co-occurrence on LL in the 45° direction”) + (5.353 × 10^(-5) × value of “Maximum-Probability of Co-occurrence on HL in the 0° direction”) + (1.010 × 10^(-4) × value of “Maximum-Probability of Co-occurrence on HL in the 45° direction”) + (4.482 × 10^(-6) × value of “Long-Run-High-Gray-Level Emphasis of Run Length on HL”) + (0.023 × value of “Entropy of GPTR in the 225° direction by two pixel steps”) + (0.116 × value of “Entropy of GPTR in the 45° direction by four pixel steps”) – (1.324 × 10^(-7) × value of “Variance of GMTR in the 90° direction by four pixel steps”) + (0.115 × value of “Entropy of GPTR in the 135° direction by four pixel steps”) – (9.290 × 10^(-9) × value of “Variance of GMTR in the 225° direction by five pixel steps”) + (2.300 × 10^(-6) × value of “Maximum diameter of tumor”).

Wavelet transform was performed on the CT image. The LL, LH, HL and HH image represent the four components after wavelet transform, respectively. In this study, we extracted all the features from two-dimensional and three-dimensional images. The computational formulas of these features we presented in detail as following. The source codes of the key features extraction have been attached in the paper.

**Contrast of Co-occurrence:**



**Maximum-Probability of Co-occurrence:**



Where the gray-level co-occurrence matrix is defined as, a matrix to describe the gray level distribution by a distance of  pixels in direction  of an image with the size of, where the th element represents the number of times the combination of intensity levels occurs in two pixels in the image. The other definitions are described as follows:

 is the co-occurrence matrix by the  and  ,

 is the number of discrete intensity levels in the image,

 is the mean of ,

 is the mean of ,

 is the mean of ,

 is the standard deviation of ,

 is the standard deviation of ,

**Long-Run-High-Gray-Level Emphasis of Run Length**



Run-length is a metrics to quantify gray level runs in an image. Since the consecutive pixels that have the same gray level value in one direction could be measured, the gray level run is defined as the length in number of pixels. In a gray level run length matrix, the th element describes the number of times the  gray level  appears consecutively in the direction specified by . The other definitions are described as follow:

 is the th point in the given run-length matrix  for a direction ,

 is the number of discrete intensity values in the image,

 is the number of different run lengths,

 is the number of voxels in the image.

Gabor filter is a linear filter used for edge detection, which is usually used in the field of face recognition. It could select valuable image information in different directions and different scales. We used eight directions (= 0°, 45°, 90°, 135°, 180°, 225°, 270° and 325°) and five scales (scale = 1, 2, 3, 4, and 5 steps) to extract Gabor features. Mean, variance, and entropy were used to construct the Gabor feature group. Gabor magnitude texture representation (GMTR) and Gabor phase-based texture representation (GPTR) are captured using the convolution between multi-scale and multi-directional Gabor wavelet function, and the feature we selected on the image in every direction and every scale. Here we use eight directions and five scales.  indicates the length of histogram of gray-level of Gabor image,  denotes the number of gray level ,  indicates the sum of image pixels and the  presents the intensity of  on the Gabor image.

**Variance of Gabor**



**Entropy of Gabor**



**Maximum diameter of tumor**

The maximum diameter of tumor is measured as the largest pairwise Euclidean distance, between voxels on the surface of tumor volume in three-dimensional.

1. **Nomogram and calibration**

Calibration is useful for assessing whether actual outcomes approximate predicted outcomes for each nomogram10. The x-axis represents the prediction PFS calculated by nomogram, while the y-axis represents the actual PFS for our patients. The 45-degree line represents the performance of an ideal nomogram; i.e., representing a predicted outcome that perfectly corresponds with actual outcome. In a well calibrated model, points on the graph are close to the 45-degree line.

The decision curve analysis was performed for comparing the net benefits at different threshold probabilities given by nomograms with and without the signature. The performance was evaluated by the net reclassification improvement (NRI) and integrated discrimination improvement (IDI). The integrated discrimination improvement (IDI) is commonly used to compare two risk prediction models. It summarizes the extent a new model increases risk in events and decreases risk in non-events. IDI averages risks across events and non-events11. NRI summarizes the net changes of allocation in clinical meaningful risk categories for events and nonevents when extending an existing prediction model with a novel marker12. It can be computed by summing up the proportions of correctly upward classified events (correctly qualifying for treatment) and downward classified nonevents (correctly abstaining from treatment), subtracted by the proportions of incorrectly downward classified events (incorrectly abstaining from treatment) and upward classified nonevents (unnecessarily qualifying for treatment). As described, the relative simplicity of NRI has undoubtedly contributed to its popularity.

**Part III: The R software packages used for statistical analysis**

R software version 3.2.3 was used in this study. The packages we used in this study are described as following.

1. LASSO Cox regression was performed using the “glmnet” package, we use the “lambda.1se” as the method to select key features, we used “nlambda = 100” and “maxit = 10000” to calculate the signature.
2. Multivariate Cox regression models, nomograms, and calibration plots were constructed with the “rms” package. “cph” was used to perform the Cox regression, and “B = 1000” for model validation.
3. C-index calculation was performed using the “Hmisc” package.
4. Independent validations were performed within the “rms” package while the“survival ROC” package was used for the time-dependent ROC curve analysis.
5. The net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were performed with the “dca” and “survIDINRI” packages, with “npert = 300”, and “IDI.INF.OUT()” was used to perform this analysis.

**Appendix Reference:**

1. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.
2. Song J, Liu Z, Zhong W, Huang Y, Ma Z, Dong D, et al. Non-small cell lung cancer: quantitative phenotypic analysis of CT images as a potential marker of prognosis. Sci Rep. Nature Publishing Group; 2016;6:38282..
3. Maxwell S, Delaney H. Designing experiments and analyzing data: A model comparison perspective. Briefings Funct. genomics proteomics. 2004.
4. Qu N, Shi R, Luo T, Wang Y, Li D. Prognostic significance and optimal cutoff of age in medullary thyroid cancer. Oncotarget. 2015;7.
5. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol. 2016;34:2157–64.
6. Simon R, Altman DG. Statistical Aspects of Prognostic Factor Studies in Oncology. Brit J Cancer 1994;69(6):979-85
7. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate Data Analysis. Int. J. Pharm. 1998..
8. Tibshirani R. Regression Selection and Shrinkage via the Lasso. J. R. Stat. Soc. B. 1996. page 267–88..
9. Zhang HH, Lu W. Adaptive Lasso for Cox’s proportional hazards model. Biometrika. 2007;94:691–703.
10. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: More than meets the eye. Lancet Oncol. 2015. page e173–80.
11. Chipman J, Braun D. Simpson’s paradox in the integrated discrimination improvement. Stat Med. 2016.
12. Leening MJG, Cook NR. Net reclassification improvement: a link between statistics and clinical practice. Eur J Epidemiol 2013;28(1):21-3.

**Supplementary Table S1.** Multivariable Cox regression analysis of various clinicopathologic characteristics and the twelve-feature-based signature with progression-free survival.

|  |  |  |
| --- | --- | --- |
| **Variables** | **Training Set (N = 117)** | |
| **p value** | **HR (95% CI)** |
| **Gender (Male vs. Female)** | 0.322 | 1.027 (0.607, 1.736) |
| **Smoke (Yes vs. No)** | 0.002 | 2.731 (1.382, 4.423) |
| **Age(≤65 vs. >65)** | 0.352 | 0.816 (0.532, 1.252) |
| **Treatment(first line vs. second line)** | 0.452 | 0.721 (0.264, 1.378) |
| **T stage T1 as reference: T2** | 0.231 | 1.634 (0.732, 3.650) |
| **T3** | 0.012 | 2.326 (1.760, 5653) |
| **T4** | 0.138 | 1.971 (0.804, 4.834) |
| **Tumor location (Right vs. Other)** | 0.322 | 0.954 (0.875, 1.132) |
| **EGFR (21L858R vs. 19Del)** | 0.559 | 1.280 (0.559, 2.932) |
| **EGFR (Other vs. 19Del)** | 0.624 | 1.232 (0.535, 2.833) |
| **N stage N0 as reference: N1** | 0.028 | 1.163 (1.170, 2.891) |
| **N2** | 0.016 | 2.206 (1.284, 3.785) |
| **N3** | 0.005 | 2.904 (1.605, 5.252) |
| **Performance status score (≥2 vs. <2)** | 0.384 | 1.258 (0.751, 2.107) |
| **Pulmonary metastasis (No vs. Yes)** | 0.587 | 0.878 (0.555,1.391) |
| **Pathology (Ade vs Other)** | 0.140 | 0.788 (0.256, 1.870) |
| **Brain metastasis (No vs. Yes)** | 0.052 | 0.646 (0.416, 1.204) |
| **Bone metastasis(No vs. Yes)** | 0.749 | 1.562 (0.659, 2.562) |
| **Liver metastasis(No vs. Yes)** | 0.882 | 0.990 (0.432, 2.271) |
| **Signature (Rapid- vs. Slow-progression)** | <0.0001 | 5.181 (3.242, 8.266) |

**Supplementary Table S2.** Detail of treatment of all the patients. Including the administration of TKIs, mean time of TKI therapy and chemotherapy of each department. Mean time represents the mean time of patients received treatment at each department. Discontinuation represents the cases of had discontinued during treatment.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Drugs** | **Patients received TKI therapy** | | | **Patients received chemotherapy (comparison)** | |
| **Training cohort:**  **Shanghai Pulmonary Hospital** | **Validation cohort 1:**  **Guangdong General Hospital** | **Validation cohort 2:**  **West China Hospital** | **Shanghai Pulmonary Hospital** | **Cancer Hospital, Chinese Academy of Medical Sciences** |
| **Icotinib** | 41 (35.0%) |  | 24 (25.0%) |  |  |
| **Gefitinib** | 76 (65.0%) | 73 (72.3%) | 49 (51.0%) |  |  |
| **Erlotinib** |  | 28 (27.7%) | 23 (24.0%) |  |  |
| **Total TKI cases** | 117 | 101 | 96 |  |  |
| **Chemotherapy** |  |  |  | 37 | 19 |
| **Mean time (months)** | 8.0 | 8.6 | 7.9 | 6.1 | 6.4 |
| **P of mean time** | 0.811 | | | 0.562 | |
| **Discontinuation** | 8 (6.8%) | 5 (4.9%) | 4 (4.2%) | 1 (2.7%) | 0 (0) |
| **Enrollment time** | May 1, 2013, to November 30, 2015. | June 1, 2010, to November 30, 2015 | January 1, 2013, to November 30, 2015 | January 1, 2012, to November 30, 2015. | November 1, 2011, to November 30, 2015 |

**Supplementary Table S3.** Univariate association of twelve features and progression-free survival in the training dataset. Cut-off values were calculated by x-tile.

|  |  |  |  |
| --- | --- | --- | --- |
| **Features** | **Cut-off** | **Training dataset (N = 117)** | |
| **p value** | **HR(95% CI)** |
| **Texture** |  |  |  |
| Contrast of Co-occurrence on LL in the 0° direction | 204864.0 | 0.0012 | 0.542 (0.383, 0.766) |
| Maximum-Probability of Co-occurrence on LL in the 0° direction | 81.0 | <0.0001 | 0.321 (0.184, 0.562) |
| Maximum-Probability of Co-occurrence on LL in the 45° direction | 62.0 | <0.0001 | 0.342 (0.210, 0.559) |
| Maximum-Probability of Co-occurrence on HL in the 0° direction | 102.0 | <0.0001 | 0.366 (0.234, 0.572) |
| Maximum-Probability of Co-occurrence on HL in the 45° direction | 107.0 | <0.0001 | 0.289 (0.174, 0.477) |
| Long-Run-High-Gray-Level Emphasis of Run Length on HL | 1124.0 | <0.0001 | 0.420 (0.278, 0.637) |
| **Gabor** |  |  |  |
| Entropy of GPTR in the 225° direction by two pixel steps | -4.79 | 0.006 | 0.601 (0.420, 0.862) |
| Entropy of GPTR in the 45° direction by four pixel steps | -4.84 | <0.0001 | 0.395 (0.274, 0.570) |
| Variance of GMTR in the 90° direction by four pixel steps | 5672.0 | 0.002 | 1.898 (1.269, 2.838) |
| Entropy of GPTR in the 135° direction by four pixel steps | -4.90 | <0.0001 | 0.395 (0.274, 0.570) |
| Variance of GMTR in the 225° direction by five pixel steps | 339170.0 | <0.0001 | 1.928 (1.345, 2.763) |
| **Volume Feature** |  |  |  |
| Maximum diameter of tumor | 11640.0 | 0.001 | 0.493 (0.327, 0.743) |

Gabor magnitude texture representation (GMTR) and Gabor phase-based texture representation (GPTR) are captured using multi-scale and multi-directional Gabor wavelet function. We extracted Run Length features and Co-occurrence features from eight directions (0°, 45, 90°, 135°, 180°, 225°, 270° and 315°), and the Gabor features were also extracted from eight directions and five pixel steps (1, 2, 3, 4, and 5) from the images after applying the Gabor filter.

**Supplementary Table S4.** Comparison of clinical variables between the rapid-progression subgroup and slow-progression subgroup in EGFR-TKI cases.

|  |  |
| --- | --- |
| **Variables**  (rapid-progression vs. slow-progression) | **Patients (N = 314)** |
| **p value** |
| **PFS** | <0.0001 |
| **Gender** | 0.518 |
| **Smoke** | 0.002 |
| **Age** | 0.772 |
| **Treatment line** | 0.380 |
| **T stage** | 0.100 |
| **Tumor location** | 0.852 |
| **EGFR mutation type** | 0.740 |
| **N stage** | 0.051 |
| **Performance status score** | 0.567 |
| **Pulmonary metastasis** | 0.302 |
| **Pathology** | 0.753 |
| **Brain metastasis** | 0.084 |
| **Bone metastasis** | 0.112 |
| **Liver metastasis** | 0.572 |



**Supplementary Figure S1. X-tile plots of the twelve selected key features.** Coloration of the plot represents the degree of the association at each division, ranging from low (dark, black) to high (bright, red or green). The red bar represents inverse association between marker expression and survival, whereas the green bar represents direct association. **FeatureA:** Contrast of Co-occurrence on LL in the 0° direction; **FeatureB:** Maximum-Probability of Co-occurrence on LL in the 0° direction; **FeatureC:** Maximum-Probability of Co-occurrence on LL in the 45° direction; **FeatureD:** Maximum-Probability of Co-occurrence on HL in the 0° direction; **FeatureE:** Maximum-Probability of Co-occurrence on HL in the 45° direction; **FeatureF:** Long-Run-High-Gray-Level Emphasis of Run Length on HL; **FeatureG:** Entropy of GPTR in the 225° direction by two pixel steps; **FeatureH:** Entropy of GPTR in the 45° direction by four pixel steps; **FeatureI:** Variance of GMTR in the 90° direction by four pixel steps; **FeatureJ:** Entropy of GPTR in the 135° direction by four pixel steps; **FeatureK:** Variance of GMTR in the 225° direction by five pixel steps; **FeatureL:** Maximum diameter of tumor.



**Supplementary Figure S2. Turning parameter (λ) selection in the LASSO model.** The solid gray vertical lines represent the partial likelihood deviance ± standard error (SE). The dotted vertical lines are drawn at the optimal values by minimum criteria and 1-SE criteria. We plotted the partial likelihood deviance versus log(λ), where λ is the tuning parameter. λ = 0.67 with log(λ) = -0.4 was chosen in our model.

****

**Supplementary Figure S3.** **Stratified analysis of the signature.** Kaplan-Meier survival analysis for all the enrolled stage IV *EGFR*-mutant NSCLC stratified by clinicopathological risk factors. The high-risk and low-risk represents the rapid-progression and slow-progression patient groups, respectively. They were categorized according to the twelve-feature-based signature. P values were calculated using the log-rank test. PS = performance status. T stage (**A, B, C**, and **D**): T1, T2, T3 and T4, respectively. N stage (**E, F, G**, and **H**): N0, N1, N2, and N3, respectively. Intrapulmonary metastasis (**I, G**): no and yes, respectively. PS score (**K, L**): lower and higher, respectively. Brain metastasis (**M, N**): no and yes, respectively. Bone metastasis (**O, P**): no and yes, respectively.

****

**Supplementary Figure S4.** **Time-dependent ROC curves of the risk characteristics.** Time-dependent ROC curves were used to compare the prognostic accuracy of the twelve-feature-based signature with clinicopathological risk factors and single feature using all the enrolled stage IV *EGFR*-mutant NSCLC patients. ROC = receiver operator characteristic. AUC = area under curve. PS = performance status. P values show the AUC at 1-year PFS for the twelve-feature-based signature vs. the AUC at 1-year PFS for the other factors. (A) represents the comparisons of the prognostic accuracy according to the twelve-feature-based signature, **FeatureA:** Contrast of Co-occurrence on LL in the 0° direction; **FeatureB:** Maximum-Probability of Co-occurrence on LL in the 0° direction; **FeatureC:** Maximum-Probability of Co-occurrence on LL in the 45° direction; **FeatureD:** Maximum-Probability of Co-occurrence on HL in the 0° direction; **FeatureE:** Maximum-Probability of Co-occurrence on HL in the 45° direction; **FeatureF:** Long-Run-High-Gray-Level Emphasis of Run Length on HL; **FeatureG:** Entropy of GPTR in the 225° direction by two pixel steps; **FeatureH:** Entropy of GPTR in the 45° direction by four pixel steps; **FeatureI:** Variance of GMTR in the 90° direction by four pixel steps; **FeatureJ:** Entropy of GPTR in the 135° direction by four pixel steps; **FeatureK:** Variance of GMTR in the 225° direction by five pixel steps; **FeatureL:** Maximum diameter of tumor. (B) represents the comparisons of the prognostic accuracy according to the twelve-feature-based signature, T stage, N stage, smoke, pulmonary metastasis, gender, and pathology.



**Supplementary Figure S5.** Progression probability of three different patient cohorts. The blue line represents slow-progression subgroup TKI patients, the red line represents rapid-progression subgroup TKI patients, the orange line represents the patients received chemotherapy and classified to high risk group by the signature, and the cyan line represents the chemotherapy patients who classified into low risk group by the signature. Significant difference is not found (p = 0.901) between the no-TKI patients with high risk (rapid-progression) and no-TKI patients with low-risk (slow-progression).