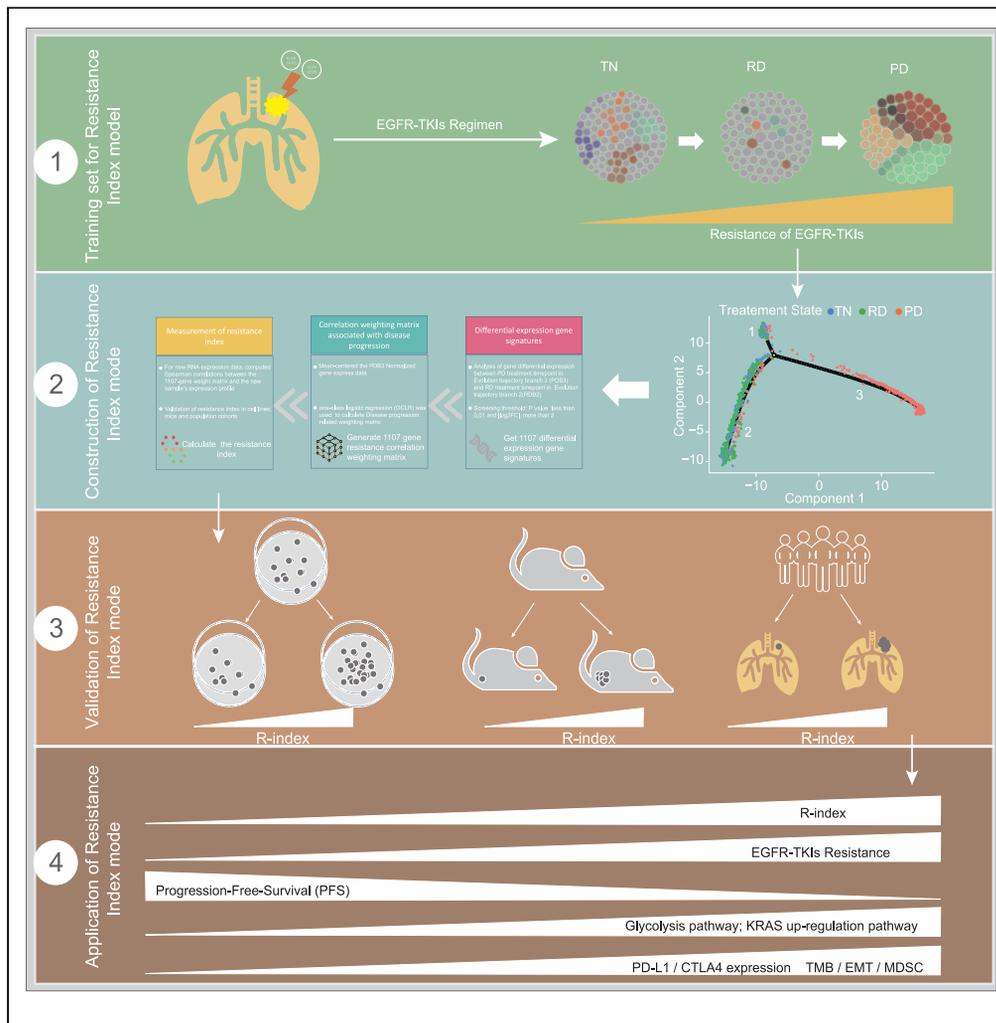**Article**

Stratification of non-small cell lung adenocarcinoma patients with EGFR actionable mutations based on drug-resistant stem cell genes

Xiaohong Xie, Lifeng Li, Liang Xie, ..., Xuefeng Xia, Zhiyi He, Chengzhi Zhou

zhiyi-river@163.com (Z.H.)
zhouchengzhi@gird.cn (C.Z.)

Highlights

Patients responded heterogeneously to EGFR-TKIs treatment

EGFR-TKIs resistance model (R-index) is built to evaluate patient resistance risk

The R-index is validated to perform well *in vivo* and *in vitro*

The R-index can explore mechanisms of unclear resistance and treatment strategies

Article

Stratification of non-small cell lung adenocarcinoma patients with EGFR actionable mutations based on drug-resistant stem cell genes

Xiaohong Xie,[1,8] Lifeng Li,[3,8] Liang Xie,[4,8] Zhentian Liu,[3] Guoliang Zhang,[3] Xuan Gao,[5,6] Wenying Peng,[7] Haiyi Deng,[1] Yilin Yang,[1] Meiling Yang,[2] Lianpeng Chang,[3] Xin Yi,[3] Xuefeng Xia,[3] Zhiyi He,[2,*] and Chengzhi Zhou[1,9,*]

SUMMARY

EGFR-TKIs were used in NSCLC patients with actionable EGFR mutations and prolong prognosis. However, most patients treated with EGFR-TKIs developed resistance within around one year. This suggests that residual EGFR-TKIs resistant cells may eventually lead to relapse. Predicting resistance risk in patients will facilitate individualized management. Herein, we built an EGFR-TKIs resistance prediction (R-index) model and validate in cell line, mice, and cohort. We found significantly higher R-index value in resistant cell lines, mice models and relapsed patients. Patients with an elevated R-index had significantly shorter relapse time. We also found that the glycolysis pathway and the KRAS upregulation pathway were related to EGFR-TKIs resistance. MDSC is a significant immunosuppression factor in the resistant microenvironment. Our model provides an executable method for assessing patient resistance status based on transcriptional reprogramming and may contribute to the clinical translation of patient individual management and the study of unclear resistance mechanisms.

INTRODUCTION

Lung cancer is the leading cause of cancer death worldwide.[1] According to previous reports, approximately 85% of patients are diagnosed with non-small cell lung cancer (NSCLC).[2] With the development of next-generation sequencing (NGS) technology and biomedicine, the treatment of NSCLC has been applied from inclusive radiotherapy and chemotherapy to personalized, targeted therapy and immunotherapy.

Epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKIs) have been used in the targeted therapy for NSCLC patients with activating mutations of *EGFR* since 2004 with a high response rate of 80%.[3,4] However, most patients treated with EGFR-TKIs developed resistance within a median of 10–14 months[5] The most common type of acquired resistance to the first and second generation of EGFR-TKIs is the *EGFR* T790M secondary point mutation.[6] Although the third-generation EGFR-TKIs have been clinically used for patients with the *EGFR* T790M mutation, the outcome remains disappointing.[5] The reported resistance mechanisms also include *PIK3CA* mutations, *BRAF* mutations, *c-MET* amplification, *AXL* overexpression, small-cell lung cancer transformation, epithelial-to-mesenchymal transition (EMT).[7–10] However, the mechanism responsible for approximately 30% of cases of resistance to EGFR-TKIs remains unclear.[11] This suggests that TKIs cannot eliminate 100% of cancer cells, and residual drug-resistant cells eventually lead to drug resistance in patients. Tumor heterogeneity provides a clear explanation for the level of treatment benefit observed in the patient cohort. Nonetheless, most decisions about lung cancer treatment are based on knowledge of a single carcinogenic driver without considering the functional consequences of continued oncogene-specific drug therapy.

Single-cell RNA sequencing (scRNA-seq) offers an opportunity to sample the whole transcriptome of individual cells and is one approach to dissect the heterogeneity of complex biological systems.[12,13] The one-class logistic regression (OCLR) machine learning algorithm provides a scalable approach to obtain stem cell signatures and determine the best base-level classification by extracting transcriptomes and

[1]Pulmonary and Critical Care Medicine, Guangzhou Institute of Respiratory Health, National Clinical Research Center for Respiratory Disease, National Center for Respiratory Medicine, State Key Laboratory of Respiratory Diseases, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong 510120, China

[2]Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi 530021, China

[3]Geneplus-Beijing, Beijing 102206, China

[4]Department of Thoracic Surgery, Guangdong Provincial People's Hospital/Guangdong Academy of Medical Sciences, Guangzhou, Guangdong 510080, China

[5]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

[6]Geneplus-Shenzhen Clinical Laboratory, Shenzhen, Guangdong 518122, China

[7]The Second Department of Oncology, Yunnan Cancer Hospital & The Third Affiliated Hospital of Kunming Medical University & Yunnan Cancer Center, Kunming 650000, China
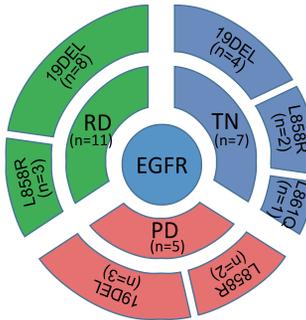
[8]These authors contributed equally

[9]Lead contact

*Correspondence:
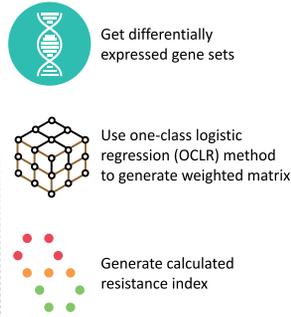zhiyi-river@163.com (Z.H.),
zhouchengzhi@gird.cn (C.Z.)
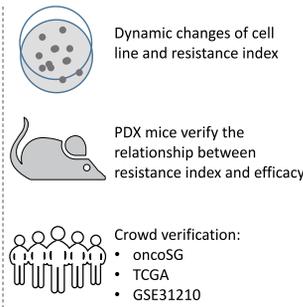https://doi.org/10.1016/j.isci.2023.106584

A Training set for Resistance Index model — Construction of Resistance Index model — Validation of Resistance Index model — Application of Resistance Index model
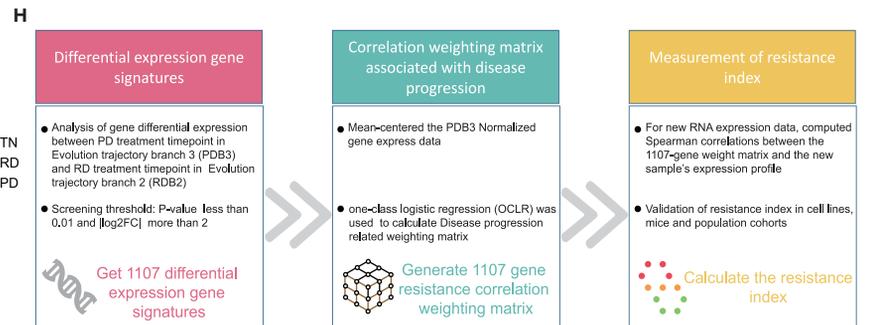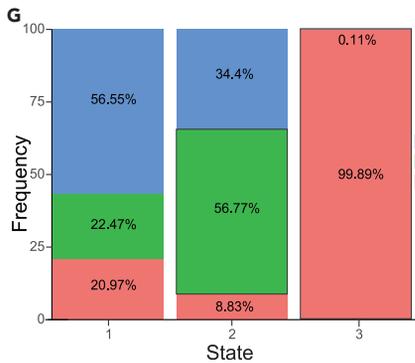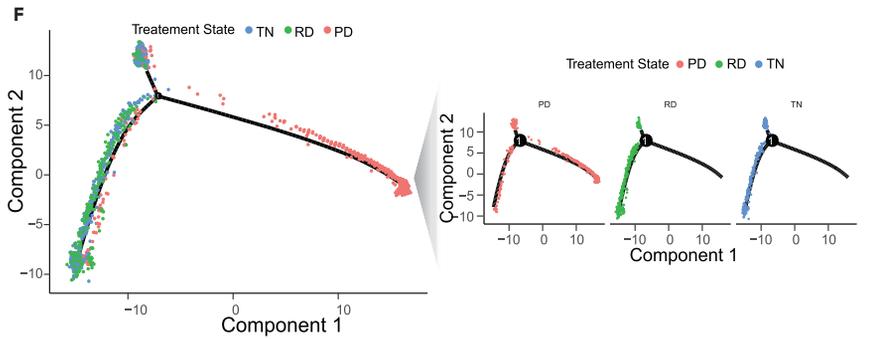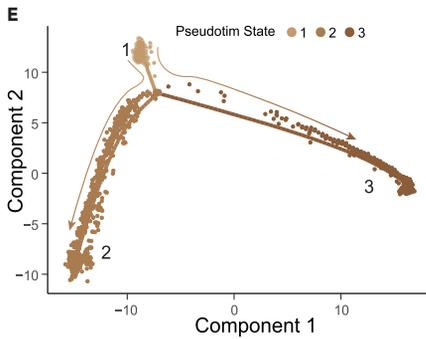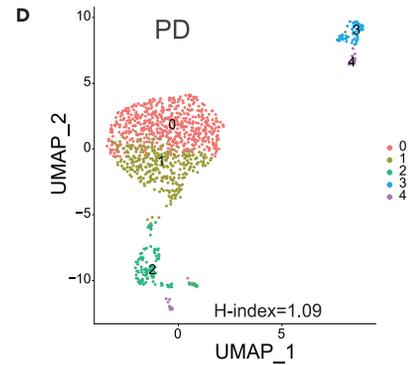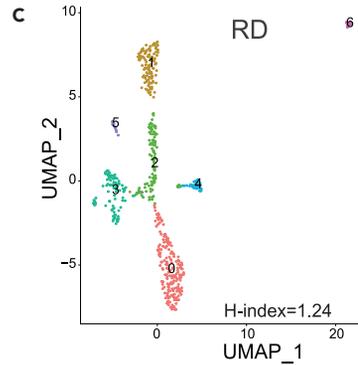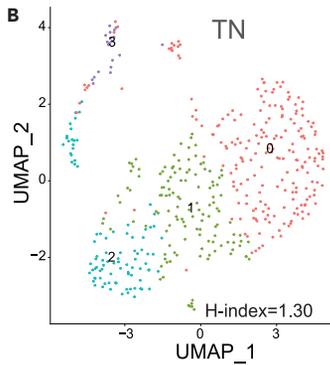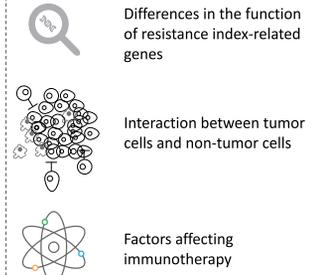
Figure 1. Development of the Resistance Index (R-index)

(A) Overall methodology. R-index training set and its development, verification, and application.

(B–D) Uniform Mani-fold Approximation and Projection (UMAP) plot and H-index of cancer cells at different treatment time points.

(E and F) The unsupervised transcriptional trajectory of cancer cells with monocle algorithm, colored by states and (F) treatment timepoint.

(G) The relative proportion of cancer cells for three treatment time points in each state.

(H) The workflow for the development and application of the R-index model.

See also Figures S1–S3, and Tables S1, S2, and S3.

differentiated progeny from non-transformed pluripotent stem cells.[14] Therefore, we aimed to detect the resistance of EGFR-TKIs at the single-cell transcriptome level with EGFR-mutated patients and fill the vacancy for preclinical research.

In this study, we utilized multiple treatment time points' single-cell RNA data to develop an EGFR-TKIs resistance index (R-index) model with an OCLR algorithm for the first time. This model is validated *in-vivo* and *in-vitro* experiments and three large real-world cohorts. Identifying an effective method to assess reprogramming-based TKI resistance in NSCLC may provide avenues for more durable clinical treatment of EGFR-mutant patients.

RESULTS

mRNA expression-based R-index model

To explore the resistance of cancer cells to EGFR-TKIs treatment, we enrolled an EGFR-TKIs treatment (Maynard et al., PRJNA591860) scRNA-seq dataset[15] which contain 14 individual patients and 23 samples with *EGFR* mutation. The sample information is displayed in Table S1 and Figure S1. We got an R-index model using the OCLR algorithm.[14] The OCLR-based R-index model was first verified in the EGFR-TKIs treatment cell line, mice model, and real-world clinical patient cohort. Then, we used R-index to explore the resistance mechanisms in aspects of the pathway, cell interaction, and immunosuppression (Figure 1A).

Identification of the drug resistance transcriptional heterogeneity of cancer cells and R-index signature genes

A total of 2,080 cancer cells were retained after quality control filtering as described in the star algorithm. All cancer cells at different treatment states were re-clustered and visualized using the UMAP (Uni-form Mani-fold Approximation and Projection) method. We hypothesize that only cells resistant to EGFR-TKIs are likely to survive and proliferate and result in reduced diversity. We find the cell clusters number at TN, RD, and PD were 4, 7, and 5, respectively (Figures 1B–1D). We also find that the 2-dimensional cell clusters of RD and PD are more diffusely clustered than TN, with each cluster having an independent distributional space. Indicating differential heterogeneity may be among cancer cells at different treatment timepoints. To quantify the heterogeneity among cancer cells, we calculated the diversity index of H-indexes in each cluster. The results showed that the H-index order was TN> RD> PD. This suggests that EGFR-TKIs interventions may affect cancer cell differentiation because of environmental screening, allowing resistant cells to obtain more competitive advantages over sensitive cells, reducing the diversity of cancer cell composition.

Trajectory analysis was executed with monocle software to project all cancer cells to explore the heterogeneity and the cells that play a significant role in governing the tumor progression (Figure 1E). The distribution of the three treatment response states (TN, RD and PD) in monocle is also distinguished (Figure 1F). Firstly, most cancer cells were located in separate trajectory branches, which marked their distinct differentiation states. For example, branch 3 was mainly occupied by cells from PD samples (PDB3, 99.89%), branch 1 by cells from TN samples (TNB1, 56.55%), and branch 2 by cells from RD samples (RDB2, 56.77%). Secondly, the differentiation direction of cancer cells in treatment time points was heterogeneous. For example, each branch had cells at various treatment time points except for branch 3 (Figure 1G).

To identify transcriptional signatures defining cellular resistance status in the trajectory, we compared differentially expressed genes between PDB3 and RDB2, and 1107 candidate genes were selected, in which 348 genes were upregulated in PDB3, and 759 genes were upregulated in RDB2 (Table S2). Considering that these genes may contain biological features that potentially distinguished the resistant states of the cancer cells, we applied the OCLR algorithm to build a trained model on PDB3 cells and produced a weighted 1107 gene matrix to extract transcriptomic features of the drug resistance signature
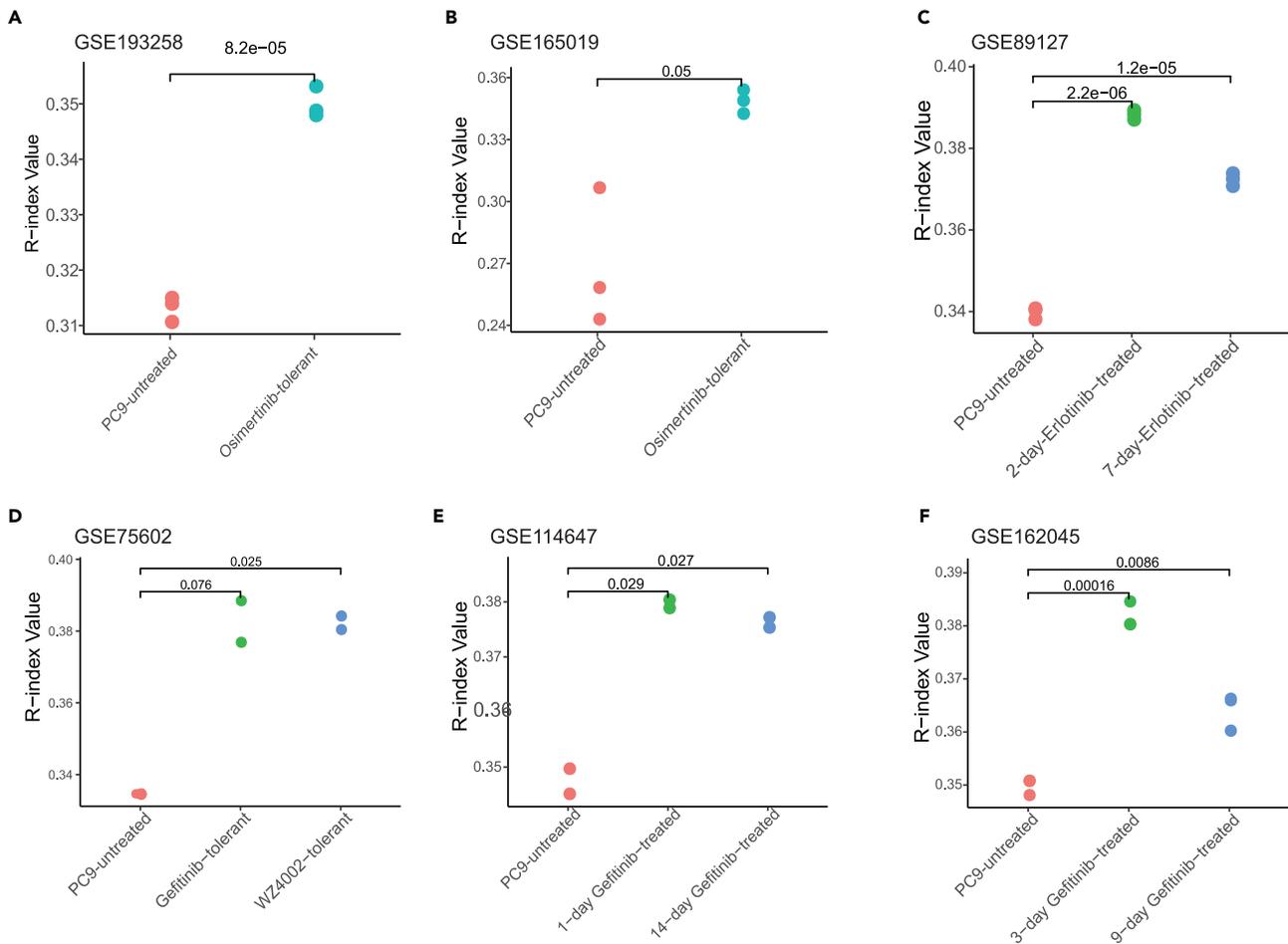
Figure 2. Validation of R-index in cell lines

(A–F) The comparison of the R-index in pro- and post EGFR-TKI treatment cell lines, and t-test was used for statistical tests. See also Figure S4.

(Table S3). The R-index score was defined as the spearman correlation coefficient of the 1107 gene signature matrix and gene expression values (Figure 1H).

Assess the predictive ability of R-index for resistance risk in cell lines and mice

Given that the R-index was hypothesized to evaluate the status of resistance of samples to EGFR inhibitors, we compared the R-index values of the EGFR-TKIs resistant PC9 cell lines with those of the untreated PC9 cell lines in GSE193258, GSE165019, GSE89127, GSE75602, GSE114647, GSE162045. We found that the EGFR-TKIs resistant cell line had significantly higher R-index values than the untreated cell line (Figures 2A–2F). We also observed a marginally significant negative correlation between the R-index and cell number in the intermittent administration (0, 1, 2, 4, 9, and 11 days) of erlotinib (Figure S4).

Because the cells in culture lacked *in-vivo* interactions, we applied R-index to investigate the response of mice to EGFR-TKIs treatment from xenograft data.[16,17] Patient-derived xenograft (PDX) models were built by implanting small pieces (3–5 mm) of adenocarcinomas specimens from patients' surgically resected tumors (SRT) with *EGFR* activating mutations (#7, #11) into the subcutaneous flank tissue of female SHO mice (Crlj: SHO-PrkdcscidHrhr, Charles River). Tumor size was measured with calipers once a week and the mice were treated by oral gavage with 25 mg/kg per day of osimertinib when tumor volume exceeded 500 mm^3. When tumor volume reached 1500 mm^3, mice were killed, and tumors were implanted into new mice. Tumor fragments #7 had *EGFR* L858R mutation, and #11 had *EGFR* exon 19 deletion mutation. The R-index
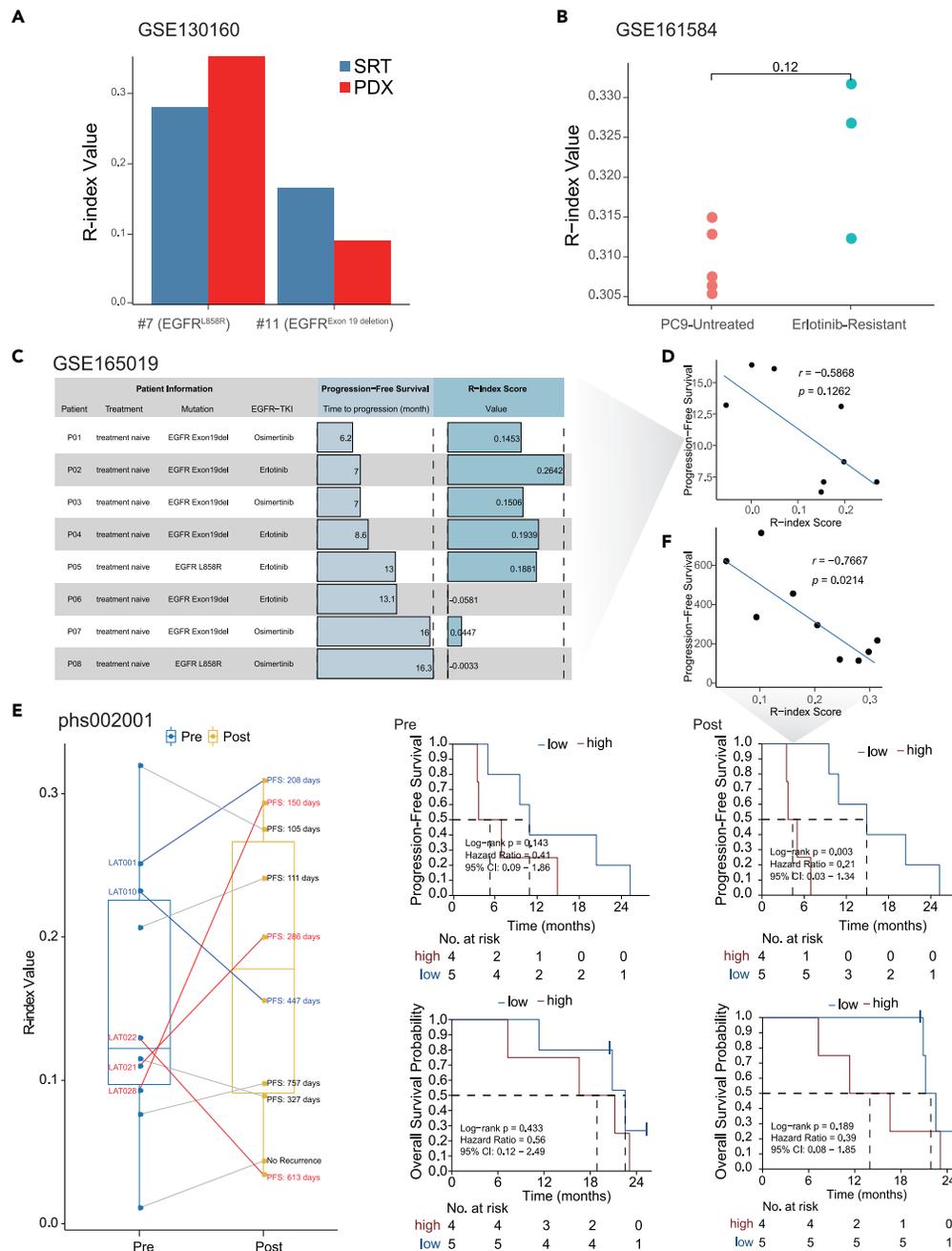
Figure 3. Validation of R-index in mice model and patients

(A and B) R-index changes in mice between treatment-naive and osimertinib or Erlotinib treatment *in-vivo*. SRT, surgically resected tumors; PDX, patient-derived xenografts.

(C) Changes of R-index and prognosis of patients pro- and post EGFR-TKI treatment.

(D) Correlation plot of R-index and PFS.

(E) Correlation of R-index and prognosis of patients' pro- and post EGFR-TKI treatment and the median R-index is used to divide high and low drug resistance groups.

(F) Correlation plot of R-index and PFS.

was calculated based on bulk mRNA gene expression. As PDX tumor in case #7 regrew during the continuous osimertinib treatment at the fifth passage, we calculated and found its R-index was higher than that of SRT. The PDX tumor in case #11 was cured at the third passage, and as expected, the R-index had an
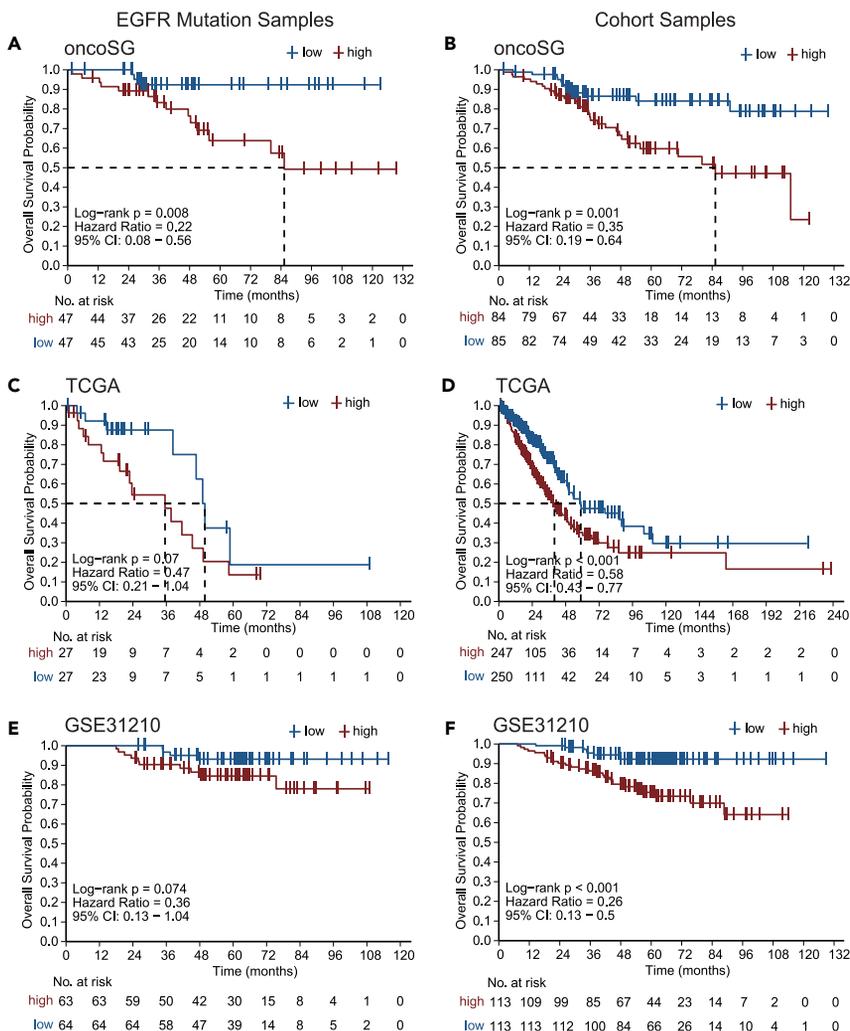
Figure 4. Validation of R-index in the human cohort
(A–F) *EGFR* mutation samples and the entire cohort samples in the OncoSG (A and B), TCGA LUAD (C and D), and GSE31210 (E and F) databases to verify the relationship between R-index and prognosis, and the threshold is the median R-index of the corresponding cohort.

opposite result compared with case #7 (Figure 3A). We also observe the same result in the GSE161584 dataset (Figure 3B).

Correlations of R-index with progression-free survival in the cohort data

After validating the feasibility of the R-index in the cell line and mice, we further verified R-index in real-world cohort. We found that the time to recurrence was shorter with high R-index values after osimertinib or erlotinib treatment (Figure 3C). The R-index is negatively correlated with the Progression-Free Survival (PFS) (Figure 3D). This result is also confirmed in an independent patient treatment cohort published by the Roper et al. (Figure 3E). Data from Roper et al. showed significant differences in PFS when stratified using the median post-treatment R-index value, with longer time to recurrence in patients with low R-index values (Figure 3E). We also observed a significant negative correlation between R-index and PFS (Figure 3F). To further verify the relationship between R-index and prognosis, we included three large cohorts for estimation. One transcriptome per patient was available for each cohort. We first calculated the R-index of *EGFR* mutant samples from the OncoSG database and dichotomized patients into two equal-size groups using the median R-index as the threshold. We found that the high R-index group showed a significantly shorter overall survival time than the low R-index group (p = 0.008, Figure 4A). Similarly, for whole samples of this cohort, the high R-index was also associated with worse outcomes (p = 0.001, Figure 4B). Another two lung
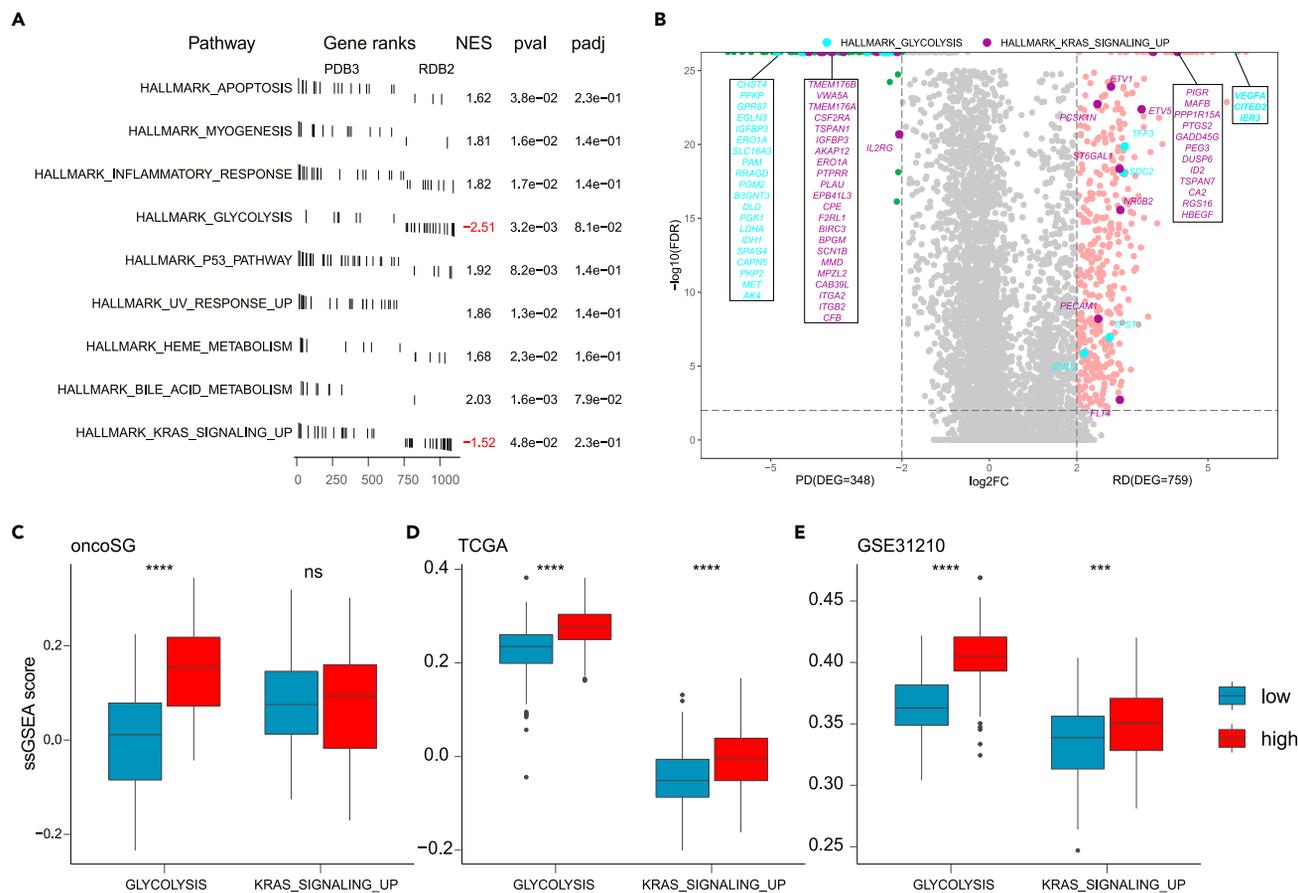
Figure 5. Enrichment of R-index-related functional gene sets

(A) R-index related function enrichment with hallmark gene sets from MSigDB analyzed by fgsea.

(B) The volcano map shows the detailed gene sets of the glycolysis pathway and KRAS upregulation pathway.

(C–E) Glycolysis and KRAS upregulation gene sets are orthogonally verified in public databases. Asterisks indicate pvalues, *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001; ****p ≤ 0.0001.

See also Figures S5 and S6.

adenocarcinoma bulk mRNA gene expression datasets were also used to validate the results further. The median resistance index was still used as the stratification threshold. As expected, in the datasets of the TCGA cohort and GSE31210 cohort, we also observed that the high R-index group showed a shorter overall survival time than the low R-index group, whether for *EGFR* mutant patients (p = 0.07, TCGA, Figure 4C; p = 0.07, GSE31210, Figure 4E) or the entire cohort patients (p< 0.001, TCGA, Figure 4D; p< 0.001, GSE31210, Figure 4F).

Analysis of R-index signature genes functional features in hallmark gene set

Our results show that the R-index related 1107 gene set may be used to assess the EGFR-TKIs resistance status of cancer cells in preclinical studies. To explore the potential functions of these genes, an fgsea script was used to analyze R-index signature genes with the hallmark gene set in MSigDB v7.4 and yielded 9 significantly enriched genesets. Metabolism-related glycolysis and signaling-related KRAS signaling up gene sets were significantly positive-enriched in PDB3 (Figure 5A), and the volcano plot highlighted related differential genes (Figure 5B).

To verify whether there were consistent results in other database, we estimated the value of each hallmark's ssGSEA profile in OncoSG, TCGA, and GSE31210. In line with expectations, glycolysis and KRAS signaling up expressed significantly higher in the high R-index group using median stratification (Figures 5C–5E). In addition, we also explored the GSEA enrichment of 1107 genes in the KEGG/GO/Reactome datasets and

found that cell cycle and stem cell related pathways are significantly enriched in PDB3 (Figure S5). The expression of the EMT geneset was also higher in the high R-index group (Figure S6).

Intercellular communication and immune microenvironment analysis

Because of the altered tumor microenvironment (TME) after EGFR-TKIs resistance, we also wanted to investigate R-index related immune microenvironment. First, we performed cell-cell communication analysis using CellPhoneDB between cancer cells and other immune cells in the TME. Based on our research purposes, we divided cancer cells into three types according to their evolutionary trajectories, namely PDB3 cells, RDB2 cells (used in the previous analysis), and Other_cancer_cells that did not contain these former two types of cells. Enriched receptor-ligand interactions network diagrams were derived based on the expression of receptors and the corresponding ligand between two connected cell types for demonstrating their extensive communication (Figure 6A). We further utilized receptor-ligand pairs to calculate the strengths of the interactions within PDB3 and RDB2, finding the close interactions with fibroblast, MF-Monocytes, endothelial, and dendritic cells in both PDB3 and RDB2. When using the odds ratio to normalize the receptor-ligand pairs of PDB3 and RDB2, we found that the Neutrophils, B-cells-M (B memory cell), B-cells-PB (B plasma cell), and T-cells had a higher ratio in the PDB3 state (Figure 6B).

Second, based on the results of quantitative analysis of the receptor-ligand pair, we specifically showed the interactions of PDB3 and RDB2 with four types of immune-related cells (Neutrophils, B-cells-M, B-cells-PB, and T-cells) (Figure 6C). We found that immunosuppressive-related receptor-ligand gene *ADORA2-B*,[18] *ENTPD1*,[19] *CXCR3*,[20] *LGALS9*[21] showed solid regulatory relationships with PDB3.

Finally, we explored the immune infiltration between RDB2 and PDB3. With immune surveillance and escape signatures,[22] we observed that PDB3 had higher immune escape ability (Figure 7A). We also examined the expression of immune checkpoint inhibitor-related genes *CD274* and *CTLA4* in the public database. We observed that the high R-index group had a significantly higher expression level by median stratification (Figure 7B). In addition, the TMB status had consistent results (Figures S7A and S7B). Considering that the EGFR-TKIs resistance had immunosuppressive features,[15] we used the TIDE algorithm to identify factors that excluded T cell infiltration into tumors from the large cohort. We found that MDSC (Myeloid-derived suppressor cells) was significantly higher in the high R-index group, and there was no significant difference between TAM.M2 (the M2 subtype of tumor-associated macrophages) and CAFs (cancer-associated fibroblasts) when using R-index median stratification (Figure 7C). We also observed that MDSC (Figure S8) and the ROS (reactive oxygen species) pathway (Figure S6) were significantly higher in the high R-index group by R-index median stratification.

DISCUSSION

EGFR-TKIs increased the five-year survival rate of late-stage patients harboring *EGFR* activating mutations. However, patients acquired resistance inevitably after a period of target treatment. At present, the exploration of EGFR-TKIs resistance mechanisms mainly focuses on off-target alterations or overexpression of several genes.[7–10] However, these factors can only be observed after acquired resistance, and there are still about 30% unknown resistance factors. An effective RNA-based EGFR-TKIs resistance index may help explore unclear resistance factors and guide treatment strategies in preclinical studies. Using the single-cell RNA-seq data at three different treatment timepoints, we developed an R-index model trained from EGFR-targeted therapy samples to evaluate the risk of EGFR-TKIs resistance. The model was validated with *in-vitro* and *in-vivo* datasets and three large cohorts. We also orthogonally verified the performance of the R-index model by single-cell data and cohort data in the aspects of resistance pathway and immune microenvironment.

In the *in-vitro* cell line and *in-vivo* PDX validation scenario, we found significant differences in R-index between EGFR-TKIs resistant and untreated cell lines. We also observed consistent results in the PDX model that R-index increased in resistant mice and decreased in cured mice. It is indicated that the R-index can be applied to EGFR-TKIs resistance prediction in cell lines and mouse models. In a cell line drug intervention model at multiple time points, the PC9 cells received intermittent erlotinib treatment at the 0, 1, 2, 4, 9, and 11th days. The results showed that R-index perfectly fit the dynamic changes of cell number under erlotinib intervention with a clear negative correlation between the cell numbers and the average R-index of the PC9 cell line. Previous studies proposed an evolution-based treatment[23] or the ''drug holiday'' phenomenon in treatment[24] as a conceptual treatment strategy based on Darwinian dynamics of
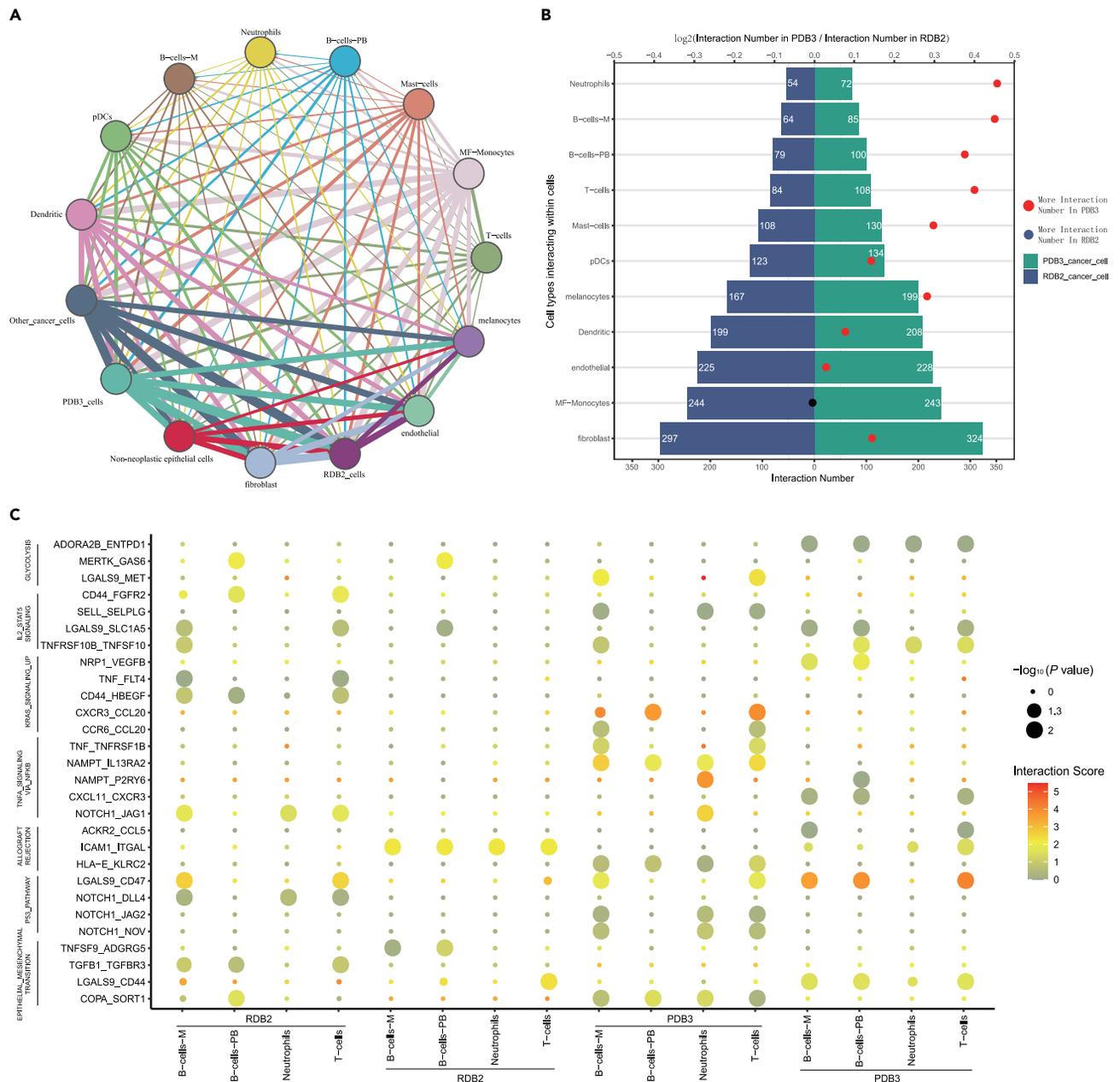
Figure 6. Interaction between cancer cells and tumor microenvironment cells

(A) Intercellular communication between cancer cells and other cells. Each line's color and thickness indicate the connection and proportion of the ligands and receptors, respectively.

(B) The bar chart shows the number of ligand-receptor pairs in cancer cells and other cells in PDB3 and RDB2 samples. The dots represent their ratio. The dots with ratios above 1 was in red and below 1 in black.

(C) Overview of selected ligand-receptor interactions in hallmark gene set between cancer cells and top-four ratio cell type.

intratumoral heterogeneity has developed for many years.[25] Clinically, many reports about salvage treatment demonstrated that patients with acquired resistance could re-respond to EGFR-TKIs re-challenge,[26–28] whereas the actual clinical benefit and timing of drug holiday have not been confirmed clearly. In our study, on day 4, when the number of cells was large and the resistance was theoretically low, the number of cells decreased significantly after receiving erlotinib. In contrast, the number of cells on day 9 was the lowest with the highest average R-index, and erlotinib intervention cannot effectively reduce the number of cells. At this time, stopping erlotinib intervention might be the best choice so that erlotinib-sensitive cells

Figure 7. The contributions of R-index stratification to the immune escape

(A) Comparison of Immune escape and Immune surveillance signature scores of PDB3 and RDB2.

(B) Comparison of *CD274* and *CTLA4* gene expression.

(C) TIDE score in external databases based on R-index stratification.

See also Figures S7 and S8.

could have a chance to proliferate and allow them to compete with erlotinib-resistant cells. Several cell line-based studies can also observe consistent results.[29] Therefore, the cell line results validated the reliability of the R-index and showed the potential value of the R-index in clinical decision-making for drug interventions.

The enrichment of EGFR-TKIs resistance-related pathways reflected R-index performance. Fgsea analysis showed that the glycolysis metabolism and KRAS upregulate pathways were significantly enriched in PDB3. We also got consistent results by R-index stratification in the three large cohorts. The Warburg effect describes the phenomenon that tumor cells increased utilization of glycolysis rather than oxidative phosphorylation to dominate ATP production despite adequate physiological oxygen conditions.[30] On the one hand, glycolysis can depress tumor cell differentiation and apoptosis to promote proliferation.[31,32]

On the other hand, glycolysis produces excessive lactate to create an acidic tumor microenvironment that promotes invasion and migration.[33] The RAS oncogene was first revealed through its ability to promote glycolysis and associated with resistance to targeted therapy.[34,35] As the downstream mechanism of the EGFR signaling pathway, the activation of KRAS-RAF-ERK plays a central role in the malignant transformation of normal cells. We also observed that the EMT geneset (Figure S6) was significantly upregulated in the high R-index group, which is an apparent EGFR-TKIs resistance factor.[36] In three cohorts, the prognosis of the high R-index group was worse than the low R-index group, which was consistent with the EGFR-TKIs resistance status, reflecting the dilemma of patient treatment after drug resistance. Several studies demonstrated that increased glucose metabolism in tumor cells was associated with resistance to EGFR-TKIs treatment. Thus, the combined use of glucose metabolism inhibitors may be a potential therapeutic strategy.[37–39] Inhibition of increased lactic acid production can also affect disease progression.[40] KRAS and its downstream stand-out signaling pathways, such as MAPK, PI3K, and RAL-GDS, have been used as essential sources to discover treatment opportunities.[35,41] These reflect the contribution of new indicators to the application of innovative therapies.

In analyzing the immune microenvironment, we used single-cell data and cohort data to verify that MDSC is a major immunosuppressive factor orthogonally. We compared the immune escape score[22] between PDB3 and RDB2, finding that the PDB3 had a higher immune escape ability. The immune checkpoint genes of *PD-L1* and *CTLA4* expression (Figure 7B) and TMB (Figure S7) showed consistent results with significantly higher values in the high R-index group. Several EGFR-TKIs treatment studies also found that the *PD-L1* expression and TMB values of resistant samples increased,[42,43] indicating that the R-index has a good predictive performance. We used the TIDE algorithm to find further that MDSC-mediated immune exclusion may be a significant factor of immune escape in three cohorts. To our knowledge, MDSC played a vital role in the resistance of tumor cells against the immune system to specific therapies.[44–46] We also observed that the IN-gamma (INFG) (Figure 7C) and ROS pathway (Figure S6), closely related to MDSC, were significantly enriched in the high R-index group. In addition, the MDSC-related receptor-ligand pairs such as *ADORA2B-ENTPD*[47] were significantly enriched in PDB3. The interaction between neutrophils and PDB3 was also closer than that of RDB2. The neutrophil is an important immune component regulating the adaptive immune responses by expressing a wide repertoire of cytokines.[48] In most human tumors, tumor-related neutrophil infiltration was associated with poor prognosis.[49] MDSC-related neutrophils were associated with immunosuppressive activity.[50] The consistency of the results of scRNA and cohort data reflected the value of the R-index in the preclinical study of EGFR-TKIs resistance.

Owing to the biological plasticity of cancer cells, effective early treatment often produces drug resistance in the later stage.[51] Several studies have explored the mechanisms by comparing baseline and re-biopsy tissue specimens.[52,53] However, performing surgery or biopsy to obtain tissue from relapsed patients has many limitations, even in large, well-designed clinical trials.[54,55] Identifying the molecular mechanism of resistance is necessary for effective posterior therapy. Combining glucose metabolism-targeted or MDSC-targeted therapies with current clinical treatment may be a possible way to overcome EGFR-TKIs

resistance. Based on the prediction accuracy and performance of the R-index, we consider that R-index can be used as a tool for preclinical research to assess the resistance status of EGFR-TKIs.

Conclusions

Our study provides a potentially clinically applicable EGFR-TKIs resistant biomarker of R-index in NSCLC patients with actionable EGFR mutations at the transcriptome level, which may contribute to clinical translation for individual patient management and for the study of mechanisms of resistance that are currently unclear.

Limitations of the study

There are still several limitations in our study. First, it is difficult to obtain tissue from advanced EGFR-TKIs resistance patients at multiple timepoints and large-scale prospective validation studies cannot be completed in a short time. Another shortcoming is that the bulk RNA data of EGFR-TKIs resistance tissue were scarce, which precluded us from setting a cohort to directly analyze the R-index's effectiveness. Moreover, our R-index model may currently only be suitable for preclinical exploratory research, and it needs further evaluation when applied to clinical decision-making.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - The maynard et al.'s single-cell RNA sequence (scRNA-seq) data
  - Maynard et al.'s scRNA-seq data quality control and R-index model
  - Cell line validation data
  - Mice model validation data
  - EGFR-TKIs treatment validation data
  - Cohort validation data
  - Cancer cell cluster diversity
  - Survival analysis
  - fgsea analysis
  - ssGSEA analysis
  - Gene set enrichment analysis
  - Cell-cell interaction analysis
  - TIDE analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106584.

AUTHOR CONTRIBUTIONS

Z.H. and C.Z. are the project managers of this study. X. Xie, L.L., and L.X. are in charge of paper writing, data analysis, and conceptualization. Z.L., G-L.Z., X.G., W-Y.P., X.Y., and X. Xia are in charge of paper revision and editing. H.D., Y.Y., M.Y., and L.C. contribute to the visualization of results.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2021). Cancer statistics, 2021. CA. Cancer J. Clin. *71*, 7–33. https://doi.org/10.3322/caac.21654.

2. Molina, J.R., Yang, P., Cassivi, S.D., Schild, S.E., and Adjei, A.A. (2008). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin. Proc. *83*, 584–594. https://doi.org/10.4065/83.5.584.

3. Lynch, T.J., Bell, D.W., Sordella, R., Gurubhagavatula, S., Okimoto, R.A., Brannigan, B.W., Harris, P.L., Haserlat, S.M., Supko, J.G., Haluska, F.G., et al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N. Engl. J. Med. *350*, 2129–2139. https://doi.org/10.1056/NEJMoa040938.

4. Paez, J.G., Jänne, P.A., Lee, J.C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F.J., Lindeman, N., Boggon, T.J., et al. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science *304*, 1497–1500. https://doi.org/10.1126/science.1099314.

5. Rosell, R., Moran, T., Queralt, C., Porta, R., Cardenal, F., Camps, C., Majem, M., Lopez-Vivanco, G., Isla, D., Provencio, M., et al. (2009). Screening for epidermal growth factor receptor mutations in lung cancer. N. Engl. J. Med. *361*, 958–967. https://doi.org/10.1056/NEJMoa0904554.

6. Arcila, M.E., Oxnard, G.R., Nafa, K., Riely, G.J., Solomon, S.B., Zakowski, M.F., Kris, M.G., Pao, W., Miller, V.A., and Ladanyi, M. (2011). Rebiopsy of lung cancer patients with acquired resistance to EGFR inhibitors and enhanced detection of the T790M mutation using a locked. Clin. Cancer Res. *17*, 1169–1180.

7. Bar, J., and Onn, A. (2012). Overcoming molecular mechanisms of resistance to first-generation epidermal growth factor receptor tyrosine kinase inhibitors. Clin. Lung Cancer *13*, 267–279. https://doi.org/10.1016/j.cllc.2011.09.001.

8. Ohashi, K., Maruvka, Y.E., Michor, F., and Pao, W. (2013). Epidermal growth factor receptor tyrosine kinase inhibitor-resistant disease. J. Clin. Oncol. *31*, 1070–1080. https://doi.org/10.1200/jco.2012.43.3912.

9. Sequist, L.V., Waltman, B.A., Dias-Santagata, D., Digumarthy, S., Turke, A.B., Fidias, P., Bergethon, K., Shaw, A.T., Gettinger, S., Cosper, A.K., et al. (2011). Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Sci. Transl. Med. *3*, 75ra26. https://doi.org/10.1126/scitranslmed.3002003.

10. Zhang, Z., Lee, J.C., Lin, L., Olivas, V., Au, V., LaFramboise, T., Abdel-Rahman, M., Wang, X., Levine, A.D., Rho, J.K., et al. (2012). Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. Nat. Genet. *44*, 852–860. https://doi.org/10.1038/ng.2330.

11. Kunimasa, K., Nagano, T., Shimono, Y., Dokuni, R., Kiriu, T., Tokunaga, S., Tamura, D., Yamamoto, M., Tachihara, M., Kobayashi, K., et al. (2017). Glucose metabolism-targeted therapy and withaferin A are effective for epidermal growth factor receptor tyrosine kinase inhibitor-induced drug-tolerant persisters. Cancer Sci. *108*, 1368–1377. https://doi.org/10.1111/cas.13266.

12. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196. https://doi.org/10.1126/science.aad0501.

13. Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., and Klein, A.M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science *360*, eaar5780. https://doi.org/10.1126/science.aar5780.

14. Sokolov, A., Paull, E.O., and Stuart, J.M. (2016). ONE-CLASS detection of cell states in tumor subtypes. Pac. Symp. Biocomput. *21*, 405–416.

15. Maynard, A., McCoach, C.E., Rotow, J.K., Harris, L., Haderk, F., Kerr, D.L., Yu, E.A., Schenk, E.L., Tan, W., Zee, A., et al. (2020). Therapy-Induced evolution of human lung cancer revealed by single-cell RNA sequencing. Cell *182*, 1232–1251.e22. https://doi.org/10.1016/j.cell.2020.07.017.

16. Kita, K., Fukuda, K., Takahashi, H., Tanimoto, A., Nishiyama, A., Arai, S., Takeuchi, S., Yamashita, K., Ohtsubo, K., Otani, S., et al. (2019). Patient-derived xenograft models of non-small cell lung cancer for evaluating targeted drug sensitivity and resistance. Cancer Sci. *110*, 3215–3224. https://doi.org/10.1111/cas.14171.

17. Marrocco, I., Romaniello, D., Vaknin, I., Drago-Garcia, D., Oren, R., Uribe, M.L., Belugali Nataraj, N., Ghosh, S., Eilam, R., Salame, T.M., et al. (2021). Upfront admixing antibodies and EGFR inhibitors preempts sequential treatments in lung cancer models. EMBO Mol. Med. *13*, e13144. https://doi.org/10.15252/emmm.202013144.

18. Liu, H., Kuang, X., Zhang, Y., Ye, Y., Li, J., Liang, L., Xie, Z., Weng, L., Guo, J., Li, H., et al. (2020). ADORA1 inhibition promotes tumor immune evasion by regulating the ATF3-PD-L1 Axis. Cancer Cell *37*, 324–339.e8. https://doi.org/10.1016/j.ccell.2020.02.006.

19. Moesta, A.K., Li, X.Y., and Smyth, M.J. (2020). Targeting CD39 in cancer. Nat. Rev. Immunol. *20*, 739–755. https://doi.org/10.1038/s41577-020-0376-4.

20. Chow, M.T., Ozga, A.J., Servis, R.L., Frederick, D.T., Lo, J.A., Fisher, D.E., Freeman, G.J., Boland, G.M., and Luster, A.D. (2019). Intratumoral activity of the CXCR3 chemokine system is required for the efficacy of anti-PD-1 therapy. Immunity *50*, 1498–1512.e5. https://doi.org/10.1016/j.immuni.2019.04.010.

21. He, Y., Jia, K., Dziadziuszko, R., Zhao, S., Zhang, X., Deng, J., Wang, H., Hirsch, F.R., and Zhou, C. (2019). Galectin-9 in non-small cell lung cancer. Lung Cancer *136*, 80–85. https://doi.org/10.1016/j.lungcan.2019.08.014.

22. Sun, Y., Wu, L., Zhong, Y., Zhou, K., Hou, Y., Wang, Z., Zhang, Z., Xie, J., Wang, C., Chen, D., et al. (2021). Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. Cell *184*, 404–421.e16. https://doi.org/10.1016/j.cell.2020.11.041.

23. Zhang, J., Cunningham, J.J., Brown, J.S., and Gatenby, R.A. (2017). Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. Nat. Commun. *8*, 1816. https://doi.org/10.1038/s41467-017-01968-5.

24. Song, T., Yu, W., and Wu, S.X. (2014). Subsequent treatment choices for patients with acquired resistance to EGFR-TKIs in non-small cell lung cancer: restore after a drug holiday or switch to another EGFR-TKI? Asian Pac. J. Cancer Prev. *15*, 205–213. https://doi.org/10.7314/apjcp.2014.15.1.205.

25. Aktipis, C.A., Kwan, V.S.Y., Johnson, K.A., Neuberg, S.L., and Maley, C.C. (2011). Overlooking evolution: a systematic analysis of cancer relapse and therapeutic resistance research. PLoS One *6*, e26100. https://doi.org/10.1371/journal.pone.0026100.

26. Watanabe, S., Tanaka, J., Ota, T., Kondo, R., Tanaka, H., Kagamu, H., Ichikawa, K., Koshio, J., Baba, J., Miyabayashi, T., et al. (2011). Clinical responses to EGFR-tyrosine kinase inhibitor retreatment in non-small cell lung

cancer patients who benefited from prior effective gefitinib therapy: a retrospective analysis. BMC Cancer *11*, 1. https://doi.org/10.1186/1471-2407-11-1.

27. Yamaguchi, O., Kaira, K., Mouri, A., Shiono, A., Hashimoto, K., Miura, Y., Nishihara, F., Murayama, Y., Kobayashi, K., and Kagamu, H. (2019). Re-challenge of afatinib after 1st generation EGFR-TKI failure in patients with previously treated non-small cell lung cancer harboring EGFR mutation. Cancer Chemother. Pharmacol. *83*, 817–825. https://doi.org/10.1007/s00280-019-03790-w.

28. Oh, I.J., Ban, H.J., Kim, K.S., and Kim, Y.C. (2012). Retreatment of gefitinib in patients with non-small-cell lung cancer who previously controlled to gefitinib: a single-arm, open-label, phase II study. Lung Cancer *77*, 121–127. https://doi.org/10.1016/j.lungcan.2012.01.012.

29. Chmielecki, J., Foo, J., Oxnard, G.R., Hutchinson, K., Ohashi, K., Somwar, R., Wang, L., Amato, K.R., Arcila, M., Sos, M.L., et al. (2011). Optimization of dosing for EGFR-mutant non-small cell lung cancer with evolutionary cancer modeling. Sci. Transl. Med. *3*, 90ra59. https://doi.org/10.1126/scitranslmed.3002356.

30. Warburg, O. (1956). On the origin of cancer cells. Science *123*, 309–314. https://doi.org/10.1126/science.123.3191.309.

31. Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science *324*, 1029–1033. https://doi.org/10.1126/science.1160809.

32. Tomiyama, A., Serizawa, S., Tachibana, K., Sakurada, K., Samejima, H., Kuchino, Y., and Kitanaka, C. (2006). Critical role for mitochondrial oxidative phosphorylation in the activation of tumor suppressors Bax and Bak. J. Natl. Cancer Inst. *98*, 1462–1473. https://doi.org/10.1093/jnci/djj395.

33. Hirschhaeuser, F., Sattler, U.G.A., and Mueller-Klieser, W. (2011). Lactate: a metabolic key player in cancer. Cancer Res. *71*, 6921–6925. https://doi.org/10.1158/0008-5472.can-11-1457.

34. Racker, E., Resnick, R.J., and Feldman, R. (1985). Glycolysis and methylaminoisobutyrate uptake in rat-1 cells transfected with ras or myc oncogenes. Proc. Natl. Acad. Sci. USA *82*, 3535–3538. https://doi.org/10.1073/pnas.82.11.3535.

35. Kitajima, S., Asahina, H., Chen, T., Guo, S., Quiceno, L.G., Cavanaugh, J.D., Merlino, A.A., Tange, S., Terai, H., Kim, J.W., et al. (2018). Overcoming resistance to dual innate immune and MEK inhibition downstream of KRAS. Cancer Cell *34*, 439–452.e6. https://doi.org/10.1016/j.ccell.2018.08.009.

36. Tulchinsky, E., Demidov, O., Kriajevska, M., Barlev, N.A., and Imyanitov, E. (2019). EMT: a mechanism for escape from EGFR-targeted therapy in lung cancer. Biochim. Biophys. Acta, Rev. Cancer *1871*, 29–39. https://doi.org/10.1016/j.bbcan.2018.10.003.

37. Tamada, M., Nagano, O., Tateyama, S., Ohmura, M., Yae, T., Ishimoto, T., Sugihara, E., Onishi, N., Yamamoto, T., Yanagawa, H., et al. (2012). Modulation of glucose metabolism by CD44 contributes to antioxidant status and drug resistance in cancer cells. Cancer Res. *72*, 1438–1448. https://doi.org/10.1158/0008-5472.can-11-3024.

38. Suzuki, S., Okada, M., Takeda, H., Kuramoto, K., Sanomachi, T., Togashi, K., Seino, S., Yamamoto, M., Yoshioka, T., and Kitanaka, C. (2018). Involvement of GLUT1-mediated glucose transport and metabolism in gefitinib resistance of non-small-cell lung cancer cells. Oncotarget *9*, 32667–32679. https://doi.org/10.18632/oncotarget.25994.

39. Kim, S.M., Yun, M.R., Hong, Y.K., Solca, F., Kim, J.H., Kim, H.J., and Cho, B.C. (2013). Glycolysis inhibition sensitizes non-small lung cancer with T790M mutation to irreversible EGFR inhibitors via translational suppression of Mcl-1 by AMPK activation. Mol. Cancer Therapeut. *12*, 2145–2156. https://doi.org/10.1158/1535-7163.mct-12-1188.

40. Xie, H., Hanai, J.I., Ren, J.G., Kats, L., Burgess, K., Bhargava, P., Signoretti, S., Billiard, J., Duffy, K.J., Grant, A., et al. (2014). Targeting lactate dehydrogenase–a inhibits tumorigenesis and tumor progression in mouse models of lung cancer and impacts tumor-initiating cells. Cell Metabol. *19*, 795–809. https://doi.org/10.1016/j.cmet.2014.03.003.

41. Lito, P., Rosen, N., and Solit, D.B. (2013). Tumor adaptation and resistance to RAF inhibitors. Nat. Med. *19*, 1401–1409. https://doi.org/10.1038/nm.3392.

42. Peng, S., Wang, R., Zhang, X., Ma, Y., Zhong, L., Li, K., Nishiyama, A., Arai, S., Yano, S., and Wang, W. (2019). EGFR-TKI resistance promotes immune escape in lung cancer via increased PD-L1 expression. Mol. Cancer *18*, 165. https://doi.org/10.1186/s12943-019-1073-4.

43. Isomoto, K., Haratani, K., Hayashi, H., Shimizu, S., Tomida, S., Niwa, T., Yokoyama, T., Fukuda, Y., Chiba, Y., Kato, R., et al. (2020). Impact of EGFR-TKI treatment on the tumor immune microenvironment in EGFR mutation-positive non-small cell lung cancer. Clin. Cancer Res. *26*, 2037–2046. https://doi.org/10.1158/1078-0432.ccr-19-2027.

44. Domagala, M., Laplagne, C., Leveque, E., Laurent, C., Fournié, J.J., Espinosa, E., and Poupot, M. (2021). Cancer cells resistance shaping by tumor infiltrating myeloid cells. Cancers *13*, 165. https://doi.org/10.3390/cancers13020165.

45. Weber, R., Fleming, V., Hu, X., Nagibin, V., Groth, C., Altevogt, P., Utikal, J., and Umansky, V. (2018). Myeloid-derived suppressor cells hinder the anti-cancer activity of immune checkpoint inhibitors. Front. Immunol. *9*, 1310. https://doi.org/10.3389/fimmu.2018.01310.

46. Li, R., Salehi-Rad, R., Crosson, W., Momcilovic, M., Lim, R.J., Ong, S.L., Huang, Z.L., Zhang, T., Abascal, J., Dumitras, C., et al. (2021). Inhibition of granulocytic myeloid-derived suppressor cells overcomes resistance to immune checkpoint inhibition in LKB1-deficient non-small cell lung cancer. Cancer Res. *81*, 3295–3308. https://doi.org/10.1158/0008-5472.can-20-3564.

47. Chen, S., Akdemir, I., Fan, J., Linden, J., Zhang, B., and Cekic, C. (2020). The expression of adenosine A2B receptor on antigen-presenting cells suppresses CD8(+) T-cell responses and promotes tumor growth. Cancer Immunol. Res. *8*, 1064–1074. https://doi.org/10.1158/2326-6066.cir-19-0833.

48. Scapini, P., and Cassatella, M.A. (2014). Social networking of human neutrophils within the immune system. Blood *124*, 710–719. https://doi.org/10.1182/blood-2014-03-453217.

49. Shaul, M.E., and Fridlender, Z.G. (2019). Tumour-associated neutrophils in patients with cancer. Nat. Rev. Clin. Oncol. *16*, 601–620. https://doi.org/10.1038/s41571-019-0222-4.

50. Bronte, V., Brandau, S., Chen, S.H., Colombo, M.P., Frey, A.B., Greten, T.F., Mandruzzato, S., Murray, P.J., Ochoa, A., Ostrand-Rosenberg, S., et al. (2016). Recommendations for myeloid-derived suppressor cell nomenclature and characterization standards. Nat. Commun. *7*, 12150. https://doi.org/10.1038/ncomms12150.

51. Yuan, S., Norgard, R.J., and Stanger, B.Z. (2019). Cellular plasticity in cancer. Cancer Discov. *9*, 837–851. https://doi.org/10.1158/2159-8290.cd-19-0015.

52. Kobayashi, S., Boggon, T.J., Dayaram, T., Jänne, P.A., Kocher, O., Meyerson, M., Johnson, B.E., Eck, M.J., Tenen, D.G., and Halmos, B. (2005). EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N. Engl. J. Med. *352*, 786–792. https://doi.org/10.1056/NEJMoa044238.

53. Yu, H.A., Arcila, M.E., Rekhtman, N., Sima, C.S., Zakowski, M.F., Pao, W., Kris, M.G., Miller, V.A., Ladanyi, M., and Riely, G.J. (2013). Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. Clin. Cancer Res. *19*, 2240–2247. https://doi.org/10.1158/1078-0432.ccr-12-2246.

54. Fukuoka, M., Wu, Y.L., Thongprasert, S., Sunpaweravong, P., Leong, S.S., Sriuranpong, V., Chao, T.Y., Nakagawa, K., Chu, D.T., Saijo, N., et al. (2011). Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS). J. Clin. Oncol. *29*, 2866–2874. https://doi.org/10.1200/jco.2010.33.4235.

55. Lee, S.M., Park, C.M., Lee, K.H., Bahn, Y.E., Kim, J.I., and Goo, J.M. (2014). C-arm cone-beam CT-guided percutaneous transthoracic needle biopsy of lung nodules: clinical experience in 1108 patients. Radiology *271*, 291–300. https://doi.org/10.1148/radiol.13131265.

56. Gogleva, A., Polychronopoulos, D., Pfeifer, M., Poroshin, V., Ughetto, M., Martin, M.J., Thorpe, H., Bornot, A., Smith, P.D., Sidders, B., et al. (2022). Knowledge graph-based recommendation framework identifies drivers of resistance in EGFR mutant non-small cell lung cancer. Nat. Commun. 13, 1667. https://doi.org/10.1038/s41467-022-29292-7.

57. Gurule, N.J., McCoach, C.E., Hinz, T.K., Merrick, D.T., Van Bokhoven, A., Kim, J., Patil, T., Calhoun, J., Nemenoff, R.A., Tan, A.C., et al. (2021). A tyrosine kinase inhibitor-induced interferon response positively associates with clinical response in EGFR-mutant lung cancer. NPJ Precis. Oncol. 5, 41. https://doi.org/10.1038/s41698-021-00181-4.

58. Rusan, M., Li, K., Li, Y., Christensen, C.L., Abraham, B.J., Kwiatkowski, N., Buczkowski, K.A., Bockorny, B., Chen, T., Li, S., et al. (2018). Suppression of adaptive responses to targeted cancer therapy by transcriptional repression. Cancer Discov. 8, 59–73. https://doi.org/10.1158/2159-8290.Cd-17-0461.

59. Hata, A.N., Niederst, M.J., Archibald, H.L., Gomez-Caraballo, M., Siddiqui, F.M., Mulvey, H.E., Maruvka, Y.E., Ji, F., Bhang, H.e.C., Krishnamurthy Radhakrishna, V., et al. (2016). Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. Nat. Med. 22, 262–269. https://doi.org/10.1038/nm.4040.

60. Raoof, S., Mulford, I.J., Frisco-Cabanos, H., Nangia, V., Timonina, D., Labrot, E., Hafeez, N., Bilton, S.J., Drier, Y., Ji, F., et al. (2019). Targeting FGFR overcomes EMT-mediated resistance in EGFR mutant non-small cell lung cancer. Oncogene 38, 6399–6413. https://doi.org/10.1038/s41388-019-0887-2.

61. Aissa, A.F., Islam, A., Ariss, M.M., Go, C.C., Rader, A.E., Conrardy, R.D., Gajda, A.M., Rubio-Perez, C., Valyi-Nagy, K., Pasquinelli, M., et al. (2021). Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. Nat. Commun. 12, 1628. https://doi.org/10.1038/s41467-021-21884-z.

62. Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-Origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell 173, 291–304.e6. https://doi.org/10.1016/j.cell.2018.03.022.

63. Okayama, H., Kohno, T., Ishii, Y., Shimada, Y., Shiraishi, K., Iwakawa, R., Furuta, K., Tsuta, K., Shibata, T., Yamamoto, S., et al. (2012). Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res. 72, 100–111. https://doi.org/10.1158/0008-5472.can-11-1403.

64. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell 177, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

65. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods 14, 979–982. https://doi.org/10.1038/nmeth.4402.

66. R Core Team (2021). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing). https://www.R-project.org/.

67. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. Preprint at bioRxiv. https://doi.org/10.1101/060012.

68. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

69. Therneau, T., and Crowson, C. (2013). Using Time Dependent Covariates and Time Dependent Coefcients in the Cox Model (Survival Package).

70. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).

71. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. 14, 7. https://doi.org/10.1186/1471-2105-14-7.

72. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation 2, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

73. Zhang, J. GseaVis: implement for 'GSEA' enrichment visualization. R package version 0.0.5. https://github.com/junjunlab/GseaVis.

74. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat. Protoc. 15, 1484–1506. https://doi.org/10.1038/s41596-020-0292-x.

75. Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., Li, Z., Traugh, N., Bu, X., Li, B., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat. Med. 24, 1550–1558. https://doi.org/10.1038/s41591-018-0136-1.

76. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420. https://doi.org/10.1038/nbt.4096.

77. Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell 173, 338–354.e15. https://doi.org/10.1016/j.cell.2018.03.034.

78. Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell 141, 69–80. https://doi.org/10.1016/j.cell.2010.02.027.

79. Ramirez, M., Rajaram, S., Steininger, R.J., Osipchuk, D., Roth, M.A., Morinishi, L.S., Evans, L., Ji, W., Hsu, C.H., Thurley, K., et al. (2016). Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. Nat. Commun. 7, 10690. https://doi.org/10.1038/ncomms10690.

80. Roper, N., Brown, A.L., Wei, J.S., Pack, S., Trindade, C., Kim, C., Restifo, O., Gao, S., Sindiri, S., Mehrabadi, F., et al. (2020). Clonal evolution and heterogeneity of osimertinib acquired resistance mechanisms in EGFR mutant lung cancer. Cell Rep. Med. 1, 100007. https://doi.org/10.1016/j.xcrm.2020.100007.

81. Yang, X., Chen, L., Mao, Y., Hu, Z., and He, M. (2020). Progressive and prognostic performance of an extracellular matrix-receptor interaction signature in gastric cancer. Dis. Markers 2020, 8816070. https://doi.org/10.1155/2020/8816070.

82. Joshi, K., de Massy, M.R., Ismail, M., Reading, J.L., Uddin, I., Woolston, A., Hatipoglu, E., Oakes, T., Rosenthal, R., Peacock, T., et al. (2019). Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. Nat. Med. 25, 1549–1559. https://doi.org/10.1038/s41591-019-0592-2.

83. Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., Hackl, H., and Trajanoski, Z. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 18, 248–262. https://doi.org/10.1016/j.celrep.2016.12.019.

84. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 1, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.

85. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Multiple EGFR-TKIs treatment timepoints scRNA-seq data (See Figure 1)	Maynard et al., 2020[15]	PRJNA591860
PC9 cell line bulk RNA-seq data (See Figure 2A)	Gogleva et al., 2022[56]	GSE193258
PC9 cell line bulk RNA-seq data (See Figure 2B)	Gurule et al., 2021[57]	GSE165019
PC9 cell line bulk RNA-seq data (See Figure 2C)	Rusan et al., 2018[58]	GSE89127
PC9 cell line bulk RNA-seq data (See Figure 2D)	Hata et al., 2016[59]	GSE75602
PC9 cell line bulk RNA-seq data (See Figure 2E)	Raoof et al., 2019[60]	GSE114647
PC9 cell line bulk RNA-seq data (See Figure 2F)	\	GSE162045
PC9 cell line scRNA-seq data (See Figure S4)	Aissa et al., 2021[61]	GSE149383
Mice model bulk RNA-seq data (See Figure 3A)	Kita et al., 2019[16]	GSE130160
Mice model bulk RNA-seq data (See Figure 3B)	Marrocco et al., 2021[17]	GSE161584
EGFR-TKIs treatment patients bulk RNA-seq data (See Figure 3C)	Gurule et al., 2021[57]	GSE165019
EGFR-TKIs treatment patients bulk RNA-seq data (See Figure 3E)	Kita et al., 2019[16]	phs002001
OncoSG bulk RNA-seq data (See Figures 4A and 4B)	Chen et al., 2020[47]	http://www.cbioportal.org/study/summary?id=luad_OncoSG_2020
TCGA bulk RNA-seq data (See Figures 4C and 4D)	Hoadley et al., 2018[62]	http://www.cbioportal.org/study/summary?id=luad_tcga_pan_can_atlas_2018
GSE31210 bulk RNA-seq data (See Figures 4E and 4F)	Kohno et al.,2011[63]	GSE31210
Software and algorithms		
R package Seurat v3.0.0	Stuart et al., 2019[64]	https://github.com/satijalab/seurat
R package Monocle v2.26.0	Qiu et al., 2017[65]	http://bioconductor.org/packages/release/bioc/html/monocle.html
R package Gelnet v1.2.1	Sokolov et al., 2016[14]	https://cran.r-project.org/web/packages/gelnet/index.html
R 4.1.2	R Core Team, 2021[66]	https://www.R-project.org
R package fgsea v1.10.1	Korotkevich et al., 2019[67]	http://bioconductor.org/packages/release/bioc/html/fgsea.html
R package DESeq2 v1.38.3	Love et al., 2014[68]	http://bioconductor.org/packages/release/bioc/html/DESeq2.html
R package Survival v3.2.11	Therneau et al., 2021[69]	https://github.com/therneau/survival
R package ggplot2 v3.3.2	Wickham et al., 2016[70]	https://cran.r-project.org/web/packages/ggplot2/index.html
R package GSVA v1.36.3	Hanzelmann et al., 2013[71]	http://bioconductor.riken.jp/packages/3.0/bioc/html/GSVA.html
R package clusterProfiler v4.2.2	Wu et al., 2021[72]	http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
R package GseaVis v0.0.5	Zhang et al., 2022[73]	https://github.com/junjunlab/GseaVis
CellPhoneDB	Efremova et al., 2020[74]	https://github.com/Teichlab/cellphonedb
TIDE	Jiang et al., 2018[75]	http://tide.dfci.harvard.edu

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Prof. Chengzhi Zhou (Zhouzhouchengzhi@gird.cn).

Materials availability

This study did not generate new unique materials.

Data and code availability

This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.

Code is available from the lead contact upon request.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

This study did not generate new unique experimental model.

METHOD DETAILS

The maynard et al.'s single-cell RNA sequence (scRNA-seq) data

To screen for EGFR-TKIs resistant cells, the Maynard et al.'s scRNA-seq data[15] of multiple EGFR-TKIs treatment timepoints were obtained from Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) under the accession number PRJNA591860. Single-cell annotation files and EGFR-TKIs therapeutic information files were downloaded from Google Cloud Disk (https://drive.google.com/drive/folders/1sDzO0WOD4rnGC7QfTKwdcQTx3L36PFwX?usp=sharing). This dataset was chosen because it is currently the largest number of scRNA-seq samples with multiple EGFR-TKIs treatment time points and high sequencing quality, which provides us with the resources to investigate the properties of EGFR-TKIs resistant cancer cells. The cohort contains 14 advanced-stage NSCLC EGFR-TKIs treatment individual patients. According to the medication situation, patients were divided into three states: TN (patients before initiating systemic targeted therapy, TKI naive state), RD (tumor was regressing or stable by clinical imaging, residual disease state), and PD (subsequent progressive disease as determined by clinical imaging, progression state). This dataset sets the stage for our investigation into the mechanisms underlying differential EGFR-TKIs efficacy. A total of 23 small tissue samples or surgical resections samples were obtained from lung tissue. Smart-seq2 technology was used to extract the expression profile of single cells (Table S1, Figure S1).

Maynard et al.'s scRNA-seq data quality control and R-index model

To ensure the quality of data analysis, we performed preliminary quality control of the Maynard et al.'s scRNA data as follows: 1) Retention of single-cell samples from patients with EGFR mutations; 2) Removal of potentially double cells; 3) Removal of mitochondrial and ribosomal genes; 4) Ensure single cell nCount >50000 & nFeature >500. Finally, we obtained 2080 cancer cells from 23 samples belonging to three states, TN, RD, and PD (Figure S2A). The Seurat v3[76] R package was used for single-cell RNA-seq analysis.[64] First, the 2080 cancer cells were scaled and normalized according to default parameters, and 2000 hypervariable genes were selected for subsequent analysis (Figure S2B). principal component analysis (PCA) was used to reduce feature gene dimension and top 20 PCA (Figure S2C) was selected to execute Shared Nearest Neighbor (SNN) algorithm cluster with 0.5 resolution (Figure S2D). Then, a 2-dimensional t-distributed Stochastic Neighbor Embedding (tSNE) and Uni-form Mani-fold Approximation and Projection (UMAP) was used to visualize the distribution of cancer cells at three time points (Figure S3). Cancer cells at each time point were displayed with UMAP. Second, to identify dynamic gene expression patterns correlated with clinical efficacy. The cell lineage trajectory was inferred by Monocle2[65] following the tutorial. We manually inspected the trajectory and selected the root nodes from the principal points with earlier development stage (EGFR-TKIs naive, Branch 1) than its nearby neighbors (EGFR-TKIs treatment, Branch 2 and Branch 3). Finally, we used the DESeq2 R package[68] to derive DEG(Differential expressed genes)

from selected branches (PD cancer cells in branch 3, PDB3 vs RD cancer cells in branch 2, RDB2) with the p-value ≤0.01 and |log2FC| > 2, and got 1107 DEG (Table S2). To extract the expression characteristics of EGRF-TKIs resistant cells, a weighted 1107 genes signature array (Table S3) was yielded using OCLR[14] algorithm performed by gelnet v1.2.1 R package according to a previous study.[77] The calculation process is as follows: 1) We first extracted the 1107 genes expression information of PDB3. 2) We then use the gelnet function for 1000 iterations to calculate the weighted value of each gene. 3) The weighted values of 1107 genes were used as expression coefficients for EGFR-TKIs resistance. 4) R-index equals the spearman correlation value of the 1107 gene expression coefficient and corresponding gene expression value.

Cell line validation data

To validate the performance of R-index in EGFR-TKIs treatment cell lines, we obtained RNA-seq sequencing of PC9 data with pre- and post-EGFR-TKI treatment. PC9 is an NSCLC cell line with EGFR 19del. A previous study has been shown that a small fraction of PC9 cells (∼0.5%) can enter a persist state to evade the intense selective pressure of high concentrations of the EGFR inhibitor erlotinib.[78,79] We hypothesize that the cells that remain viable after EGFR-TKIs treatment will be resistant and that the corresponding R-index will be significantly higher. EGFR-TKIs treatment data of the PC9 cell line was obtained from Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo) under the accession number GSE149383,[61] GSE193258,[56] GSE165019,[57] GSE89127,[58] GSE75602,[59] GSE114647,[60] GSE162045. The GSE193258 PC9 cell lines were treated with 500 nM osimertinib for day 0 and day 21. The total RNA was extracted and performed RNA-seq. The GSE165019 PC9 cell lines were treated with 300 nM Osimertinib for 3 days and total RNA was extracted and performed RNA-seq. This dataset also contains 8 EGFRmutant patients treated with EGFR-TKIs (Osimertinib or Erlotinib), bulk RNA-seq was performed and progression information was recorded. The GSE89127 PC9 cell lines were treated with Erlotinib at day 0, day 2, day 7 and total RNA was extracted and performed RNA-seq. The GSE75602 PC9 cell lines were treated with 1 mM gefitinib and 1 mM WZ4002 for 24 h, respectively. The total RNA was extracted and performed RNA-seq. The GSE114647 PC9 cell lines were treated with 300 nM gefitinib at day 0, day 1 and day 14. The total RNA was extracted and performed RNA-seq. The GSE162045 PC9 cell lines were treated with gefitinib at day 0, day 3 and day 9. The total RNA was extracted and performed RNA-seq. The GSE149383 PC9 cell lines were treated with 2 µM erlotinib at day 0, day 1, day 2, day 4, day9, day 11 and single-cell RNA-seq was performed using Drop-seq.

Mice model validation data

To validate the performance of R-index in mice. Patient-derived NSCLC xenograft models bulk RNA-seq data were downloaded from the GEO database under the accession numberGSE161584,[17] GSE130160.[16] The GSE161584 mice model were injected with PC9 cells. When tumors reached 500 mm^3, mice were treated daily with erlotinib (50 mg/kg) until resistant (started relapsing or reached 800 mm^3). Then, the tumor tissue RNA was extracted and performed RNA-seq. The GSE130160 mice model was implanted with NSCLC tumor specimens (Exon 19 deletion and L858R mutation) and treated by oral gavage with 25 mg/kg per day of osimertinib when tumor volume exceeded 500 mm^3 until resistant (reached 1500 mm^3).

EGFR-TKIs treatment validation data

To validate the performance of R-index in the real-world clinical treatment of patients. The NSCLC patients with pre- and post-EGFR-TKIs treatment independent bulk RNA-seq data were also downloaded from the GEO database under the accession number GSE165019,[57] and the dbGaP database under the accession number phs002001.[80] The GSE165019 contains 8 EGFRmutant patients treated with EGFR-TKIs (Osimertinib or Erlotinib), bulk RNA-seq and progression information was obtained. The phs002001 contains 9 first line of EGFR-TKIs treatment patients, bulk RNA-seq and progression information was obtained.

Cohort validation data

To validate the performance of R-index in prognosis. The OncoSG (Singapore Oncology Data Portal),[81] TCGA[62] and GSE31210[63] bulk RNA-seq data with clinic information was obtained. In these cohorts, one patient had only one bulk RNA-seq data. The OncoSG cohort was downloaded from cbioportal at (http://www.cbioportal.org/study/summary?id=luad_OncoSG_2020). This cohort contains 169 lung adenocarcinoma patients and 94 patients had EGFR mutations. The TCGA cohort were downloaded from cbioportal at (http://www.cbioportal.org/study/summary?id=luad_tcga_pan_can_atlas_2018). This cohort contains

510 lung adenocarcinoma patients and 54 patients with EGFR mutations. The GSE31210 cohort was downloaded from the GEO database. This cohort contains 226 Japan early lung adenocarcinomas patients and 127 patients with EGFR mutations.

Cancer cell cluster diversity

We hypothesize that EGFR-TKIs interventions may have a screening effect on cancer cells, resulting in structural changes in the composition and distribution of cancer cells that may be assessed from a diversity perspective. We calculated the shannon entropy of cancer cells at the three time points to capture the contribution of each cancer cell cluster.[82] The shannon entropy index is given by

$$H - index = \frac{-\sum_{i=1}^{n} P_i log_2(P_i)}{ln(N)}$$

where p_i represents the relative abundance ratio of the ith cluster cell number and N is the total number of clusters. p_i is obtained by dividing the number of cancer cells belonging to the ith cluster by the total number of cancer cell numbers in the states. A large H-index indicates high diversity.

Survival analysis

In the bulk validation cohort OncoSG, TCGA, and GSE31210, Cox proportional hazard models were used to investigate the association between R-index and patient survival. The samples were grouped into high and low expression groups by the median value. The Kaplan-Meier survival curves were plotted to show differences in survival time, and log-rank p values reported by the Cox regression models implemented in the R package survival v3.2.11 were used to determine the statistical significance.[69]

fgsea analysis

Fgsea was used to explore the MSigDB v7.4 hallmark gene set properties of 1107 EGFR-TKIs resistant genes in Maynard et al.'s scRNA-seq data with fgsea v1.10.1 R package,[67] which had accurate standard approaches to multiple hypothesis correction, making more permutations, and getting more fine-grained p-values.

ssGSEA analysis

The ssGSEA algorithm[71] was used to quantify the relative abundance of 28 immune cell types[83] and 50 hallmark gene sets[84] with the GSVA v1.36.3 R package. The relative abundance value was represented by an enrichment score, which was normalized to unity distribution from zero to one.

Gene set enrichment analysis

Gene set enrichment analysis was performed with the R package clusterProfiler v4.2.2.[72] The KEGG/GO/Reactome gene set collection from the Molecular Signature Database[85] was used to identify the biological pathways. Pathways with adjusted p< 0.05 were included and visualized by GseaVis v0.0.5 R package.

Cell-cell interaction analysis

We used CellPhoneDB[74] to identify significant ligand-receptor pairs within cells from PD patient and cells from RD patient. CellPhoneDB is a Python program calculating the interaction between the receptors and ligands. The cell-type-specific receptor-ligand interactions among cell types were identified based on the specific expression of a receptor by one cell type and a ligand by another cell type. The interaction score refers to the total mean of the individual ligand-receptor partner average expression values in the corresponding interacting pairs of cell types. The expression of complexes output by CellPhoneDB was calculated as the sum of the expression of the component genes.

TIDE analysis

The T cell dysfunction and exclusion status was estimated with Tumor Immune Dysfunction and Exclusion (TIDE) algorithm[75] (http://tide.dfci.harvard.edu) using bulk transcriptome profiles of OncoSG, TCGA and GSE31210 cohort.

QUANTIFICATION AND STATISTICAL ANALYSIS

The differences between R-index median stratification groups were analyzed using Mann–Whitney U tests. The consistency between R-index and cell number was assessed using Spearman correlation analysis. ggplot2 was used for data visualization.[70] All statistical analyses and presentations were performed using R v4.1.2. Statistical significance was set at $p < 0.05$.