Identification of predictive biomarkers for ZD-6474 in lung cancer

Liang-Bo Wang[1], Eric Y. Chuang[1], Tzu-Pin Lu[2]

[1]Graduate Institute of Biomedical Electronics and Bioinformatics, [2]Department of Public Health, Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan

Contributions: (I) Conception and design: All authors; (II) Administrative support: All authors; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Tzu-Pin Lu. Department of Public Health, Institute of Epidemiology and Preventive Medicine, National Taiwan University, No. 17, Xu-Zhou Road, Taipei 10055, Taiwan. Email: tplu@ntu.edu.tw.

Background: Lung cancer is the leading cause of cancer-related death globally. Even though numerous research efforts have been devoted to improving the treatment of patients with lung cancer, the overall five-year survival rate is still below 20%. The major challenge in improving this survival rate is the highly heterogeneous cancer genome. Recently, targeted therapy, and especially drugs that target *EGFR*, has been shown to be a promising therapeutic method against lung cancer. Challenges arise, however, in trying to classify patients as responders or non-responders to the drugs. Patients who receive no benefits from the treatment may still suffer from its adverse effects. One way to address this issue is to identify genomic biomarkers that predict the responses of patients to drugs before treatment.

Methods: This study demonstrated the identification of predictive biomarkers for responses to drugs by analyzing the gene expression profiles of lung cancer cell lines. For each cell line, microarray data were normalized using the quantile algorithm. A linear regression model was applied to select probes that are associated with drug efficacy, and a prediction model was developed using the support vector machine algorithm.

Results: ZD-6474 has the strongest inhibitory effects of the three drugs that target *EGFR*, and a total of 24 probes displayed significant associations with the efficacy of ZD-6474 ($P<2\times10^{-5}$). The prediction model demonstrated approximately 80% accuracy in the leave-one out cross-validation test of 89 lung cancer cell lines.

Conclusions: In conclusion, gene expression profiles can serve as potential predictive biomarkers for predicting patients' responses to drugs so the treatment plans for different individuals may be improved by considering their genetic statuses.

Keywords: Predictive biomarkers; drug response; lung cancer; gene expression microarray

Introduction

Lung cancer is the leading cause of cancer-related death globally (1-3). With advances in biotechnology, several high-throughput methods, including microarray and next-generation sequencing (NGS), have been utilized to elucidate the etiology and molecular mechanisms of lung cancer (4-6). Although these studies have improved out understanding of lung cancer, the overall 5-year survival rate is still below 20% (3). This low survival rate is attributable to several causes, such as heterogeneity, difficulties in early diagnosis, and a high recurrence rate. The most important challenge in improving the survival rate is to identify patients who are susceptible to treatments. Traditional drugs and treatments are designed for all individuals with specific diseases, regardless of their clinical and genomic differences. However, recent data about the survival rates of patients with

cancer have shown that such "one size fits all" approach is ineffective. Patients can be classified into different subtypes, even if they are diagnosed with the same cancer, and the development of drugs and therapeutic methods should take intratumor heterogeneity into consideration (7).

In past decades, several targeted therapies have been developed for different cancers, and many of them have been approved by the Food Drug Administration (FDA) of the United States of America (USA) (8). The uniqueness of targeted therapy is that such treatment is designed to target malfunctioning molecules and pathways to improve effectiveness, rather than to attack tumor cells directly (9). In practice, targeted therapy requires the selection of a subgroup of patients who can benefit most from the therapy. Many genomic features such as single nucleotide polymorphisms (SNPs), copy number variations (CNVs) and gene expression profiles, have been used to identify subgroups of patients (4,5,10), and those features are called biomarkers. Biomarkers can be divided into two major types, which are prognostic biomarkers and predictive biomarkers (11). Prognostic biomarkers are predictors of overall outcomes for a patient with a specific disease, regardless of treatment procedure. For example, the MammaPrint utilizes the expression levels of 70 genes to predict the probability of recurrence of breast cancer and the survival outcomes of patients with it (12), and the Vysis UroVysion test predicts the recurrence probability of bladder cancer by examining CNVs using the multi-target fluorescence *in situ* hybridization (FISH) method (13). Predictive biomarkers are used to predict patients' responses to a particular treatment. For instance, Gefitinib, also known as Iressa, has been approved by the US-FDA for treating non-small-cell lung cancer (NSCLC) with exon 19 deletions or exon 21 L858R substitutions in *EGFR* (14). Trastuzumab, also known as Herceptin, provides significant clinical benefit for the subpopulation of breast cancer patients with HER2 amplification (15). In conclusion, relevant studies have indicated that genomic features are effective biomarkers for predicting overall survival outcomes and responses the of cancer patients to treatment.

The US-FDA defines personalized medicine as "*the right patient with the right drug at the right dose at the right time*" (16). However, identify the right patient with the right drug poses a major challenge. Ideally, a randomized clinical trial of each drug would be performed, and the genomic features of patients measured to identify the most susceptible populations. Yet, such an approach is difficult to execute for reasons of medical ethics and the cost. To address this issue, cancer cell lines may be used as study models and their

**Table 1** The 24 drugs studied in the cancer cell line encyclopedia (CCLE) dataset

Target gene	Drug name	Target gene	Drug name
ABL	AZD0530	*MDM2*	Nutlin-3
ABL	Nilotinib	*MEK*	AZD6244
ALK	TAE684	*MEK*	PD-0325901
CDK4	PD-0332991	*RAF*	PLX4720
EGFR	Erlotinib	*RAF*	RAF265
EGFR	Lapatinib	*RTK*	Sorafenib
EGFR	ZD-6474	*TOP1*	Irinotecan
FGFR	TKI258	*TOP1*	Topotecan
GS	L-685458	*TUBB1*	Paclitaxel
HDAC	Panobinostat	*XIAP*	LBW242
HSP90	17-AAG	*c-MET*	PF2341066
IGF1R	AEW541	*c-MET*	PHA-665752

responses to drugs and genomic changes measured. In this study, we analyzed the cancer cell line encyclopedia (CCLE) project (17) to demonstrate the possibility of identifying predictive biomarkers using gene expression data for lung cancer cell lines. A linear regression model was applied to the gene expression profiles of 89 lung cancer cell lines along with their drug efficacy data. Potential predictive biomarkers were identified for the drug that targets *EGFR*, ZD-6474, and a prediction model was developed using the support vector machine (SVM) algorithm. The prediction model had an accuracy of approximately 80%, based on the leave one out cross-validation test, suggesting the possibility that the results can be practically applied.

Methods

Microarray dataset

The gene expression microarrays in the CCLE project with the accession number GSE36133 (17) were retrieved from the Gene Expression Omnibus (GEO) (18), and the samples were analyzed using the Affymetrix U133 Plus 2.0 platform. Of the 475 cell lines for which microarrays were available, 89 were lung cancer. For each probe, the corresponding gene symbols were obtained according to the annotation file that was provided by Affymetrix. The CCLE dataset provides efficacy data for 24 drugs that target 17 distinct genes (*Table 1*). Drug efficacy was measured by computing the area under the cumulative curve of the surviving fraction after treatments of drugs in different concentrations, which is called the activity

326

Wang et al. Predictive biomarkers for ZD-6474 in lung cancer
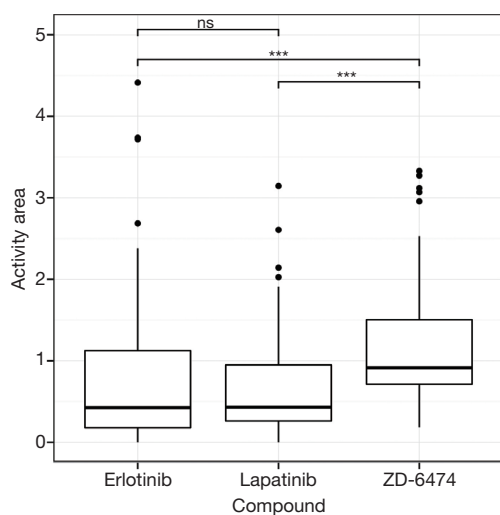
Figure 1 The measurement of drug efficacy in the three drugs targeting *EGFR*. The y-axis represents the activity area, which represents the inhibitory effects in tumor cell growth. The Wilcoxon rank sum test was performed to assess the differences between the three drugs. ***, $P<10^{-8}$. ns, $P>0.05$.
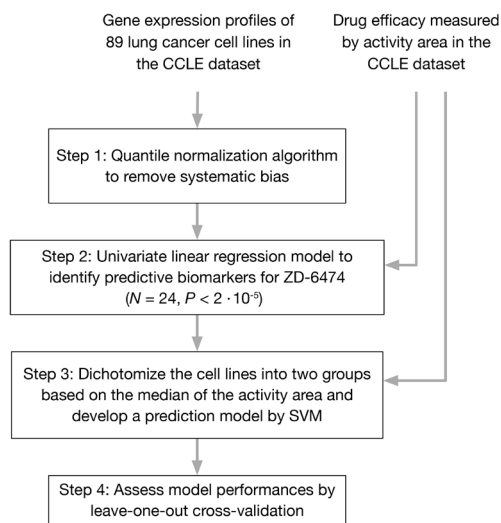
Figure 2 Flowchart for identification of predictive biomarkers and development of a prediction model. CCLE, cancer cell line encyclopedia; SVM, support vector machine.

area. For each drug, a larger activity area indicates a stronger suppression of tumor cell growth.

Identification of predictive biomarkers for ZD-6474

In the CCLE dataset, the highest number of cell lines

(N=89) were lung cancer, on which further analyses therefore focused. Among the 24 drugs in *Table 1*, the three that target *EGFR* (Erlotinib, Lapatinib and ZD-6474) were selected for advanced comparison. The Wilcoxon rank sum tests were performed on the activity area for the three drugs in the 89 lung cancer cell lines (*Figure 1*), and revealed that ZD-6474 has the highest inhibitory effect. Therefore, ZD-6474 was the drug target in this study.

Figure 2 presents a protocol for identifying predictive biomarkers for ZD-6474. The raw CEL files of the gene expression microarrays were imported into the Partek Genomic Suite and the quantile normalization algorithm was utilized to remove systematic biases (step 1). A linear regression model was used to select probes that were associated with the efficacy of ZD-6474 (step 2). A total of 24 significant probes were identified ($P<2\times10^{-5}$) and utilized to develop a prediction model for the efficacy of ZD-6474 using the SVM algorithm (19) (step 3). The 89 cell lines were split into two groups based on the median activity area of ZD-6474.

Assessment of performance of prediction model

To evaluate the predictive performance of the model, a leave-one out cross-validation (LOOCV) test was performed (step 4). The 89 lung cancer cell lines were divided into a training group with 88 samples and a testing group with only one sample. The procedure was conducted 89 times until a prediction had been made for every cell line. The quality of prediction was captured using the following indexes; sensitivity, specificity, positive prediction value (PPV), negative prediction value (NPV) and accuracy (*Table 2*). Lastly, a resampling test was performed to ensure that the 24 predictive biomarkers were not identified by chance. A null baseline was developed by randomly selecting 24 probes from the original pool of probes (N=54,675) and a prediction model was developed and evaluated using the procedure (steps 3-4) in *Figure 2*. The empirical P value was determined by comparing the indexes with the null baseline; that is, the ranking of the indexes.

Results

Gene expression profiles and drug responses according to the CCLE dataset

The microarray data in the CCLE dataset (GSE36133) were retrieved from the GEO database (17), and the

Table 2 The prediction performances in our model and the resampling test. The resampling test was repeated for 10,000 trials

Feature	Accuracy	Sensitivity	Specificity	PPV	NPV
Our model	0.820	0.800	0.841	0.837	0.804
Resampling test[a]	0.500	0.508	0.491	0.505	0.494
Empirical P	<0.0001	0.0004	<0.0001	<0.0001	<0.0001

PPV, positive prediction value; NPV, negative prediction value; [a], The mean values of the 10,000 trials are shown here.

gene expression profiles of 475 cell lines in 27 distinct primary sites were examined using the Affymetrix u133plus 2.0 platform (*Table S1*). More of the cell lines in the CCLE dataset were lung cancer (N=89) than were other cancers, so these 89 cell lines were the target of further analyses. In addition to gene expression profiles, the CCLE dataset includes measurements of the efficacy of 24 drugs, including activity area, IC-50, and EC-50. The activity area was selected to represent the response of each drug because fewer activity area data for the 89 lung cancer cell lines were missing than other data. Three of the 24 drugs (Erlotinib, Lapatinib and ZD-6474) were designed to target *EGFR*. Since *EGFR* has been shown to be an important player and a critical therapeutic target in lung cancer (20,21), the following analyses focus on these three drugs, whose activity areas are presented in *Figure 2*. The Wilcoxon rank sum test was performed to determine whether their efficacy in lung cancer cell lines varied. The results thus obtained showed that ZD-6474 had significantly stronger inhibition effects than the other two drugs (P<0.05), suggesting that ZD-6474 may perform best in killing lung cancer cells in these three drugs.

Identification of predictive biomarkers ZD-6474

To identify potential biomarkers that are associated with the drug efficacy of ZD-6474 in the 89 lung cancer cell lines, a linear regression model was applied to the gene expression level for each gene. A total of 24 probes with significant associations were identified (P<2×10^{-5}) and summarized in *Table 3*. Notably, multiple probes with the same gene symbol were identified, suggesting the consistency and reproducibility of the identified genes. With respect to their functional significances, several studies have shown their importance in lung cancer (25,30,39). For example, three probes were located in *KIAA0494*, which has been reported to be differentially expressed between *EGFR/KRAS* groups in at least four of five lung adenocarcinoma cohorts (25). Moreover, *PDGF* is a platelet-derived growth factor and

its overexpression or mutation is able to drive the growth of cancer cells (39). Intriguingly, the expressions of *PDGF* and *PDGF* receptors have been reportedly associated with poor prognosis in lung cancer patients (30). Therefore, these identified potential predictive biomarkers may play important roles in regulating the biological functions and the cellular responses of ZD-6474.

Performance of prediction model for ZD-6474 in lung cancer cell lines

To evaluate the prediction performance of those identified biomarkers, a prediction model was developed using the SVM algorithm. The LOOCV procedures were performed on the 89 lung cancer cell lines and the results are summarized in *Table 2*. The prediction model was 82% accurate in the LOOCV test, suggesting its effectiveness. Notably, the indexes of sensitivity, specificity, PPV and NPV were all higher than 0.8, indicating that the prediction model was a general model without a systematic bias toward responding or non-responding groups of ZD-6474. Lastly, the prediction model was compared with the null baselines that were developed using the resampling procedure. As shown in *Table 2*, the empirical p-values for the five indexes were significant (P<5×10^{-4}), showing that the 24 predictive biomarkers were not identified by chance. In conclusion, the prediction model is effective in predicting the efficacy of ZD-6474 using gene expression profiles.

Discussion

In past decades, a major breakthrough in the biomedical research field has been the completion of the human genome project (HGP), which cost a total of US $3 billion and took more than ten years. Its extremely high price and long running time prohibit the whole-genome sequencing for an individual. Recent advances in high-throughput experimental methods have enabled researchers simultaneously to investigate gene expression

Table 3 Identified 24 predictive biomarkers showing significant associations with the efficacy of ZD-6474

Gene Symbol	Probe	P value	Related studies
CTNND1	1557944_s_at	1.32E-05	Involved in the *EGFR*/Src-P pathway (22-24)
CYSLTR1	216288_at	4.87E-08	
EPS15L1	231926_at	7.18E-06	
ERP27	227450_at	8.8E-06	
INPP4B	223878_at	2.7E-07	
KCNK1	1566465_at	8.56E-06	
KIAA0494	201775_s_at	1.61E-06	Differentially expressed between *EGFR/KRAS* groups in lung
	201776_s_at	1.51E-05	adenocarcinoma (25)
	201778_s_at	1.35E-05	
KLF8	230986_at	1.39E-05	A member of the Sp/KLF family of transcription factors (26)
MAL	204777_s_at	2.27E-06	
MAP3K1	214786_at	2.65E-08	Identified based on the survival analysis of SNPs in the NSCLC samples (27)
PADI1	223739_at	8.59E-06	Significant overexpression of PADI4 is observed in NSCLC (28,29)
PDGFB	204200_s_at	4.11E-06	Associated with poor prognosis in lung cancer patients (30-34)
PDGFB	216061_x_at	3.12E-06	
PRRG4	207291_at	2.74E-06	
PRRG4	238513_at	2.54E-06	
RASEF	243862_at	9.06E-06	A novel tissue biomarker and a therapeutic target for lung cancer (35,36)
RNF144B	235549_at	7.26E-06	Mediating degradation of p21^{WAF1}, which is associated with the treatment of Gefitinib in NSCLC (37,38)
TNKS1BP1	224792_at	1.3E-05	
USP30	227572_at	1.38E-05	
VGLL1	205487_s_at	1.55E-05	
–	241301_at	1.59E-05	
–	244116_at	4.86E-07	

NSCLC, non-small cell lung cancer.

profiles and to characterize genetic changes in dysregulated genomes that are affected by different diseases (4-6). Since many diseases, such as cancers, are highly heterogeneous, biotechnological progress has helped to shed light on how differences among individuals can be considered in planning treatment. Consequently, personalized medicine has become an important issue not only for drug companies but also for the regulatory agencies of governments. More targeted therapies and *in vitro* diagnostic tests (IVDs) have recently been approved by the US-FDA (8,16), suggesting the importance of using genomic features in classifying patients based on disease status and response to treatment.

A well-known, major challenge in the development of drugs is the high failure rate in several clinical trials—especially for drugs for treating cancers. As described above, these low success rates may be attributable to heterogeneous genomes. Large variations among individuals may cause studied drugs to be effective only for a small proportion of examined patients, making their effectiveness non-significant across the whole population. However, effectiveness may become significant if responders can be identified from their genomic features. Accordingly, the *adaptive design clinical study* is relevant; it has been officially defined by the US-FDA as "*a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study*" (40). Higher success rates in the development of drugs can thus be obtained using adaptively designed clinical trials, which provide a better understanding of the effects of treatment

than conventional trials. Therefore, for each drug, effectively identifying predictive biomarkers to determine the response of a patient has become an important task.

The best procedure for identifying predictive biomarkers for a drug is to perform a randomized clinical trial and a genome-wide genetic study simultaneously. However, such a design is problematic ethically, and so difficult to execute in practice. This issue can be addressed in two ways. The first is to perform retrospective analyses of clinical trials, which take a long time to accumulate the data. Alternatively, cancer cell lines may serve as a good model to identify potential predictive biomarkers owing to their ease of manipulations. This study demonstrates the potential of analyzing lung cancer cell lines to select genes that are significantly associated with drug efficacy (*Table 3*). Several studies have also shown the importance of identified genes in regulating biological functions and signaling pathways in response to the drug treatment (25,30,39). These results suggest that analyzing the gene expression profiles of cancer lines to identify predictive biomarkers for drug efficacy is feasible.

This study has some limitations. Although the favorable performance of the prediction model was verified by using the LOOCV test, further analyses of external datasets are required to demonstrate reproducibility. Lung cancer cell lines were grouped by the median efficacy (activity area) of ZD-6474. However, the drug efficacy is a continuous variable so this grouping may introduce biases into the prediction model. Lastly, even though approximately 80% accuracy was achieved in the cell lines herein, the results of real clinical trials can further demonstrate the feasibility of applying the efficacy model of ZD-6474).

ZD-6474, also known as Vandetanib, was selected as the prediction target in this study because ZD-6474 had the strongest inhibitory effects of the three drugs herein that target *EGFR*. Although ZD-6474 has not been approved to treat NSCLC patients, it has been used to treat medullary thyroid cancer (41,42). A large-scale meta-analysis of 14 clinical trials demonstrated that NSCLC patients who received ZD-6474 had better progression-free survival (43), suggesting the possibility of using ZD-6474 in lung cancer. Subsequently, to clarify whether different drugs that target the same gene have similar predictive biomarkers, linear regression models were utilized in Erlotinib and Lapatinib (*Figure S1*). Only one probe exhibited significant associations with the efficacies of all three drugs, suggesting that these drugs have distinct predictive biomarkers. Therefore, the procedures for identifying predictive biomarkers should be performed individually for each drug, even if they have the same target gene.

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editors (Lyudmila Bazhenova and Ajay Pal Singh Sandhu) for the series "Recent advances in radiotherapy and targeted therapies for lung cancer" published in *Translational Cancer Research*. The article has undergone external peer review.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.3978/j.issn.2218-676X.2015.08.10). The series "Recent advances in radiotherapy and targeted therapies for lung cancer" was commissioned by the editorial office without any funding or sponsorship. EYC serves as the Editor-in-Chief of *Translational Cancer Research*. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Neither IRB permission nor informed consent is required since the data involved is from public sources.

References

1. Malvezzi M, Bertuccio P, Levi F, et al. European cancer mortality predictions for the year 2013. Ann Oncol 2013;24:792-800.
2. Molina JR, Yang P, Cassivi SD, et al. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin Proc 2008;83:584-94.
3. Siegel R, Ma J, Zou Z, et al. Cancer statistics, 2014. CA Cancer J Clin 2014;64:9-29.
4. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816-24.
5. Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma, Shedden K, Taylor JM, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nat Med 2008;14:822-7.
6. Lu TP, Tsai MH, Lee JM, et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. Cancer Epidemiol Biomarkers Prev 2010;19:2590-7.
7. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. Cancer Cell 2015;27:15-26.
8. National Cancer Institute. Targeted Cancer Therapies. 2015. Available online: http://www.cancer.gov/about-cancer/treatment/types/targeted-therapies/targeted-therapies-fact-sheet#q8
9. Sawyers C. Targeted cancer therapy. Nature 2004;432:294-7.
10. Lu TP, Lai LC, Tsai MH, et al. Integrated analyses of copy number variations and gene expression in lung adenocarcinoma. PLoS One 2011;6:e24829.
11. Chen JJ, Lu TP, Chen DT, et al. Biomarker adaptive designs in clinical trials. Transl Cancer Res 2014;3:279-92.
12. Slodkowska EA, Ross JS. MammaPrint 70-gene signature: another milestone in personalized medical care for breast cancer patients. Expert Rev Mol Diagn 2009;9:417-22.
13. Zellweger T, Benz G, Cathomas G, et al. Multi-target fluorescence in situ hybridization in bladder washings for prediction of recurrent bladder cancer. Int J Cancer 2006;119:1660-5.
14. Gefitinib Approved for EGFR-Mutated NSCLC. Cancer Discov 2015. [Epub ahead of print].
15. Gajria D, Chandarlapaty S. HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies. Expert Rev Anticancer Ther 2011;11:263-75.
16. U.S. Food and Drug Administration. Personalized Medicine. 2015. Available online: http://www.fda.gov/scienceresearch/specialtopics/personalizedmedicine/default.htm
17. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483:603-7.
18. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207-10.
19. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2011;2.
20. Carrera S, Buque A, Azkona E, et al. Epidermal growth factor receptor tyrosine-kinase inhibitor treatment resistance in non-small cell lung cancer: biological basis and therapeutic strategies. Clin Transl Oncol 2014;16:339-50.
21. Mok TS, Lee K, Leung L. Targeting epidermal growth factor receptor in the management of lung cancer. Semin Oncol 2014;41:101-9.
22. Li J, Bennett K, Stukalov A, et al. Perturbation of the mutated EGFR interactome identifies vulnerabilities and resistance mechanisms. Mol Syst Biol 2013;9:705.
23. Zhang L, Gallup M, Zlock L, et al. p120-catenin modulates airway epithelial cell migration induced by cigarette smoke. Biochem Biophys Res Commun 2012;417:49-55.
24. Zhang Y, Zhao Y, Jiang G, et al. Impact of p120-catenin isoforms 1A and 3A on epithelial mesenchymal transition of lung cancer cells expressing E-cadherin in different subcellular locations. PLoS One 2014;9:e88064.
25. Planck M, Edlund K, Botling J, et al. Genomic and transcriptional alterations in lung adenocarcinoma in relation to EGFR and KRAS mutation status. PLoS One 2013;8:e78614.
26. Wang X, Zhao J. KLF8 transcription factor participates in oncogenic transformation. Oncogene 2007;26:456-61.
27. Li Y, Sun Z, Cunningham JM, et al. Genetic variations in multiple drug action pathways and survival in advanced stage non-small cell lung cancer treated with chemotherapy. Clin Cancer Res 2011;17:3830-40.
28. O'Byrne KJ, Barr MP, Gray SG. The role of epigenetics in resistance to Cisplatin chemotherapy in lung cancer. Cancers (Basel) 2011;3:1426-53.
29. Ulivi P, Mercatali L, Casoni GL, et al. Multiple marker detection in peripheral blood for NSCLC diagnosis. PLoS One 2013;8:e57401.
30. Donnem T, Al-Saad S, Al-Shibli K, et al. Prognostic impact of platelet-derived growth factors in non-small

cell lung cancer tumor and stromal cells. J Thorac Oncol 2008;3:963-70.

31. Donnem T, Al-Saad S, Al-Shibli K, et al. Co-expression of PDGF-B and VEGFR-3 strongly correlates with lymph node metastasis and poor survival in non-small-cell lung cancer. Ann Oncol 2010;21:223-31.

32. Lazar V, Suo C, Orear C, et al. Integrated molecular portrait of non-small cell lung cancers. BMC Med Genomics 2013;6:53.

33. Navab R, Strumpf D, Bandarchi B, et al. Prognostic gene-expression signature of carcinoma-associated fibroblasts in non-small cell lung cancer. Proc Natl Acad Sci U S A 2011;108:7160-5.

34. Yatabe Y, Takahashi T, Mitsudomi T. Epidermal growth factor receptor gene amplification is acquired in association with tumor progression of EGFR-mutated lung cancer. Cancer Res 2008;68:2106-11.

35. Daigo Y, Takano A, Nakamura Y. Abstract 1765: Ras and EF-hand domain containing as a novel tissue biomarker and a therapeutic target for lung cancer. Cancer Res 2014;74:1765.

36. Oshita H, Nishino R, Takano A, et al. RASEF is a novel diagnostic biomarker and a therapeutic target for lung

cancer. Mol Cancer Res 2013;11:937-51.

37. Conforti F, Yang AL, Piro MC, et al. PIR2/Rnf144B regulates epithelial homeostasis by mediating degradation of p21WAF1 and p63. Oncogene 2013;32:4758-65.

38. Zhao YF, Wang CR, Wu YM, et al. P21 (waf1/cip1) is required for non-small cell lung cancer sensitive to Gefitinib treatment. Biomed Pharmacother 2011;65:151-6.

39. Heldin CH. Targeting the PDGF signaling pathway in tumor treatment. Cell Commun Signal 2013;11:97.

40. U.S. Food and Drug Administration. Guidance for Industry. Adaptive Design Clinical Trials for Drugs and Biologics. 2010. Available online: http://www.fda.gov/downloads/drugs/.../guidances/ucm201790.Pdf

41. Chu CT, Sada YH, Kim ES. Vandetanib for the treatment of lung cancer. Expert Opin Investig Drugs 2012;21:1211-21.

42. Leboulleux S, Bastholt L, Krause T, et al. Vandetanib in locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 2 trial. Lancet Oncol 2012;13:897-905.

43. Wu X, Jin Y, Cui IH, et al. Addition of vandetanib to chemotherapy in advanced solid cancers: a meta-analysis. Anticancer Drugs 2012;23:731-8.

Table S1 The 475 cell lines studied by the gene expression microarrays in the cancer cell line encyclopedia (CCLE) dataset

Primary site	Sample number	Primary site	Sample number
Autonomic ganglia	10	Ovary	26
Biliary tract	1	Pancreas	28
Bone	10	Pleura	7
Breast	28	Prostate	3
Central nervous system	26	Salivary gland	1
Endometrium	20	Sf8657	0
Haematopoietic and lymphoid tissue	67	Skin	39
Kidney	9	Snuc2b	0
Large intestine	22	Soft tissue	11
Liver	19	Stomach	18
Lung	89	Thyroid	5
M059j	0	Upper aerodigestive tract	7
Nervous system	1	Urinary tract	13
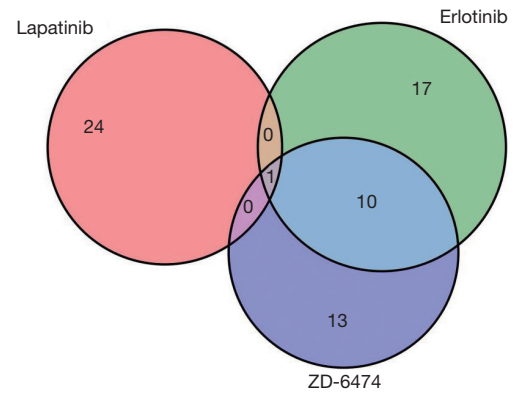
Figure S1 Venn diagram for identified predictive biomarkers for the three drugs targeting *EGFR*. Different P value thresholds were used for the three drugs. Lapatinib, 5×10^{-7}; Erlotinib, 5×10^{-6}; ZD-6474, 2×10^{-5}.