

Keywords: predictive signature; relative expression orderings; bioinformatics; clinical application; platinum-based adjuvant chemotherapy; lung adenocarcinoma

An individualised signature for predicting response with concordant survival benefit for lung adenocarcinoma patients receiving platinum-based chemotherapy

Lishuang Qi^{1,3}, Yang Li^{1,3}, Yuan Qin¹, Gengen Shi¹, Tianhao Li¹, Jiasheng Wang¹, Libin Chen¹, Yunyan Gu¹, Wenyuan Zhao^{*,1} and Zheng Guo^{*,1,2}

¹College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China and ²Key Laboratory of Ministry of Education for Gastrointestinal Cancer, Department of Bioinformatics, Fujian Medical University, Fuzhou 350001, China

Background: For lung adenocarcinoma (LUAD) patients receiving platinum-based adjuvant chemotherapy (ACT), predictive signatures extracted from survival data solely are not directly associated with platinum response. Another limitation of reported signatures, commonly based on risk scores summarised from gene expressions, is that they could not be applied directly to samples measured by different laboratories due to experimental batch effects.

Methods: Using 60 samples of LUAD patients receiving platinum-based ACT in TCGA, we pre-selected gene pairs whose within-samples relative expression orderings (REOs) were significantly associated with both pathological response and 5-year survival, from which we selected an optimal signature whose within-samples REOs could identify responders with improved 5-year survival rate.

Results: A predictive signature consisting of three gene pairs was developed. In an independent data set integrated from five small data sets, the predicted responders had a significantly higher 5-year survival rate than the predicted non-responders if and only if they received platinum-based ACT (log-rank $P=0.0006$). The predicted responders showed a 22% absolute benefit of platinum-based ACT in 5-year survival rate compared with untreated patients (log-rank $P=0.0019$).

Conclusions: The REO-based signature can individually predict response to platinum-based ACT with concordant survival benefit directly for LUAD samples measured by different laboratories.

Lung adenocarcinoma (LUAD) is the most common type of non-small-cell lung cancer with the greatest mortality (Siegel *et al*, 2015). Stage I patients usually receive surgical resection only because platinum-based adjuvant chemotherapy (ACT) has not shown to be beneficial for IA patients (Crino *et al*, 2015) and controversial for IB patients (Tsuboi *et al*, 2007). Stage II–IIIa patients usually receive platinum-based ACT after surgical resection, but only 4–15% survival benefit after ACT has been observed (Wallerek and

Sorensen, 2015). Heterogeneity of response to platinum-based ACT significantly confounds treatment of LUAD patients. Therefore, it is necessary to develop a clinically feasible signature to distinguish patients who might benefit from platinum-based ACT and those who should be spared the side effects of unnecessary treatment (Subramanian and Simon, 2010).

Recently, considerable efforts have been devoted to developing predictive transcriptional signatures for platinum-based ACT.

*Correspondence: Dr W Zhao; E-mail: zhaowenyuan@ems.hrbmu.edu.cn or Professor Z Guo; E-mail: guoz@ems.hrbmu.edu.cn

³These authors contributed equally to this work.

Received 26 September 2016; accepted 12 October 2016; published online 17 November 2016

© 2016 Cancer Research UK. All rights reserved 0007–0920/16

Most researchers firstly developed prognostic signatures for patients not receiving ACT and then demonstrated that only the high-risk patients predicted by the signatures showed significant survival benefit after platinum-based ACT (Zhu *et al*, 2010; Chen *et al*, 2011; Tang *et al*, 2013). Obviously, such prognostic signatures do not directly associate with patients' response to platinum-based ACT. Different from this strategy, Ryan (Van Laar, 2012) developed a 37-gene signature for distinguishing patients with shorter and longer survival after receiving platinum-based ACT and then defined them as non-responders and responders, respectively. However, the responders (or non-responders) predicted by the signature might include patients resistant (or sensitive) to platinum-based ACT but the tumour cells had a low (or high) degree of malignancy (Earl *et al*, 2015). Therefore, to increase the relevance of signatures to platinum-based ACT, patient pathological response should be utilised. However, according to current RECIST criteria, a certain percentage of pathological response states of LUAD patients may be misclassified by the conventional iconographies especially near the cutoff points for the short-term reduction of tumour size after chemotherapy (Gruber *et al*, 2013; Wu and Wang, 2015). This misclassification may blur the survival difference between pathological responders and non-responders after platinum-based ACT. Thus, we propose to use survival information to complement pathological response states to identify patients responsive to platinum-based ACT with survival improvement.

Another limitation of current transcriptional signatures is that they stratify patients based on risk scores summarised from the signature gene expression measurements (Zhu *et al*, 2010; Chen *et al*, 2011; Van Laar, 2012; Tang *et al*, 2013). However, as demonstrated in our recent work (Qi *et al*, 2016), the risk-score based signatures tend to be impractical for clinical settings because their application requires the pre-collection of samples for data normalisation to remove the experimental batch effects, whereas sample risk classification will be influenced by the risk composition of other samples adopted for normalisation. In contrast, the relative expression orderings (REOs) of gene pairs within a sample have been reported to be robust against to experimental batch effects (Wang *et al*, 2015; Qi *et al*, 2016) and variations of probe designs used in different platforms (Guan *et al*, 2016), rendering them promising for developing robust gene pair signatures (GPSs) (Eddy *et al*, 2010; Zhao *et al*, 2016). In particular, the small sample sizes of LUAD patients receiving platinum-based ACT included in the publically available data sets deter the development and validation of a robust signature. Most data integration approaches are limited by the technical problem that batch effects widely exist among data sets produced by different laboratories (Leek *et al*, 2010). In contrast, different data sets could be directly integrated based on the within-samples REOs (Geman *et al*, 2004).

In this study, we combined the pathological response states and 5-year survival data to extract a REO-based signature for predicting platinum responders with improved 5-year survival rate after receiving platinum-based ACT. The signature was tested in an integrated independent data set and compared with three signatures developed based on patients' survival solely.

MATERIALS AND METHODS

Data sources and data pre-processing. All gene expression profiling data of LUAD tissues, as described in Table 1 and Supplementary Table S1, were downloaded from TCGA (<http://cancergenome.nih.gov/>) and the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>). For TCGA data (The Cancer Genome Atlas Research Network, 2014), the normalised count values of level 3 gene expression data derived from Illumina

Table 1. Data sets analysed in this study

GEO ID(s) or data source	ACT	non-ACT	Platform
TCGA-LUAD (The Cancer Genome Atlas Research Network, 2014)	60	—	Illumina HiSeq V2
GSE42127 (Tang <i>et al</i> , 2013)	38	61	Illumina WG-6 V3
GSE14814 (Zhu <i>et al</i> , 2010)	31	26	Affymetrix U133A
GSE29013 (Xie <i>et al</i> , 2011)	19	9	Affymetrix Plus 2
GSE68465 (Shedden <i>et al</i> , 2008)	18	117	Affymetrix U133A
GSE37745 (Botling <i>et al</i> , 2013)	13	29	Affymetrix Plus 2

Abbreviations: ACT = patients receiving adjuvant chemotherapy; LUAD = lung adenocarcinoma; non-ACT = patients not receiving adjuvant chemotherapy.

HiSeqV2 were extracted as gene expression measurements. The RNA-seq expression profiles of 60 patients receiving cisplatin or carboplatin ACT, including 43 pathological complete response (pCR) and 17 pathological non-response (pNR: stable disease or progressive disease) patients recorded with overall survival data, were used to extract a REO-based predictive signature. We found no other publically available data sets providing patients' pathological response states for ACT. Therefore, we collected five gene expression data sets with the overall survival data of patients receiving ACT to test the predictive ability of the REO-based signature, based on the assumption that platinum response states would affect survival of patients receiving ACT. The five data sets were GSE42127 (Tang *et al*, 2013), GSE14814 (Zhu *et al*, 2010), GSE29013 (Xie *et al*, 2011), GSE68465 (Shedden *et al*, 2008) and GSE37745 (Botling *et al*, 2013). Notably, 42 LUAD samples were included in both the GSE14814 and GSE68465 data sets. After excluding the repeated 42 LUAD samples, 119 stage IB–III patients receiving ACT and 242 stage IB–III patients not receiving ACT were obtained, which were used as the testing ACT and non-ACT groups, respectively (Supplementary Data SA). The data set of 59 cell lines of various cancer types with gene expression profiles and growth inhibition of 50% (GI₅₀) data of cisplatin and carboplatin, respectively, were downloaded from the NCI60 (<http://dtp.nci.nih.gov/>; Supplementary Table S2).

For data generated by the Affymetrix platforms, the Robust Multi-array Average algorithm (Irizarry *et al*, 2003) was used for preprocessing the raw data. For a data set generated by the Illumina microarray platform, the originally processed data were used. All gene expression measurements were log₂ transformed. The Entrez IDs were used to map genes across microarray platforms.

Differential expression and survival analyses. The RankProd algorithm (Hong *et al*, 2006) was used to identify differentially expressed genes (DE genes) between two response groups. The *P*-values were adjusted using the Benjamini–Hochberg procedure for multiple testing to control the false discovery rate (FDR; Benjamini *et al*, 1995).

The 5-year survival rate of patients was used as the end point of interest. Patients with more than 5 years follow-up were censored at 5 years because deaths occurring past five years were not likely to be related to ACT. A multivariate Cox proportional-hazards regression model was used to evaluate independent associations between predictive factors and patient survival after adjusting for stage, age and gender. We adopted the concordance index (C-index; Harrell *et al*, 1996) to estimate the predictive performance of a signature for patient survival. Survival curves were estimated using the Kaplan–Meier method and were compared using the log-rank test (Bland and Altman, 2004).

Developing the predictive GPS for platinum-based ACT. First, for a pair of genes, a and b, with expression values G_a and G_b,

respectively, including at least one of the DE genes between pCR and pNR groups, we used Fisher's exact test to evaluate whether the frequency of pCR samples with a specific REOs pattern ($G_a > G_b$) was significantly higher than that in pNR samples. The gene pairs detected with $P < 0.01$ were defined as response-related gene pairs. Then, for each of the response-related gene pairs, we estimated the association between its specific REO and patient survival by performing the multivariate Cox regression models, and its predictive performance by calculating the C-index values. Gene pairs identified as $P < 0.01$ and C-index > 0.75 were selected as candidates to develop a predictive GPS. For a candidate gene pair, the specific REO pattern ($G_a > G_b$) in a cancer sample voted for response and the other REO pattern ($G_a < G_b$) voted for non-response. Finally, based on the candidate gene pairs, we applied a forward selection procedure to search a signature that achieved the largest C-index value for predicting patient survival. We chose each candidate gene pair as a seed and added another candidate gene pair to a set one at a time until the C-index did not increase. The majority voting rule was adopted as follows: a sample was a responder or non-responder if more than half of the REOs of gene pairs in the set voted for platinum response or non-response. Among the results derived from all seeds, a set of gene pairs with the largest C-index was chosen as a predictive GPS for platinum-based ACT.

Clustering and functional enrichment analyses. The K-means clustering algorithm was used to stratify patients into two subgroups ($k=2$). Sample similarity was estimated by the Euclidean distance based on the expression measurements of interesting genes.

The functional terms for enrichment analysis were downloaded from Gene Ontology (GO) in November 2015. The hypergeometric distribution model was used to test whether a set of genes observed in a term was significantly more than what expected by random chance. All statistical analyses were performed using the R2.15.3.

RESULTS

Identification of a predictive GPS for platinum-based ACT.

Figure 1 describes the flowchart of this study. For the 60 LUAD patients receiving platinum-based ACT in TCGA, we found no significant difference in 5-year survival rate between the 43 pCR and 17 pNR patients (log-rank $P=0.3250$; Supplementary Figure S1A). We suspected that some patients might be misclassified according to RECIST criteria and thus confound the survival difference between the true responders and non-responders. Nevertheless, the biological differences between the real responders and non-responders, especially those large differences, could be identified by comparing the two pathological response groups classified according to the conventional iconographies, given that a majority of samples were classified correctly. Therefore, we combined the data of pathological response and overall survival of patients receiving platinum-based ACT to select gene pairs that were related with both pathological response and overall survival.

Firstly, using the RankProd algorithm with 5% FDR control, we extracted 1352 DE genes between the 43 pCR and the 17 pNR patients. Then, from all the gene pairs consisting of DE genes, we extracted 24 305 response-related gene pairs, whose specific REOs patterns ($G_a > G_b$) occurred more frequently in the pCR samples than in the pNR samples (Fisher's exact test, $P < 0.01$). From these gene pairs, we further extracted 12 candidate gene pairs whose specific REOs were significantly correlated with improved 5-year survival of patients (multivariate Cox regression model, $P < 0.01$) with C-index values above 0.75. Using each of the 12 gene pairs as a seed, we obtained 12 sets of gene pairs using a forward selection procedure, among which a set of three gene pairs derived from the seed of the *LOC81691-ZMYND10* pair reached the largest C-index of 0.84 based on the majority voting rule. Thus, the three gene pairs were selected as the predictive signature for platinum-based ACT, denoted as 3-GPS (Table 2).

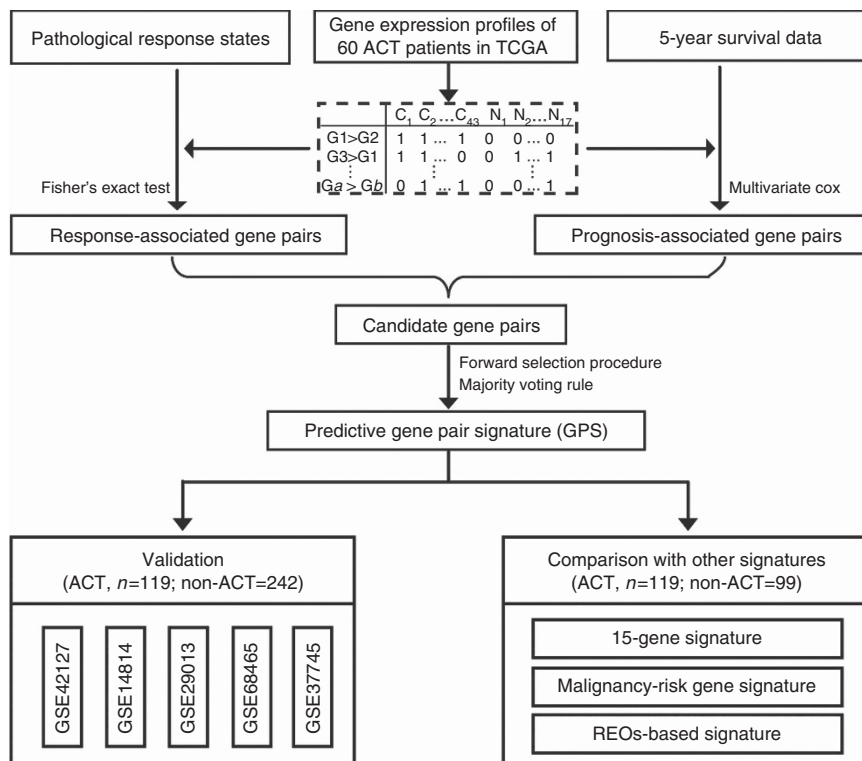


Figure 1. The flowchart of this study. Combining the pathological response and 5-year survival data of 60 patients receiving platinum-based ACT in TCGA, we develop a REO-based signature. The signature was tested in an integrated independent data set and compared with three signatures developed based on patients' survival solely.

3-GPS classified 37 and 23 patients into complete response (CR) and non-response (NR) groups, respectively, with significantly different 5-year survival rate (log-rank $P < 0.0001$; Supplementary Figure S1B). As predicted by 3-GPS, five pNR patients reclassified as responders had tentatively higher 5-year survival rate than the 11 pCR patients reclassified as non-responders (log-rank $P = 0.1710$; Supplementary Figure S1C). We also compared the gene expression patterns of the reclassified samples with the other samples. First, we identified 1239 DE genes (RankProd algorithm, $FDR < 0.05$) between the 32 responders and 12 non-responders consistently classified by 3-GPS and their original pathological response states. Second, based on the expression measurements of the top 50 significant DE genes, 60 samples were classified into two subgroups using the K-means clustering algorithm (Figure 2). The result showed that all the five pNR patients were clustered with 30 of the 32 consistent responders, providing transcriptional evidence that the five patients might be responders. Similarly, 8 of the 11 pCR patients were clustered with 10 consistent non-responders. These results suggested that the two groups identified by 3-GPS had more distinct transcriptional patterns than the original pathological response groups, suggesting a better platinum response classification of 3-GPS.

Between the two response groups predicted by 3-GPS, 1814 DE genes were identified (RankProd, $FDR < 0.05$). These DE genes

were significantly enriched in 73 GO functional terms (hypergeometric test, $FDR < 0.05$; Supplementary Table S3), including cell adhesion (Dalton, 2000), chemokine-mediated signalling pathway (Yin *et al*, 2013) and positive regulation of cell proliferation (Biliran *et al*, 2005), which have been reported to be related with platinum resistance. In the cell adhesion term, more than 80% of DE genes were upregulated in the predicted non-responders, among which *LAMB2* is a known extracellular matrix glycoprotein promoting platinum resistance (Kim *et al*, 2010). The other dysregulated functions, such as calcium ion transmembrane transport (Helson, 1984) and regulation of interferon-gamma production (Marth *et al*, 1997), might also be related with platinum sensitivity. In addition, we found 18 of the 1814 DE genes were also significantly correlated with GI_{50} values of cancer cell lines for carboplatin (Spearman correlation analysis, $FDR < 0.2$) based on the NCI60 data set. The concordance score, as described in detail in Supplementary Data SB, was 94.44% (binomial test, $P < 0.0001$), which meant that 94.44% of the up- or downregulations of the 18 DE genes in the predicted non-responders compared with the predicted responders could be explained by their correlations with platinum resistance. Similarly, we found 5 of the 1814 DE genes were also significantly correlated with GI_{50} values of cancer cell lines for cisplatin in the NCI60 data set, and the concordance score was 100% (binomial test, $P < 0.0001$).

Table 2. Composition of 3-GPS

Signature ^a	REOs (Ga > Gb) ^b	β^c	P^d	C-index
Gene pair 1	<i>LOC81691 > ZMYND10</i>	-5.56	0.0133	0.80
Gene pair 2	<i>TREM1 > ANPEP</i>	-2.43	0.0450	0.79
Gene pair 3	<i>ERN1 > FA2H</i>	-3.95	0.0213	0.77

^aPredictive model: a sample was determined to be responder if any two of the three gene pairs with the specific REOs (Ga > Gb); otherwise, the non-responder.
^bRepresents the REO (Ga > Gb) more frequently occur in the pCR group than in the pNR; The statistics are calculated from multivariate Cox regression model.
^c β represents risk coefficient of REO (Ga > Gb).
^d P represents the significance of the correlation between the REO (Ga > Gb) and 5-year survival rate.

Validation of the signature. First, in the testing ACT group, the 82 patients classified into the CR group by 3-GPS had significantly higher 5-year survival rate than the 37 patients classified into the NR group (log-rank $P = 0.0006$, C-index = 0.63; Figure 3A). Specifically, 3-GPS could also discriminate 5-year survival of 55 stage IB ACT patients (log-rank $P = 0.0055$; Figure 3B) and 64 stage II-III ACT patients (log-rank $P = 0.0211$; Figure 3C), respectively. Multivariate Cox analysis showed that 3-GPS remained significantly associated with 5-year survival rate of ACT patients after adjusting for stage, age and gender ($P = 0.0008$, hazard ratio = 0.33, 95% CI = 0.17-0.63; Supplementary Table S4). In contrast, in the non-ACT group, the predicted 156 CR and 86 NR patients had no survival difference (log-rank $P = 0.3796$;

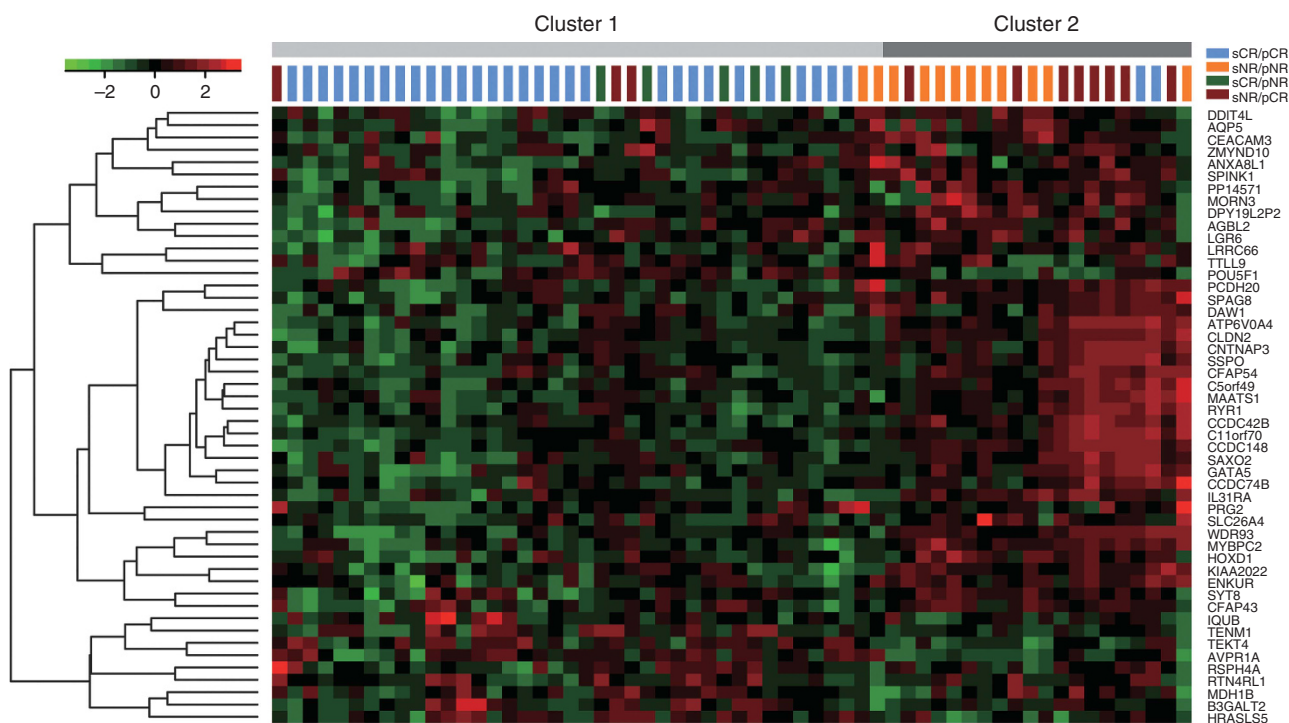


Figure 2. The K-means clustering of 60 patients receiving platinum-based ACT based on the top 50 differentially expressed genes between the responders and non-responders which were consistently classified by 3-GPS and their pathological response states.

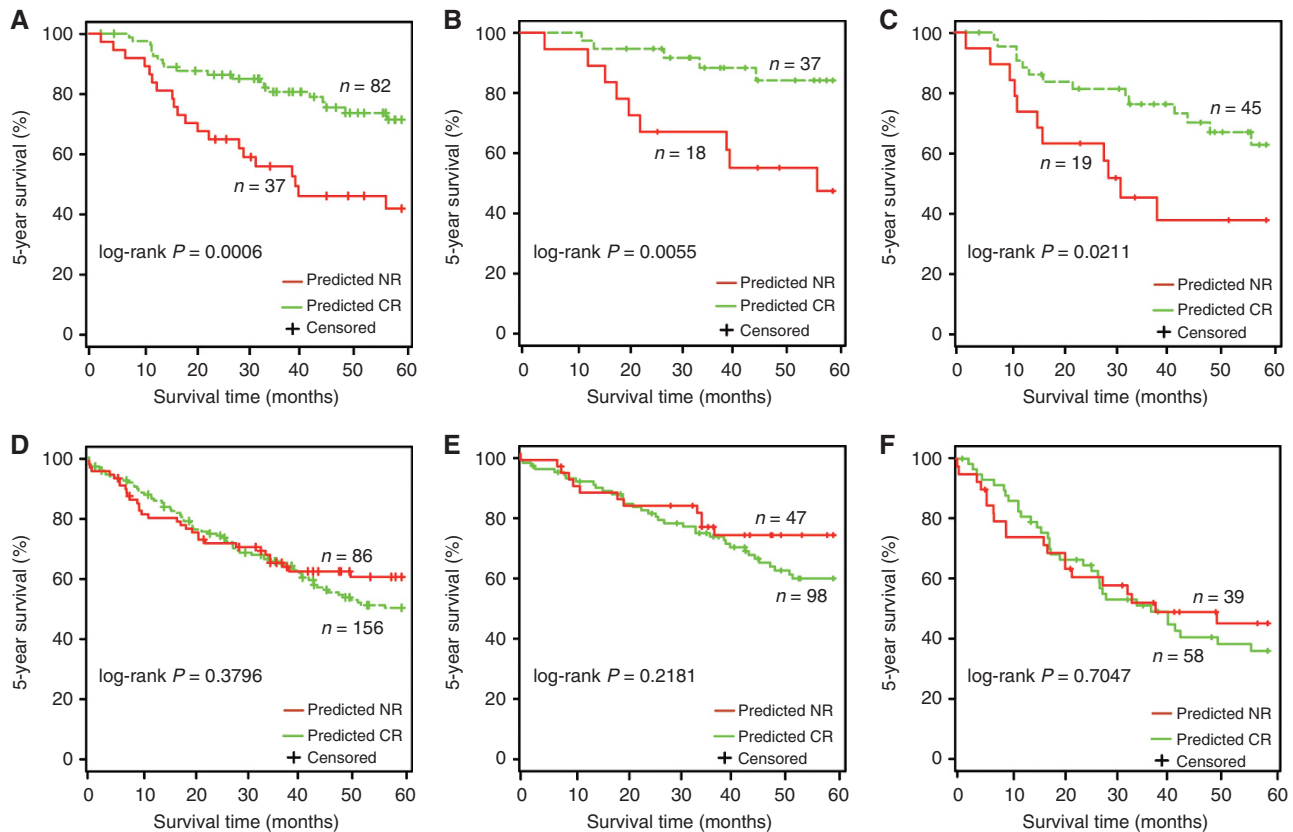


Figure 3. The Kaplan–Meier curves of 5-year survival between different response groups predicted by 3-GPS in the testing data set. (A) All patients receiving platinum-based ACT. (B) Stage IB patients receiving platinum-based ACT. (C) Stage II–III patients receiving platinum-based ACT. (D) All patients not receiving ACT. (E) Stage IB patients not receiving ACT. (F) Stage II–III patients not receiving ACT.

Figure 3D). Similar results were observed for the 145 stage IB patients (log-rank $P = 0.2181$; Figure 3E) and the 97 stage II–III patients (log-rank $P = 0.7047$; Figure 3F) in the non-ACT group.

Then, we estimated the 5-year survival benefit from platinum-based ACT. In the predicted CR group, the 5-year survival rate of the 82 ACT patients was 72.30%, significantly higher than 50.30% for the 156 non-ACT patients, with a 22% absolute benefit (log-rank $P = 0.0019$; Figure 4A). Specifically, the predicted responders showed 25.50 and 26.60% absolute benefits of ACT in the 5-year survival rate for stage IB (log-rank $P = 0.0142$; Figure 4B) and stage II–III patients (log-rank $P = 0.0073$; Figure 4C), respectively. In contrast, in the predicted NR group, patients receiving ACT had no significant survival benefit when compared with patients not receiving ACT (log-rank $P = 0.1348$; Figure 4D). Similar results were observed for stage IB (log-rank $P = 0.1052$, Figure 4E) and stage II–III patients (log-rank $P = 0.7317$; Figure 4F).

Comparison of 3-GPS with three other signatures. We also compared 3-GPS with two published signatures, the 15-gene signature (Zhu *et al*, 2010) and the malignancy-risk gene signature (Chen *et al*, 2011), which were both developed based on survival of patients without receiving ACT. The 12-gene signature (Tang *et al*, 2013) cited in the Introduction was not analysed because its application to independent data needs resetting risk thresholds, which makes it a non-independent validation. The 37-gene signature (Van Laar, 2012) developed based on survival of patients receiving ACT were not analysed because the author did not provide the predictive model. Briefly, for each of the 15-gene signature and the malignancy-risk gene signature, a risk score for a sample was calculated as a weighted sum of principal components based on the expression measurements of the signature genes (Zhu *et al*, 2010; Chen *et al*, 2011). Notably, 26 and 117 non-ACT

samples from the GSE14814 and GSE68465 data sets were excluded from the integrated testing data set, as they were used for training the two signatures. In the remained 119 ACT patients and 99 non-ACT patients in the integrated testing data set, the two signatures classified all samples into high-risk group when no Z-score normalisation was performed. This provided further evidence that this type of signatures would be unsuitable for direct application to individual samples without pre-collection of a set of samples for data normalisation, as reported in our previous study (Qi *et al*, 2016). Even when Z-score normalisation was performed, the high-risk patients predicted by the 15-gene signature and the malignancy-risk gene signature showed 11.50 and 11.30% absolute benefits of ACT in 5-year survival rate, respectively, when compared with high-risk patients without receiving ACT. These absolute benefits were much lower than the 16.8% 5-year absolute benefit of ACT for the responders predicted by 3-GPS (Supplementary Figure S2). Developed based on survival of patients without receiving platinum-based ACT, the two signatures were just able to identify patients with poor prognoses who need adjuvant treatment but unable to identify patients who might be sensitive to platinum-based ACT. Thus, the high-risk patients predicted by them included ACT-resistant patients who would not benefit from ACT. In addition, we also developed a REO-based signature for predicting prognoses of patients receiving platinum-based ACT, following the same approach for developing 3-GPS but using only patients' survival (Supplementary Data SC). The developed REO-based signature, consisting of five gene pairs, was denoted as 5-GPS. The result showed that the low-risk patients ('responders') predicted by 5-GPS showed a 12.4% absolute benefit in 5-year survival rate after receiving ACT, which was also lower than the 16.8% 5-year survival absolute benefit of ACT for the responders predicted by

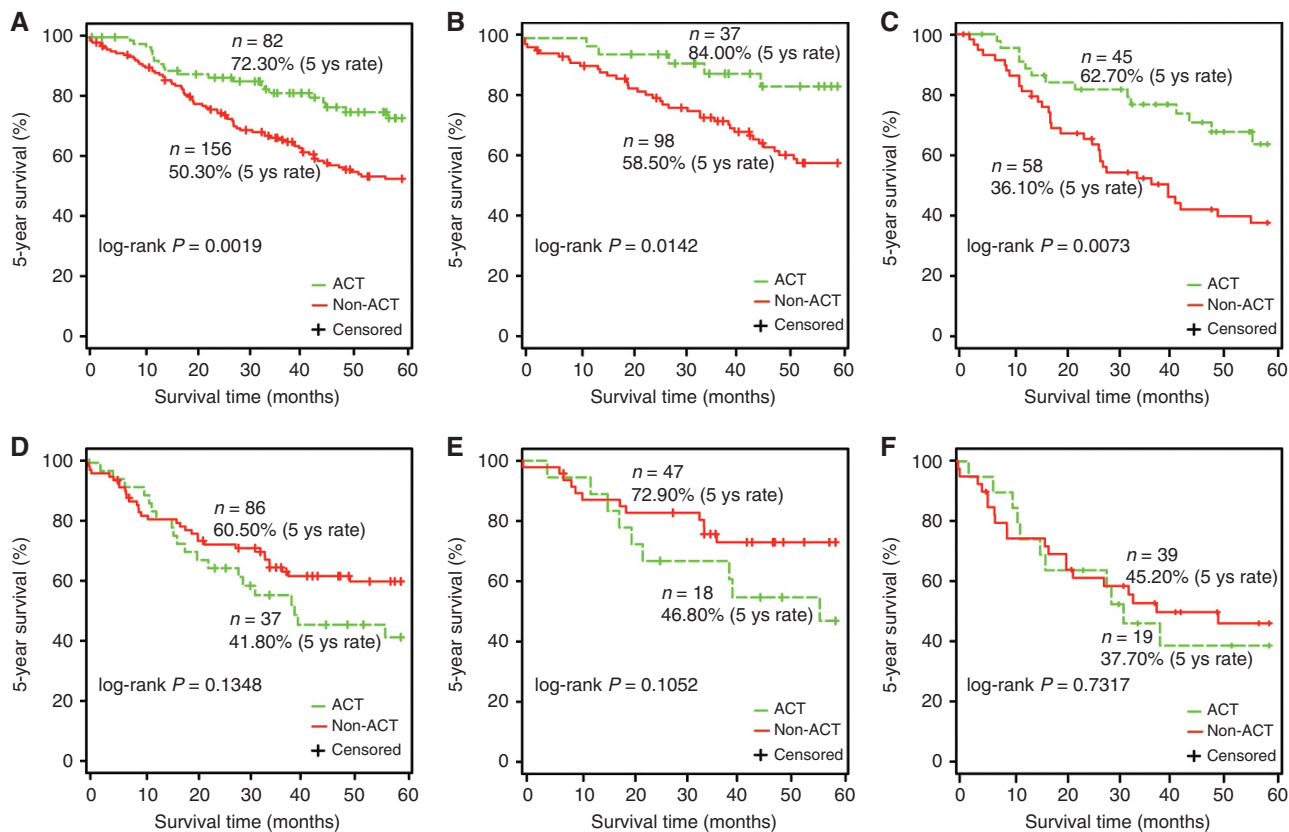


Figure 4. The benefits of ACT in 5-year survival rate for the different response groups predicted by 3-GPS in the testing data set. (A) All patients in the predicted CR group. (B) Stage IB patients in the predicted CR group. (C) Stage II-III patients in the predicted CR group. (D) All patients in the predicted NR group. (E) Stage IB patients in the predicted NR group. (F) Stage II-III patients in the predicted NR group. The 5 yr rate represents 5-year survival rate.

3-GPS (Supplementary Figure S2). This could be explained by the possibility that the responders predicted by the signature developed only based on survival of patients receiving platinum-based ACT might include patients resistant to platinum-based ACT but the tumour cells had a low degree of malignancy, and these patients would not benefit from ACT. These results suggested that 3-GPS, developed by combining pathological response and survival data, had a better predictive performance than the signatures developed based on survival data solely.

DISCUSSION

In this study, combining the pathological response and 5-year survival data, we developed a REO-based signature consisting of three gene pairs (3-GPS), which could individually identify the platinum responders with higher 5-year survival rate after receiving platinum-based ACT. Currently, the survival benefit of platinum-based ACT for stage IB LUAD patients remains controversial (Tsuboi *et al*, 2007). Our results showed that platinum-based ACT significantly improved 5-year survival rate of the predicted stage IB responders (log-rank $P=0.0142$; Figure 4B), while not beneficial to the predicted non-responders (log-rank $P=0.1052$; Figure 4E). Thus, the application of 3-GPS might allow clinicians to select stage IB patients who should adopt platinum-based ACT. Although platinum-based ACT is a standard treatment for stage II-III patients after surgical resection, the non-responders identified by 3-GPS could not benefit from platinum-based ACT (log-rank $P=0.7317$; Figure 4F), suggesting that these patients should adopt other treatments.

Notably, current platinum-based ACT is generally administered with other drugs for LUAD patients (Wang and Lippard, 2005). Recently, our study have proved that genes related to a single drug sensitivity could be identified in clinical samples of patients administered with combination ACT, given that the drugs used in combination had no or limited pharmacological antagonism (Tong *et al*, 2015). In this study, based on the cell lines data for drug resistance, we demonstrated that DE genes between the responders and non-responders predicted by 3-GPS were concordantly correlated with both carboplatin and cisplatin sensitivity, providing a circumstantial evidence of the relevance between 3-GPS and platinum ACT.

In conclusion, the REO-based predictive signature for platinum-based ACT in LUAD patients can identify patients who may benefit from platinum-based ACT. Because the within-sample REOs are robust against to experimental batch effects (Wang *et al*, 2015) and differences of probes designed for different platforms (Guan *et al*, 2016), it would be possible to develop a customised microarray kit, with a similar reaction conditions for the microarray platforms analysed in this study, to measure the REOs of the three gene pairs for clinical application. It would also be possible to design a reverse transcriptase PCR kit to measure the REOs robustly (Leek *et al*, 2010), which deserves our future experimental research. The robustness and simplicity of the REO-based signature would make it convenient in clinical settings and merits further validation in a prospective clinical trial.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Benjamini A, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* **57**: 289–300.
- Biliran Jr H, Wang Y, Banerjee S, Xu H, Heng H, Thakur A, Bollig A, Sarkar FH, Liao JD (2005) Overexpression of cyclin D1 promotes tumor cell growth and confers resistance to cisplatin-mediated apoptosis in an elastase-myc transgene-expressing pancreatic tumor cell line. *Clin Cancer Res* **11**(16): 6075–6086.
- Bland JM, Altman DG (2004) The logrank test. *BMJ* **328**: 1073.
- Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M, Berglund A, Ekman S, Bergqvist M, Ponten F, König A, Fernandes O, Karlsson M, Helenius G, Karlsson C, Rahnenführer J, Hengstler JG, Micke P (2013) Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clin Cancer Res* **19**: 194–204.
- Chen DT, Hsu YL, Fulp WJ, Coppola D, Haura EB, Yeatman TJ, Cress WD (2011) Prognostic and predictive value of a malignancy-risk gene signature in early-stage non-small cell lung cancer. *J Natl Cancer Inst* **103**: 1859–1870.
- Crino L, Weder W, van Meerbeeck J, Felip E (2015) Early stage and locally advanced (non-metastatic) non-small-cell lung cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-up. *Ann Oncol* **21**: v103–v115.
- Dalton WS (2000) The tumor microenvironment as a determinant of drug response and resistance. *Curr Opin Oncol* **12**(6): 557–563.
- Earl H, Provenzano E, Abraham J, Dunn J, Vallier AL, Gounaris I, Hiller L (2015) Neoadjuvant trials in early breast cancer: pathological response at surgery and correlation to longer term outcomes – what does it all mean? *BMC Med* **13**: 234.
- Eddy JA, Sung J, Geman D, Price ND (2010) Relative expression analysis for molecular cancer diagnosis and prognosis. *Technol Cancer Res Treat* **9**: 149–159.
- Geman D, d'Avignon C, Naiman DQ, Winslow RL (2004) Classifying gene expression profiles from pairwise mRNA comparisons. *Stat Appl Genet Mol Biol* **3**: Article19.
- Gruber, Rueckert M, Kagan KO, Staebler A, Siegmann KC, Hartkopf A, Wallwiener D, Hahn M (2013) Measurement of tumour size with mammography, sonography and magnetic resonance imaging as compared to histological tumour size in primary breast cancer. *BMC Cancer* **13**: 328.
- Guan Q, Chen R, Yan H, Cai H, Guo Y, Li M, Li X, Tong M, Ao L, Li H, Hong G, Guo Z (2016) Differential expression analysis for individual cancer samples based on robust within-sample relative gene expression orderings across multiple profiling platforms. *Oncotarget*; e-pub ahead of print 13 September 2016; doi:10.18632/oncotarget.11996.
- Harrell Jr FE, Lee KL, Mark DB (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* **15**: 361–387.
- Helson L (1984) Calcium channel blocker enhancement of anticancer drug cytotoxicity—a review. *Cancer Drug Deliv* **1**(4): 353–361.
- Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J (2006) RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* **22**: 2825–2827.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249–264.
- Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* **11**(10): 733–739.
- Kim YS, Hwan JD, Bae S, Bae DH, Shick WA (2010) Identification of differentially expressed genes using an annealing control primer system in stage III serous ovarian carcinoma. *BMC Cancer* **10**: 576.
- Marth C, Widschwendter M, Kaern J, Jørgensen NP, Windbichler G, Zeimet AG, Tropé C, Daxenbichler G (1997) Cisplatin resistance is associated with reduced interferon-gamma-sensitivity and increased HER-2 expression in cultured ovarian carcinoma cells. *Br J Cancer* **76**(10): 1328–1332.
- Qi L, Chen L, Li Y, Qin Y, Pan R, Zhao W, Gu Y, Wang H, Wang R, Chen X, Guo Z (2016) Critical limitations of prognostic signatures based on risk scores summarized from gene expression levels: a case study for resected stage I non-small-cell lung cancer. *Brief Bioinform* **17**: 233–242.
- Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, Misek DE, Chang AC, Zhu CQ, Strumpf D, Hanash S, Shepherd FA, Ding K, Seymour L, Naoki K, Pennell N, Weir B, Verhaak R, Ladd-Acosta C, Golub T, Gruidl M, Sharma A, Szoke J, Zakowski M, Rusch V, Kris M, Viale A, Motoi N, Travis W, Conley B, Seshan VE, Meyerson M, Kuick R, Dobbin KK, Lively T, Jacobson JW, Beer DG (2008) Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**: 822–827.
- Siegel RL, Miller KD, Jemal A (2015) Cancer statistics, 2015. *CA Cancer J Clin* **65**: 5–29.
- Subramanian J, Simon R (2010) Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* **7**: 464–474.
- Tang H, Xiao G, Behrens C, Schiller J, Allen J, Chow CW, Suraokar M, Corvalan A, Mao J, White MA, Wistuba, Minna JD, Xie Y (2013) A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin Cancer Res* **19**: 1577–1586.
- The Cancer Genome Atlas Research Network (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**: 543–550.
- Tong M, Zheng W, Lu X, Ao L, Li X, Guan Q, Cai H, Li M, Yan H, Guo Y, Chi P, Guo Z (2015) Identifying clinically relevant drug resistance genes in drug-induced resistant cancer cell lines and post-chemotherapy tissues. *Oncotarget* **6**: 41216–41227.
- Tsuboi M, Ohira T, Saji H, Miyajima K, Kajiwara N, Uchida O, Usuda J, Kato H (2007) The present states of postoperative adjuvant chemotherapy for completely resected non-small cell lung cancer. *Ann Thorac Cardiovasc Surg* **13**: 73–77.
- Van Laar RK (2012) Genomic signatures for predicting survival and adjuvant chemotherapy benefit in patients with non-small-cell lung cancer. *BMC Med Genomics* **5**: 30.
- Wallerek S, Sorensen JB (2015) Biomarkers for efficacy of adjuvant chemotherapy following complete resection in NSCLC stages I-IIIa. *Eur Respir Rev* **24**: 340–355.
- Wang D, Lippard SJ (2005) Cellular processing of platinum anticancer drugs. *Nat Rev Drug Discov* **4**: 307–320.
- Wang H, Sun Q, Zhao W, Qi L, Gu Y, Li P, Zhang M, Li Y, Liu SL, Guo Z (2015) Individual-level analysis of differential expression of genes and pathways for personalized medicine. *Bioinformatics* **31**(1): 62–68.
- Wu D, Wang X (2015) Application of clinical bioinformatics in lung cancer-specific biomarkers. *Cancer Metastasis Rev* **34**: 209–216.
- Xie Y, Xiao G, Coombes KR, Behrens C, Solis LM, Raso G, Girard L, Erickson HS, Roth J, Heymach JV, Moran C, Danenberg K, Minna JD, Wistuba II (2011) Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. *Clin Cancer Res* **17**: 5705–5714.
- Yin F, Liu X, Li D, Wang Q, Zhang W, Li L (2013) Bioinformatic analysis of chemokine (C-C motif) ligand 21 and SPARC-like protein 1 revealing their associations with drug resistance in ovarian cancer. *Int J Oncol* **42**(4): 1305–1316.
- Zhao W, Chen B, Guo X, Wang R, Chang Z, Dong Y, Song K, Wang W, Qi L, Gu Y, Wang C, Yang D, Guo Z (2016) A rank-based transcriptional signature for predicting relapse risk of stage II colorectal cancer identified with proper data sources. *Oncotarget* **7**: 19060–19071.
- Zhu CQ, Ding K, Strumpf D, Weir BA, Meyerson M, Pennell N, Thomas RK, Naoki K, Ladd-Acosta C, Liu N, Pintilie M, Der S, Seymour L, Jurisica I, Shepherd FA, Tsao MS (2010) Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol* **28**: 4417–4424.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)