Dear Editor,

Lung cancer—predominantly non-small cell lung cancer (NSCLC)—is the most prevalent cancer in the world, in terms of both incidence and mortality. Squamous cell lung cancer (SCLC) is the second most common type of NSCLC, making up about 30% of all cases. The average 5-year survival rate among NSCLC patients is barely 15% (Herbst et al., 2008). Among post-surgery patients, prognosis varies widely, even among patients with similar clinicopathological characteristics, demonstrating the need for improved ways to predict treatment outcomes (Chen et al., 2007).

Previous global gene-expression studies have identified genes that could be referred to as signatures for classifying patients with significantly different prognostic outcomes. However, these studies generally analyzed NSCLC patients with mixed histopathologic subsets and some of them only focused on adenocarcinomas (ADC) or stage I disease (Potti et al., 2006; Chen et al., 2007; Guo et al., 2008; Shedden et al., 2008; Tomida et al., 2009; Jeong et al., 2010; Kadara et al., 2011). Interestingly, each study identified different sets of gene signatures. This could reflect the genetic heterogeneity of NSCLC and thus suggests the necessity of stratifying the prognosis of lung cancer with respect to more accurate grouping information, such as stages, histopathologic subsets, or other demographic factors. Correspondingly, currently known NSCLC blood biomarkers are rare and of little use in the prediction of prognosis. For the purpose of patient-tailored therapeutics, we herein attempted to identify relapse-related signatures that allow the selection of late stage SCLC with a high probability of rapid relapse. By harnessing current powerful approaches in shotgun and targeted proteomics (based on selected reaction monitoring, SRM) (Huttenhain et al., 2009), we have succeeded in profiling the pre-therapy serum proteome of patients with an improved analysis depth. Moreover, we elaborated a directed approach to avoid possible confounding factors, such as clinical and pathological features that might complicate efforts to delineate relapse discrepancies. The triple non-invasive blood signatures discovered here may be promising for refining the prognosis of stage IIb and IIIa SCLC.

Fifty patients with SCLC who had disease recurrence within 10 months (7.44 ± 2.08 months) after surgery were selected and defined as the ‘rapid recurrence’ group (RR group). Fifty-six patients who showed no evidence of recurrent disease (ND) >20 months (>27.04 ± 5.73 months) post-surgery were defined as the ND group. To identify relapse-related blood proteins, all RR patients were selected with different genders, smoking history, metastasis symptoms, varied ages (from 41 to 73) and pathological features determined TNM stages and, therefore, may present fast relapse SCLC diseases with varied clinical and pathological features. Both RR and ND patients had either stage IIb or IIIa disease. All the ND patients under the same adjuvant-therapy and follow-up surveillance post-surgery were accordingly selected to match RR cases by gender, age, smoking habits and the exact pathological features determined TNM stage, except for a much better prognostic outcome (Figure 1A). A total of 106 patients were separated into two cohorts for this study: Cohort I, 30 pairs of RR/ND and Cohort II, 20RR/26ND. Patient demographics are summarized in Supplementary Table S1, with all clinicopathologic details listed in Supplementary Table S2.

Five pairs of RR/ND samples in Cohort I were used and compared in the discovery phase (Figure 1B). Collectively, 711012 peptide hits from 4139 unique peptides, assigned to 595 protein groups, were credibly identified by shotgun proteomics with the final protein-level false discovery rate (FDR) <1% in each pair. And 386–439 serum proteins were identified from each patient. The final peptide FDR for the whole data set was only 0.016%. By applying a label-free method termed localized statistics of protein abundance distribution (LSPAD) (Li et al., 2008; Luo et al., 2011), we identified 64–117 significantly up-regulated proteins and 70–108 down-regulated proteins in RR cases compared with ND cases (P < 0.05, Figure 1C and Supplementary Figure S1). To distinguish the true protein abundance changes from random fluctuations, we used the following stringent criteria: (i) an LSPAD, P < 0.05 in all the five pairs; (ii) an average of P < 0.01; (iii) a fold-change of >1.3 between the RR and ND groups; and (iv) at least 20 MS/MS spectra had to be assigned to the candidate due to the sensitivity issue of spectral counting. Twenty-five proteins were thus identified as relapse-related signatures. Literature mining and biological interests then led us to focus on three proteins—C4b-binding protein alpha (C4BP), leucine-rich alpha-2-glycoprotein (LRG1), and serum amyloid A protein (SAA) for the next phase of...
Figure 1 The discovery of serum prognosis biomarkers for incidental N2 squamous cell lung cancer. (A) Serum samples were collected, paired, depleted, and extensively analyzed by shotgun proteomics (left panel). SRM assay-based targeted proteomics were adopted to verify three biomarkers filtered from the discovery phase (right panel). SCX, strong cation exchanger; HSA, human serum albumin; IgG, immunoglobulin.
Six pure heavy labeled peptides were synthesized for developing the SRM assay (Supplementary Figure S2), which confidently and reproducibly quantified proteins down to μg/ml with extremely low inter-assay coefficient of variation (1.20%–1.90%, Supplementary Table S3). Samples were run in duplicates and the technical reproducibility was excellent ($R^2 = 0.9820–0.9998$, Supplementary Figure S3); therefore, the averaged data were used for each sample. Using SRM coupled with absolute quantification strategy, we validated the altered levels of C4BP, LRG1 and SAA in pre-treatment serum from 60 pair-wise SCLC patients (i.e., Cohort I, see Supplementary Table S2). The results illustrated in Figure 1D confirmed their clinical significance (C4BP, $P = 2.01E - 6$; LRG1, $P = 1.55E - 6$; SAA, $P = 0.0033$; paired Student's t-test). The relative averaged ratio between RR and ND cases are 1.47, 1.68 and 3.55 for C4BP, LRG1 and SAA, respectively. We then plotted the ROC curve for each candidate and found all their corresponding area under ROC curve (AUC) were >0.8 (C4BP, 0.887; LRG1, 0.878; SAA, 0.808, Supplementary Figure S4), indicating a strong discriminative power. The recommended cutoffs were assessed by minimizing the distance from the ideal point in the ROC plot (see Supplementary Method) and were then applied for the further Kaplan–Meier (KM) analysis in a second independent patient cohort (Cohort II, 20 RR and 26 ND cases). The clinicopathological characters were also well balanced in this cohort. Intriguingly, significant differences of the cumulative relapse probability (Figure 1E) could have been predicted if we had partitioned Cohort II patients before surgery according to either cutoff obtained from Cohort I (C4BP, $P = 0.0140$; LRG1, $P = 0.0136$; SAA, $P = 0.0009$; log-rank test).

To combine the identified candidates, we trained a classification and regression tree (CART) model using the data of 60 SCLC in Cohort I, which identified that all the three proteins should be included as the optimal panel with six terminals. Overall, the CART model correctly classified 27 out of 30 (90.0%) RR and 26 out of 30 (86.6%) ND patients (Figure 1F). In the test set of Cohort II, 16 out of 20 RR cases were successfully assigned to high-risk nodes, and 19 of 26 ND sera were binned to low-risk nodes (sensitivity, 80.0%; specificity, 73.1%). The high classification power of terminal 6 in both the training and test sets suggests that C4BP in combination with SAA could serve as the foundation of a very powerful prognostic blood test for resectable SCLC in the future.

Of note, the biomarker panel created by logistic regression which incorporated the three proteins had an AUC of 0.911 in Cohort I, which was better than using either protein alone (Figure 1G). Moreover, the most significant prediction result in Cohort II could be achieved if we use CART rules as a classifier (Figure 1H, $P = 0.0004$). These data indicate that combining the candidates is helpful to improve the prediction performances.

We next distributed the serum levels in all SCLC cases in comparison with 72 age-matched control specimens. Importantly, the serum levels of LRG1 and SAA between the ND and normal cases were also remarkably different (LRG1, $P = 1.51E - 9$; SAA, $P = 0.0013$), while the levels of C4BP were not ($P = 0.441$, Figure 1). Thus, LRG1 and SAA deserve further evaluation to serve as multimarkers for both locally advanced SCLC diagnosis and prognosis, while C4BP could be applied as a specific prognosis-related protein for stage Ib and IIa SCLC. Finally, though the therapy after recurrence varied among patients, significant differences in cumulative survival probability could also be predicted if we had partitioned the patient groups before surgery based on our triple biomarkers ($P < 0.0001$, Figure 1).

Lung cancer is a genetically heterogeneous disease. Indeed, several studies have indicated that prognostic predictors for ADC sometimes have minor or no effects in identifying high-risk SCLC. Taken together, our study has devoted an important effort towards personalized therapeutics that allows the identification of a group of late stage SCLC patients with a high probability of fast relapse. The usage of pre-therapy blood proteins to predict the prognosis of stage Ib and IIa SCLC has not been reported to the best of our knowledge. Notably, the overall prediction accuracy in this study is comparable to previous results of genetic signatures for lung cancer prognosis, whereas the genetic signatures reported are different between the studies even when various works are conducted to cross-validate one another (Potti et al., 2006; Chen et al., 2007; Guo et al., 2008; Shedden et al., 2008; Tomida et al., 2009; Jeong et al., 2010; Kadara et al., 2011). In addition, gene predictors and those microRNA signatures (Hu et al., 2010) cannot provide clinical cut-offs in an absolute quantitative scale. In this regard, we were able to directly quantify the absolute serum levels in patients using targeted proteomic approach. This absolute quantification may offer a better option for a clinical application. The SCLC SRM assay we developed is totally antibody-free, sensitive.
enough, easily transferable and exquisitely reproducible.

Our results underscore the feasibility and importance of establishing novel clinical decision rules based on the levels of serum proteins, at least to stage IIb and IIIa of SCLC. The application of these proteins, especially a combination of C4BP, LRG1 and SAA (Figure 1F) or C4BP alone (Figure 1I) are not only powerful in prognosis determination but also promising in treatment optimization and overtreatment prevention (Cho et al., 2010). Future confirmatory studies and clinical trials are warranted to evaluate the final outcomes of possible patient-tailored adjuvant therapies based on these serum signatures.

[Supplementary material is available at Journal of Molecular Cell Biology online. This project was supported by grants from the National Natural Science Foundation of China (91029301, 81101760, 81172218, and 81101761), Science and Technology Commission of Shanghai Municipality (09JC1416302), Fudan University (09FQ78), and Fudan University Cancer Hospital (YJ200804, YJ200701).]

Yan-Sheng Liu1,2,3, Xiao-Yang Luo2,3,4, Qing-Run Li5, Hong Li3, Chen Li1, Hong Ni6, Rong-Xia Li3, Rui Wang2,3, Hai-chuan Hu2,3, Yun-jian Pan2,3, Hai-Quan Chen2,3,4, and Rong Zeng1,*

1Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
2Department of Thoracic Surgery, Fudan University Shanghai Cancer Center, Shanghai 200032, China
3Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China
4These authors contributed equally to this work.

*Correspondence to: Rong Zeng,
E-mail: zr@sibs.ac.cn; Hai-Quan Chen,
E-mail: chenhq@shca.org.cn

References


