Machine Learning to Build and Validate a Model for Radiation Pneumonitis Prediction in Patients with Non-Small-Cell Lung Cancer

Hao Yu[1,2], Huanmei Wu[2], Weili Wang[3], Shruti Jolly[4], Jianyue Jin[3], Chen Hu[5], and Feng-Ming (Spring) Kong[3,6,7] *

1 Biomedical Engineering, Shenzhen Polytechnic, Shenzhen, China

2 BioHealth Informatics, School Of Informatics and Computing, IUPUI, Indianapolis, IN

3 Department of Radiation Oncology, University Hospitals Cleveland Medical Center/Seidman Cancer Center and Case Comprehensive Cancer Center of Case Western Reserve University, Cleveland, OH

4 Radiation Oncology, University of Michigan, Ann Arbor, MI

5 Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD

6 Department of Clinical Oncology, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China

7 Department of Clinical Oncology, The University of Hong Kong and Shenzhen Hospital, Hong Kong, China

Running head

Machine learning build and validate RP2 prediction in NSCLC

Keywords

Machine learning; Non-small-cell lung cancer; cytokine; radiation pneumonitis;

Financial supports

The Project was supported in part by NIHNCI R01 CA142840 (F.M. Kong), and supported by Province Natural Science Foundation of Guangdong Grant No. 2015A030313593 (H. Yu); Science and Technology Planning Project of Shenzhen, China, Grant No. JCYJ20160413164156155 (H. Yu).

Corresponding author

Feng-Ming (Spring) Kong, MD, PhD, FACR, FASTRO

Department of Clinical Oncology, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China; Clinical Oncology Center, The University of Hong Kong-Shenzhen Hospital, Shenzhen, China

Department of Radiation Oncology, University Hospitals Cleveland Medical Center /Seidman Cancer Center and Case Comprehensive Cancer Center of Case Western Reserve University, OH

Tel: 1-317-944-2524/1-317-944-1303

Fax: 1-317-944-2486

E-mail: kong0001@hku.hk; fxk132@case.edu

A conflict of interest disclosure statement

The author(s) indicated no potential conflicts of interest associated with this project.

This is the author's manuscript of the article published in final edited form as:

Yu, H., Wu, H., Wang, W., Jolly, S., Jin, J.-Y., Hu, C., & Kong, F.-M. (Spring). (2019). Machine Learning to Build and Validate a Model for Radiation Pneumonitis Prediction in Patients with Non–Small Cell Lung Cancer. Clinical Cancer Research, 25(14), 4343–4350. https://doi.org/10.1158/1078-0432.CCR-18-1084

Statement of translational relevance

Radiation pneumonitis is a dose limiting toxicity of thoracic radiation therapy. Combining patient factors like hypertension, the lung dosimetric parameters like mean lung dose, and plasma levels of biomarkers like IL-8 and CCL2, this study built and validated internally a predictive model for radiation pneumonitis grade≥2 with AUC of 0.863 and accuracy 80% in patients with non-small-cell lung cancer who underwent radiation therapy. Should these be validated by an external database, this study will provide an opportunity to guide clinicians for personalized radiation dose prescription in future trials or clinical practice, to improve patient's survival while limiting the risk of radiation pneumonitis.

Abstract

Purpose

Radiation pneumonitis is an important adverse event in patients with non-small-cell lung cancer (NSCLC) receiving thoracic radiation therapy (RT). However, the risk of radiation pneumonitis grade⩾2 (RP2) has not been well predicted. This study hypothesized that inflammatory cytokines or the dynamic changes during-RT can improve predictive accuracy for RP2.

Materials and Methods

Levels of 30 inflammatory cytokines and clinical information in patients with stages I-III NSCLC treated with RT were from our prospective studies. Statistical analysis was used to select predictive cytokine candidates and clinical covariates for adjustment. Machine learning algorithm was used to develop the generalized linear model for predicting risk RP2.

Results

A total of 131 patients were eligible, 17 (13.0%) developed RP2. IL-8 and CCL2 had significantly (Bonferroni) lower expression levels in patients with RP2 than without RP2. But none of the changes in cytokine levels during RT was significantly associated with RP2. The final predictive GLM model for RP2 was established including IL-8 and CCL2 at baseline level and two clinical variables. Nomogram was constructed based on the GLM model. The model's predicting ability was validated in the completely independent test-set (area under curve=0.863, accuracy=80.0%, sensitivity=100%, specificity=76.5%).

Conclusion

By machine learning, this study has developed and validated a comprehensive model integrating inflammatory cytokines with clinical variables to predict RP2 before RT which provides an opportunity to guide clinicians.

Introduction

Radiation therapy (RT) plays an important role in the treatment of lung cancer, the leading cause of cancer death. Radiation induced lung toxicity (RILT) is a common and dose limiting adverse effect of thoracic RT in lung cancer patients, which may decrease quality of life, lead to pulmonary failure, and become life-threatening.[1,2] Radiation pneumonitis (RP), one of the commonly reported RILT, usually occurs within 1 to 6 months after completion of RT.[3] In patients treated with concurrent chemoradiation therapy,[4-6] 7.0 to 32.0% of patients have grade 2 and above (RP2) while 2.6 to 18.0% with severe RP Grade $\geqslant$3.

We and others have previously demonstrated that the risk of RILT is correlated with radiation dosimetric factors, like mean lung dose (MLD)[2,7,8] with AUC<0.60. Proteomic analysis demonstrated that molecules associated with inflammation pathways such as C4b-binding protein alpha chain (C4BPA), Complement C3 (C3) and vitronectin (VTN) had substantially higher expression levels in patients with grade $\geqslant$2 RILT.[9] Addition of C4BPA+VTN to MLD improved the RILT predictive accuracy (AUC=0.71).[10] We also found radiation-induced elevation in plasma TGF-β1 level during RT had predictive ability of RILT.[11] TGF-β1 combined with MLD stratified patients for high risk of RILT.[12] Additionally, we demonstrated that combining IL-8, TGF-β1, and MLD into a single model yielded a good predictive ability

(AUC=0.80).[8] In addition, baseline pulmonary function, including FEV1, FVC, and DLCO, may be related to the risk of RILT2.[13]

The pathogenesis of RILT is described as multiple inter-reacting cellular activities such as hypoxia, fibrogenesis, inflammation, and angiogenesis.[14] It is known that RILT combined the events of RP and radiation induced lung fibrosis (RILF) together, though RP and RILF have different biopathophysiological mechanisms. RP is associated with inflammatory reaction, while the latter is direct results of fibrosis and scar formation. The biomarkers of RP have been studied more extensively for clinical prediction. Kim JY, et al.[15] found that TGF-β1 level became significantly higher at 4 weeks after RT ($p = 0.007$). Variations of circulating IL-1A, IL-6, and IL-10 were also significant with RP ($p<0.05$).[16,17] Serum superoxide dismutase (SOD) has the predictive ability of RP with a sensitivity of about 0.8, and a specificity of about 0.7.[18,19] Additionally, some groups revealed a number of single-nucleotide polymorphism markers (SNPs)[20-23] were significantly correlated with the incidence of RP, including TGF-β1 rs1982073 with RP2 (hazard ratio = 0.489);[20] TGF-β1 rs11466345 with RP3 (hazard ratio = 2.295).[21] Genetic variation in the pro-inflammatory genes IL-1A, IL-8, TNF, TNFRSF1B, and MIF also significantly increased the risk of RP;[22] MTHFR rs1801131 with RP2 (hazard ratio = 0.37).[23] These studies suggested that individual patient's genetic makeup and cytokine milieu may play critical roles

in an individual's response to RP2 development. However, no study to date has reported good validated models to predict the risk of RP.

In this study, we hypothesized that cytokines may play a vital role in predicting RP. We measured the plasma levels of representative cytokines of having immunomodulating and inflammatory effects, including interleukin, colony stimulating factor, interferon, tumor necrosis factor, transforming growth factor, growth factor and chemokine families, and even their changing dynamic during the course of radiation. This study aimed to build and validate a model to predict RP2 by using plasma cytokine in patients with NSCLC who underwent radiation therapy.

Materials and Methods

Study population

Eligible subjects included patients with stages I-III NSCLC undergoing radiation alone or combined radiation with chemotherapy (UMCC 2003.073, UMCC 2003.076, NCT00603057, NCT01190527). Patients with a life expectancy of less than 6 months were excluded as they might not benefit from local radiation and might not be assessable for late lung toxicity. No restrictions were placed on either the degree of weight loss or pulmonary compromise, or oxygen dependency. All clinical data, including clinical parameters, grading of RP, and blood samples, were prospectively collected.

Radiation therapy

All patients received daily fractionated 3D conformal external beam radiotherapy technique with or without sequential or concurrent chemotherapy. No patients treated with stereotactic body RT were included in the analysis.[3] In general, the radiation dose prescription is limited by MLD of 20 Gy or normal lung tissue complication probability of 15%-17.5%. The lung dosimetric factors were computed with subtraction of gross tumor volume (GTV) overlapping with normal lung.

Cytokines measurement

A total 30 inflammation modulating cytokines were measured, including: EGF, VEGF, CCL2, CCL3, CCL4, CCL11, CX3CL1, CXCL10, G-CSF, GM-CSF, IFN-γ, TNF-α, IL-1a, IL-1b, IL-1r, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12p40, IL-12p70, IL-13, IL-15, IL-17, sCD40l, TGF-α, and TGF-β1. Cytokine measurements were performed in plasma samples at 3 time-points: at baseline (within 2 weeks before the start of RT) and at 2 and 4 weeks during RT. Cytokines were measured in pg/mL as previously described.[24,25]

Evaluation of radiation pneumonitis

The primary endpoint of radiation therapy was RP2, defined as RP grade$\geqslant$2. RP was diagnosed and graded based on a modified criteria combining

RTOG/SWOG/CTCAE. The detailed grading definitions were previously described in our previous publications,[2,12] consistent with a recent update from the expert panel of an AAPM task (Supplementary Table S1).[26]

Statistical analysis

Patients with detectable levels of all 30 cytokines were eligible. Chi-square test, Fisher's exact test, and logistic regression were applied for univariate clinical variables' analyses, in order to select covariates available for model development. GLMM (generalized linear mixed model) with Bonferroni multiplicity correction was used to assess the potential importance of each cytokine as well as its dynamics (at baseline, 2 and 4 weeks during-RT), after adjusting for potential prognostic covariates as identified in the univariate clinical variables' analyses. Machine learning algorithm was used for developing the final prediction GLM (generalized linear model) model of risk RP2 based on selected cytokines and clinical covariates, as described in the next section. The nomogram was built based on the GLM model for risk RP2. ROC curves and its corresponding AUC value, accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated to show the performance of the GLM model. All analyses were performed after data normalization: Log-transformed all cytokine levels and GTV. All statistical tests were 2-sided, and the overall adjusted p-value threshold was 0.05. All analyses were performed using the "R" version

3.3.2.[27]

Machine learning

Given the concern about generalization performance, machine learning was adopted here to build the generalized linear model (GLM). The methodology of the process can be divided into the following steps. 1) Divide the dataset into a training cross-validation set (trainCV-set) and a test-set (80% and 20%). 2) Use ten times 10-fold cross-validation (CV) in the trainCV-set to avoid over-fitting. In the progress of CV, the trainCV-set was divided into training-set and CV-set (9 folds and 1 fold); the most fitted GLM model was generated on training-set by model performance criteria AIC; and the mean standard error (MSE) of the GLM model was calculated on the CV-set. 3) Select the model, which had an MSE value closest to the mean MSE value, as the final model for predicting RP2. 4) Tested the final GLM model in test-set, which is unused and is completely independent with the trainCV-set. Given the imbalanced data (17 cases with RP2), we random resample those datasets carefully in keeping the ratio of RP2 not to zero.

Results

Patients Characteristics and RP2

A total of 131 consecutive patients with NSCLC met the study criteria. Median age was 65.7 years (range 59.4-73.9), 22.9% were female. 84.7% were

treated with a combination of chemotherapy and RT. Seventeen of 131 patients (13.0%) developed RP2 at a minimum follow-up of 12 months.

Table 1 shows the clinical variables included in this study, including gender, age, Karnofsky performance status (KPS), smoking history, Chronic obstructive pulmonary disease (COPD), Cardiovascular disease (CVD), Hypertension, tumor location, tumor clinical stage, gross tumor volume (GTV), whether received chemotherapy, mean lung dose (MLD), mean heart dose (MHD), and equivalent dose in 2Gy fraction (EQD2). Under univariate analysis, MLD was significantly correlated with the risk of RP2 (p-value= 0.029). Three candidate clinical variables with p-value<0.1, including MLD, Hypertension, and GTV were selected as covariates for further multivariable analysis.

Analysis of single cytokine with RP2

To identify the most influential inflammatory cytokines in predicting RP2, the associations between the risk of RP2 and cytokine expression level and rate of level change were analyzed using GLMM models. Time points (baseline, 2 and 4 weeks during RT) were defined as random effects in GLMM model. The suitable GLMM model for every single cytokine, adjusted by the three selected clinical variables (Hypertension, GTV, and MLD), was determined by the minimum AIC score. Analysis showed that two cytokines (IL-8 and CCL2)

had significantly (p-value < 0.0017, Bonferroni) lower expression levels in patients with RP2 than those without RP2 (Figure 1 and Table 2). However, none of the temporal rate of cytokine's level during RT was statistically significant for risk RP2 after multiplicity adjustment. The full data was supplied in Supplementary Table S2.

Multivariable model for predicting RP2

Based on the above analysis, the expression levels of IL-8 and CCL2 at baseline as possible early predictors of RP2, along with the three clinical variables (Hypertension, GTV, and MLD), might be the candidates for the final multivariable GLM models for predicting the risk of RP2. To guarantee the predictors following a normal distribution and orthogonality, their distributions (IL-8, CCL2, and GTV in log transform) and correlations were shown in Figure 2. It can be seen that there were no strong correlations among them, while a straightforward relationship between MLD and GTV was found. According to their p-values (Table 1) and further clinical usage of the model, MLD was retained as one predictive candidate in the final multivariable model.

The final multivariable logistic prediction model was generated by machine learning as described above. This included two cytokines (IL-8 and CCL2) and two clinical variables (Hypertension and MLD), as shown in Table 2. The nomogram for predicting risk RP2 was constructed using the GLM model, as

shown in Figure 3. To evaluate the final model's generalization performance, this model was validated in test-set which was completely independent with trainCV-set. The predictive performances on test-set were the following: AUC=0.863 (95%CI=0.676~1, p-value=0.027), accuracy=80.0%, sensitivity=100%, specificity=76.5%, PPV=42.9%, NPV=100%. On the full dataset, the final GLM model had classified performances as the following: AUC=0.881 (95%CI=0.799~0.963, p-value=1.299e-6), accuracy=82.0%, sensitivity=86.7%, specificity=82.0%, PPV=44.8%, NPV=97.3%. The ROC curve of the full dataset was compared with those based on univariate models and the model from our previous study for RILT, as shown in Figure 4.

Discussion

In this study, we studied 30 cytokines from 131 NSCLC patients enrolled in prospective clinical trials. Although our initial hypothesis was the temporal change of cytokines during RT can predict RP2 better than baseline measurements alone, the results of this study demonstrated that the expression levels of two cytokines (IL-8 and CCL2) were statistically significant for RP2, while none of the rates in the change of cytokines was significant for RP2. This has an important meaning that the individual's intrinsic micro-environment of patients, especially inflammation cytokine levels before RT, plays an important role in the following progress of RP2 in the patient with NSCLC underwent RT.

As described above previous groups[15-21,28,29] have reported that biologic factors for RP prediction. In smaller datasets, TGF-β1[15], IL-6[16,17], IL-1A[16], SOD[18], GPX[18] have shown significant p-values with RP.. Specifically, IL-6 and IL-1A had been studied for their prediction ability as reported sensitivities of 50% and 53% respectively.[16] Rs1982073 in TGF-β1, rs11466345 in TGF-β1 and rs10898880 in ATG16L2 had hazard ratios (0.489, 2.2 and 1.8 respectively). Other research groups[4,30,31] used simple clinical variables to predicting the RP progression, such as MLD (odds ratio=2.02) and lung receiving 20 Gy of radiotherapy (odds ratio=1.41). Notably, Valdes G et al.[32] used machine learning in predicting RP. They found radiation dosimetric parameters and patients' race were important features in RUSBoost algorithm, however the accuracy of their classification is limited and the algorithm is difficult for a clinical usage. We compared them in details as shown in Supplementary Table S3. While in our study, we developed and validated a predictive model for RP2 with AUC=0.863, using stringent statistical method and machine learning approach. This GLM model included IL-8 and CCL2 at baseline level and two clinical variables (MLD and Hypertension) as early predictors of RP2. We also validated in the completely independent test-set, with the model predictive values of over 80% (AUC=0.863, Sensitivity=100%, Specificity=0.765%), numerically better than those previous reports for RP.

Furthermore, since it predicts RP2 based on cytokines at the baseline, our model may provide an opportunity to personalize radiation treatment guidance before RT start. To our knowledge, this is the first study to validate that IL-8 and CCL2 as the early predictors for RP2, particularly to predict RP2 before RT start. Using the nomogram in Figure 3, the risk of RP2 can be calculated based on the patient's IL-8 and CCL2 levels, hypertension, and MLD values before RT start. Even the MLD value can be modulated to control the RP2 risk, which may contribute to patients' overall survival. Therefore, predicting the risk of RP2 before RT may provide guidance for the aggressiveness of the RT treatment or prescribing anti-inflammation treatment.

Our model is promising with a predictive accuracy of 0.86. The relatively high accuracy of our final GLM model may partially contribute to consideration of clinically important variables. MLD's contribution in RP2 is consistent with previous literatures studied in RILT.[7,8,10-12] Interestingly, hypertension was found that have the contribution to RP2. The biologic mechanism is unclear regarding the relationship between hypertension and RP2. It is possible that the progression of hypertension is associated with inflammation and fibrosis. However, whether inflammation is the cause or effect of hypertension is not well understood.[33] Of additional note, the effect of Tumor Location on RP2 was also tested here as it was previously reported in both patient[34] and

243 animal studies.[35,36] Although we considered it as a candidate predictor, tumor
244 location parameter was not significant so as to be included in multivariate
245 consideration, but not included in the final predictive GLM model. This result
246 does not necessarily imply that tumor location is insignificant, since our
247 sample included 20% tumor with unknown location. More so, this may be due
248 to the fact that the majority of our patients were stage III with some
249 component of central diseases.
250
251 It is known that cytokines play an important role in RP2. In our study, lower IL-8
252 level was found statistically significant with higher risk of RP2, which is
253 consistent with previous literature on RILT. Both of our previous studies[8,37] and
254 Hart et al.[38] had reported that low IL-8 was correlated with an increased risk of
255 RILT. IL-8 has chemotactic activity for leukocytes and induces collagen
256 synthesis and cell proliferation[39] in animal studies, but it has been consistently
257 found to have an anti-inflammatory effect in humans.[12,37,38,40] Furthermore, it
258 has been shown that neutrophils penetrate the injury site and perform the
259 critical tasks of dismantling injured vessels and creating channels for new
260 vascular regrowth, which is important for full repair of the sterile injury.[41] This
261 discovery strongly supported our results that higher level of IL-8 before RT was
262 correlated with lower risk of RP. We believe that higher level of IL-8 can
263 chemotaxis more neutrophils, to migrate toward the site of injury caused by
264 radiation. As enough neutrophils be recruited and activated in the repair

process, the progress of RP will not be happened because the radiation injury was almost being repaired.

This study is the first to demonstrate that low CCL2 level was associated with the increased risk of RP2. CCL2 was also known as involved in attracting neutrophils in animal studies. These factors may work together with IL-8 in RP process, as they are recognized in other conditions like inflammation of vascular disease. It may be reasonable to hypothesize that long-term overexpression of CCL2 in humans may play the same anti-inflammation role as IL-8. This needs to be validated and be a focus of future research.

It is interesting to note that neither the rate of change in TGF-β1 during treatment nor the baseline level of TGF-β1 was significantly associated with RP2; and its' AUC for RP2 on the dataset was 0.507, as shown in Figure 3. This was different from some previous studies that investigated plasma TGF-β1 as a predictor for RILT, including our own studies.[8,11-13,15,20,22,37,42] This controversial result may be multifactorial. Firstly, definitions of RP and RILT were not consistent as described in the introduction section, which may have confounded the results. The role of TGF-β1 may be more prominent for fibrosis.[37,43] Secondly, TGF-β1 can be produced by both tumor and normal tissues, which seriously confound its role on RP2. The insignificant results of TGF-β1 on RP2 does not override its effect on RILF or RILT. Studies with

larger numbers of events or stratified analysis with consideration of TGF-β1 effect on tumor are needed.

While this study may be somewhat limited in the number of RP2 events and number of cytokines tested, this has been corrected by Bonferroni test, a stringent methodology for correction as well as by the use of machine learning algorithms, a stringent methodology for ensuring model's generalization. Our study continues to show that inflammatory cytokines play an important role in the evolution of radiation pneumonitis and further clinical studies leveraging these relationships are warranted.

In summary, this study demonstrated that two inflammatory cytokines (IL-8 and CCL2) have strong correlation with RP2, while the temporal rate of cytokine levels had no statistical significance with RP2. According to machine learning algorithms, we established a predictive GLM model which included mean lung dose, hypertension and both IL-8 and CCL2 at baseline levels as early predictors of radiation pneumonitis. The predictive performance of the model was validated in the independent test-set with an AUC=0.863, Sensitivity=100%, and Specificity= 76.5%. This model and its nomogram, if further validated externally, can provide an opportunity of guiding personalized lung cancer treatment plan according to individual's inflammatory cytokines.

Reference

1. Tsujino K, Hirota S, Endo M, Obayashi K, Kotani Y, Satouchi M, et al. Predictive value of dosevolume histogram parameters for predicting radiation pneumonitis after concurrent chemoradiation for lung cancer. Int J Radiat Oncol Biol Phys 2003;55:110-115.

2. Kong FM, Hayman JA, Griffith KA, Kalemkerian GP, Arenberg D, Lyons S, et al. Final toxicity results of a radiation-dose escalation study in patients with non-smallcell lung cancer (NSCLC): predictors for radiation pneumonitis and fibrosis. Int J Radiat Oncol Biol Phys 2006; 65:1075-1086.

3. Kim TH, Cho KH, Pyo H Lee JS, Zo JI, Lee DH, R, et al. Dose-volumetric parameters for predicting severe radiation pneumonitis after three-dimensional conformal radiation therapy for lung cancer. Radiology 2005;235:208-215.

4. Barriger RB, Fakiris AJ, Hanna N, Yu M, Mantravadi P, McGarry RC. Dose-volume analysis of radiation pneumonitis in non-small-cell lung cancer patients treated with concurrent cisplatinum and etoposide with or without consolidation docetaxel. Int J Radiat Oncol Biol Phys 2010;78:1381-1386.

5. Phernambucq EC, Spoelstra FO, Verbakel WF, Postmus PE, Melissant CF, Maassen van den Brink KI, et al. Outcomes of concurrent chemoradiotherapy in patients with stage III non-small-cell lung cancer and significant comorbidity. Ann Oncol 2011;22:132-138.

6. Werner-Wasik M, Paulus R, Curran WJ Jr, Byhardt R. Acute esophagitis and late lung toxicity in concurrent chemoradiotherapy trials in patients with locally advanced non-small-cell lung cancer: analysis of the radiation therapy oncology group (RTOG) database. Clin Lung Cancer 2011;12: 245-251.

7. Wang W, Xu Y, Schipper M, et al: Effect of normal lung definition on lung dosimetry and lung toxicity prediction in radiation therapy treatment planning. Int J Radiat Oncol Biol Phys 2013;86: 956–963.

8. Stenmark MH, Cai XW, Shedden K, Hayman JA, Yuan S, Ritter T, et al. Combining Physical and Biologic Parameters to Predict Radiation-Induced Lung Toxicity in Patients With Non-Small-Cell Lung Cancer Treated With Definitive Radiation Therapy. Int J Radiat Oncol Biol Phys 2012;84:e217-222.

9. Cai XW, Shedden K, Ao X, Davis M, Fu XL, Lawrence TS, et al. Plasma proteomic analysis may identify new markers for radiation-induced lung toxicity in patients with non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2010;77:867-876.

10. Cai XW, Shedden KA, Yuan SH, Davis MA, Xu LY, Xie CY, et al. Baseline Plasma Proteomic Analysis to Identify Biomarkers that Predict Radiation-Induced Lung Toxicity in Patients Receiving Radiation for Non-small Cell Lung Cancer. J Thorac Oncol 2011; 6:1073-1078.

11. Zhao L, Sheldon K, Chen M, Yin MS, Hayman JA, Kalemkerian GP, et al. The predictive role of plasma TGF-beta1 during radiation therapy for radiation-induced lung toxicity deserves further study in patients with non-small cell lung cancer. Lung Cancer 2008;59:232-239.

12. Zhao L, Wang L, Ji W, Wang X, Zhu X, Hayman JA, et al. Elevation of plasma TGF-beta1 during radiation therapy predicts radiation-induced lung toxicity in patients with non-small-cell lung cancer: a combined analysis from Beijing and Michigan. Int J Radiat Oncol Biol Phys 2009;74:1385-1390.

13. Wang J, Cao J, Yuan S, Ji W, Arenberg D, Dai J, et al. Poor Baseline Pulmonary Function May Not Increase the Risk of Radiation Induced Lung Toxicity. Int J Radiat Oncol Biol Phys 2013;85:798-804.
14. Fleckenstein K, Gauter-Fleckenstein B, Jackson IL, Rabbani Z, Anscher M, Vujaskovic Z. Using biological markers to predict risk of radiation injury. Semin Radiat Oncol 2007;17:89-98.
15. Kim JY, Kim YS, Kim YK, Park HJ, Kim SJ, Kang JH, et al. The TGF-beta1 dynamics during radiation therapy and its correlation to symptomatic radiation pneumonitis in lung cancer patients. Radiat Oncol 2009;27:59-67.
16. Chen Y, Hyrien O, Williams J, Okunieff P, Smudzin T, Rubin P. Interleukin (IL)-1A and IL-6: applications to the predictive diagnostic testing of radiation pneumonitis. Int J Radiat Oncol Biol Phys 2005;62:260-266.
17. Arpin D, Perol D, Blay JY, Falchero L, Claude L, Vuillermoz-Blas S, et al. Early variations of circulating interleukin-6 and interleukin-10 levels during thoracic radiotherapy are predictive for radiation pneumonitis. J Clin Oncol 2005;23:8748-8756.
18. Park EM, Ramnath N, Yang GY, Ahn JY, Park Y, Lee TY, et al. High superoxide dismutase and low glutathione peroxidase activities in red blood cells predict susceptibility of lung cancer patients to radiation pneumonitis. Free Radic Biol Med 2007;42:280-287.
19. Wang D, Zhu J, Sun M, Li B, Wei L, Wang Z, et al. Serum superoxide dismutase, a potential predictor for radiation pneumonitis following chemoradiotherapy in non-small cell lung cancer patients. Biomarkers 2012;17:455-462.
20. Yuan X, Liao Z, Liu Z, Wang LE, Tucker SL, Mao L, et al. Single nucleotide polymorphism at rs1982073:T869C of the TGFbeta1 gene is associated with the risk of radiation pneumonitis in patients with non-small-cell lung cancer treated with definitive radiotherapy. J Clin Oncol 2009;27:3370-3378.
21. Niu X, Li H, Chen Z, Liu Y, Kan M, Zhou D, et al. A Study of Ethnic Differences in TGFβ1 Gene Polymorphisms and Effects on the Risk of Radiation Pneumonitis in Non–Small-Cell Lung Cancer. J Thorac Oncol 2012;7:1668-1675.
22. Hildebrandt MA, Komaki R, Liao Z, Gu J, Chang JY, Ye Y, et al. Genetic variants in inflammation-related genes are associated with radiation-induced toxicity following treatment for non-small cell lung cancer. 2010;PLoS One 5:e12402.
23. Mak RH, Alexander BM, Asomaning K, Heist RS, Liu CY, Su L, et al. A single-nucleotide polymorphism in the MTHFR (methylene tetrahydrofolate reductase) gene is associated with risk of radiation pneumonitis in lung cancer patients treated with thoracic radiation therapy. Cancer 2012;118:3654-3665.
24. Ellsworth SG, Rabatic BM, Chen J, Zhao J, Campbell J, Wang W, et al. Principal component analysis identifies patterns of cytokine expression in non-small cell lung cancer patients undergoing definitive radiation therapy. PLoS One 2017;12:e0183239.
25. Kong FS, Zhao L, Wang L, Chen Y, Hu J, Fu X, et al. Ensuring sample quality for blood biomarker studies in clinical trials: a multicenter international study for plasma and serum sample preparation. Transl Lung Cancer Res 2017;6:625-634.
26. Kong FS, Moiseenko V, Zhao J, Milano MT, Li L, Rimner A, et al. Organs at Risk Considerations for Thoracic Stereotactic Body Radiation Therapy: What Is Safe for Lung Parenchyma? Int J Radiat Oncol Biol Phys 2018;s0360-3016: 34014-34018.

27. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
28. Wen J, Liu H, Wang L, Wang X, Gu N, Liu Z, et al. Potentially Functional Variants of ATG16L2 Predict Radiation Pneumonitis and Outcomes in Patients with Non-Small Cell Lung Cancer after Definitive Radiotherapy. J Thorac Oncol 2018;13:660-675.
29. Li P, Wang X, Liu Z, Liu H, Xu T, Wang H, et al. Single Nucleotide Polymorphisms in CBLB, a Regulator of T-Cell Response, Predict Radiation Pneumonitis and Outcomes After Definitive Radiotherapy for Non-Small-Cell Lung Cancer. Clin Lung Cancer 2016;17:253-262.
30. Zheng Y, Shi A, Wang W, Yu H, Yu R, Li D, et al. Posttreatment Immune Parameters Predict Cancer Control and Pneumonitis in Stage I Non-Small-Cell Lung Cancer Patients Treated With Stereotactic Ablative Radiotherapy. Clin Lung Cancer 2018;S1525-7304:30351.
31. Briere TM, Krafft S, Liao Z, Martel MK. Lung Size and the Risk of Radiation Pneumonitis. Int J Radiat Oncol Biol Phys 2016;94:377-84..
32. Valdes G, Solberg TD, Heskel M, Ungar L, Simone CB 2nd. Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy. Phys Med Biol 2016;61:6105-6120.
33.Dinh QN, Drummond GR, Sobey CG, Chrissobolis S. Roles of inflammation, oxidative stress, and vascular dysfunction in hypertension. Biomed Res Int 2014;2014:406960.
34. Seppenwoolde Y, De Jaeger K, Boersma LJ, Belderbos JS, Lebesque JV. Regional differences in lung radiosensitivity after radiotherapy for non–small-cell lung cancer. Int J Radiat Oncol Biol Phys 2004;60:748–758.
35. Kang KH, Okoye CC, Patel RB, Siva S, Biswas T, Ellis RJ, et al. Complications from Stereotactic Body Radiotherapy for Lung Cancer. Cancers (Basel) 2015;7:981-1004.
36. Du S, Lockamy V, Zhou L, Xue C, LeBlanc J, Glenn S, et at. Stereotactic Body Radiation Therapy Delivery in a Genetically Engineered Mouse Model of Lung Cancer. Int J Radiat Oncol Biol Phys 2016;96:519-537.
37. Wang S, Campbell J, Stenmark MH, Zhao J, Stanton P, Matuszak MM, et al. Plasma Levels of IL-8 and TGF-β1 Predict Radiation-Induced Lung Toxicity in Non-Small Cell Lung Cancer: A Validation Study. Int J Radiat Oncol Biol Phys 2017;98:615-621.
38. Hart JP, Broadwater G, Rabbani Z, Moeller BJ, Clough R, Huang D, et al. Cytokine profiling for prediction of symptomatic radiation-induced lung injury. Int J Radiat Oncol Biol Phys 2005;63:1448-1454.
39. Ward PA, Hunninghake GW. Lung inflammation and fibrosis. Am J Respir Crit Care Med 1998;157:S123-S129.
40. Meirovitz A, Kuten M, Billan S, Roxolyana Abdah-Bortnyak, Anat Sharon, Tamar Peretz, et al. Cytokines levels, severity of acute mucositis and the need of PEG tube installation during chemoradiation for head and neck cancer: A prospective pilot study. Radiat Oncol 2010;5:16.
41. Wang J, Hossain M, Thanabalasuriar A, Gunzer M, Meininger C, Kubes P. Visualizing the function and fate of neutrophils in sterile injury and repair. Science 2017;358:111-116.
42. De Jaeger K, Seppenwoolde Y, Kampinga HH, Boersma LJ, Belderbos JS, Lebesque JV. Significance of plasma transforming growth factor-beta levels in radiotherapy for non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2004;58:1378-1387.
43. Tsoutsou PG, Koukourakis MI. Radiation pneumonitis and fibrosis: Mechanisms underlying

its pathogenesis and implications for future research. Int J Radiat Oncol Biol Phys 2006;66:1281-1293.

Tables

Table 1. Association between clinical variables and RP2

Characteristic		All patients n (%)	Without RP2 n (%)	With RP2 n (%)	p-value
Gender $	male	101 (77.1%)	87 (76.3%)	14 (82.4%)	0.761
	female	30 (22.9%)	27 (23.7%)	3 (17.6%)	
Age #	Median	65.7	65.7	66.01	0.911
	1st – 3rd Qu	(59.4 – 73.9)	(59.4 – 73.5)	(60.1 - 77.2)	
KPS #	Median	90.0	90.0	90.0	0.328
	1st – 3rd Qu	(80.0 - 90.0)	(80.0 - 90.0)	(85.0 - 90.0)	
Smoking status $	Never	6 (4.6%)	5 (4.4%)	1 (5.9%)	0.573
	Current/former	125 (95.4%)	109 (95.6%)	16 (94.1%)	
COPD &	No	72 (55.0%)	62 (54.4%)	10 (58.8%)	0.935
	Yes	59 (45.0%)	52 (45.6%)	7 (41.2%)	
CVD &	No	86 (65.6%)	77 (67.5%)	9 (52.9%)	0.363
	Yes	45 (34.3%)	37 (32.5%)	8 (47.1%)	
Hypertension $	No	61 (46.6%)	57 (50.0%)	4 (23.5%)	0.066 •
	Yes	70 (53.4%)	57 (50.0%)	13 (76.5%)	
GTV (cm^3) #	Median	135.5	126.2	186.3	0.088 •
	1st – 3rd Qu	(65.7 – 268.0)	(57.1 – 251.5)	(109.6 – 337.6)	
Tumor location $	Central	80 (61.1%)	67 (58.8%)	13 (76.5%)	0.380
	Peripheral	23 (17.6%)	22 (19.3%)	1 (5.9%)	
	Unknown	28 (21.4%)	25 (21.9%)	3 (17.6%)	
Clinical stage $	I	15 (11.4%)	13 (11.4%)	2 (11.8%)	1
	II	12 (9.2%)	11 (9.7%)	1 (5.8%)	
	III	104 (79.4%)	90 (78.9%)	14 (82.4%)	
MLD (Gy) #	Median	17.2	16.9	18.8	0.029 *
	1st – 3rd Qu~	(14.2 – 19.8)	(13.6 – 19.3)	(17.5 – 20.8)	
MHD (%) #	Median	11.3	10.8	16.9	0.194
	1st – 3rd Qu~	(4.5 – 19.1)	(4.3 – 18.6)	(9.6 – 19.8)	
EQD2 (Gy) #	Median	70.0	70.0	76.2	0.308
	1st – 3rd Qu~	(65.0 - 78.0)	(65.0 - 77.9)	(65.0 - 81.9)	
Chemotherapy $	No	20 (15.3%)	17 (14.9%)	3 (17.6%)	0.724
	Yes	111 (84.7%)	97 (85.1%)	14 (82.4%)	

Abbreviation: KPS = Karnofsky Performance Status; COPD = Chronic obstructive pulmonary disease; CVD = Cardiovascular disease; GTV = gross tumor volume; MLD = mean lung dose; MHD = mean heart dose; EQD2 = equivalent dose in 2 Gy fractions;

$ Fisher's exact test; & Chi-square test; # logistic regression; ~QU=quartile

* p-value < 0.05; • < 0.1

Table 2. Single cytokines' level and the final GLM model for predicting risk RP2

variables		Estimate coefficient	Standard error	Odds ratio [#]	confidence interval 95%	p-value
Single Cytokines	IL-8	-1.177	0.281	0.308	0.178 – 0.535	2.8e-5 **
	CCL2	-0.979	0.295	0.376	0.211 – 0.669	8.9e-4 **
The final GLM model	(Intercept)	-0.569	3.038	0.566	0.001 – 226.743	0.851
	IL-8	-0.887	0.371	0.412	0.183 – 0.810	0.017 *
	CCL2	-1.190	0.471	0.304	0.107 – 0.734	0.011 *
	Hypertension	1.993	1.086	7.337	1.128 – 90.945	0.066
	MLD	0.321	0.161	1.378	1.058 – 2.017	0.046 *

[#] Ratio of levels in patient with RP2 / without RP2

* p-value < 0.05; ** < 0.0017 (Bonferroni)

Figure Legends

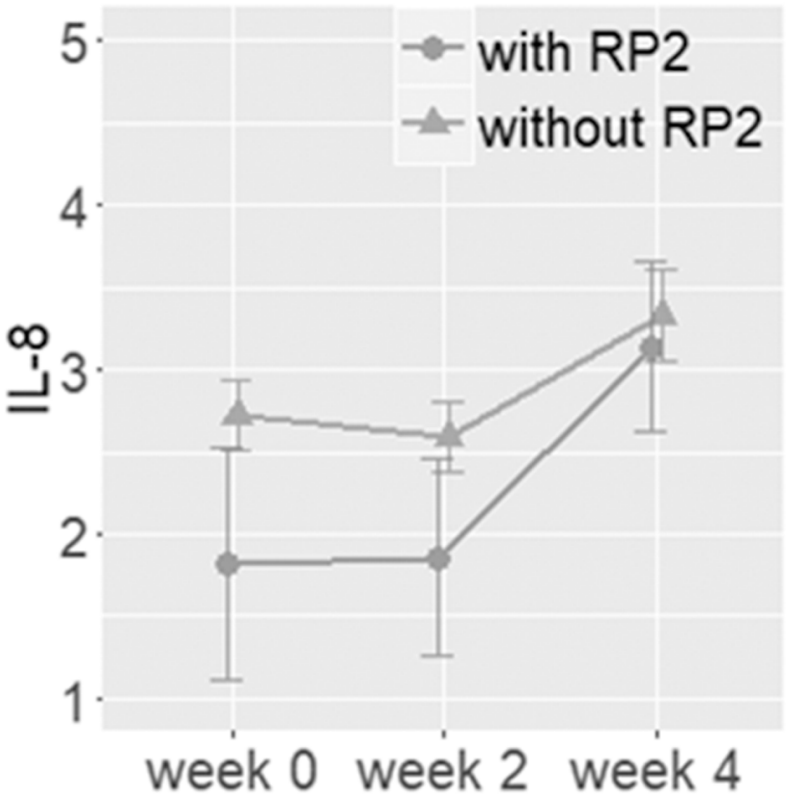
Figure 1. Temporal data of cytokine expression levels: (left) IL-8, (right) CCL2 (means ± 95% confidence interval)

Figure 2. Visualization of the continuous predicting candidates' distribution and correlation. Corr= Pearson Correlation Coefficient. (x- and y-coordinate are predictors' level).

Figure 3. The nomogram for risk RP2, constructed based on the final GLM model. It based on two cytokines (IL-8 and CCL2) and two clinical variables (Hypertension, MLD).

Figure 4. ROC curves for risk RP2 on the whole dataset, comparing the final GLM model with single cytokines (IL-8, CCL2, TGF-β1); and comparing with IL-8+TGF-β1+MLD, which were referenced from our previous study for predicting RILT.[8]

Figure 1



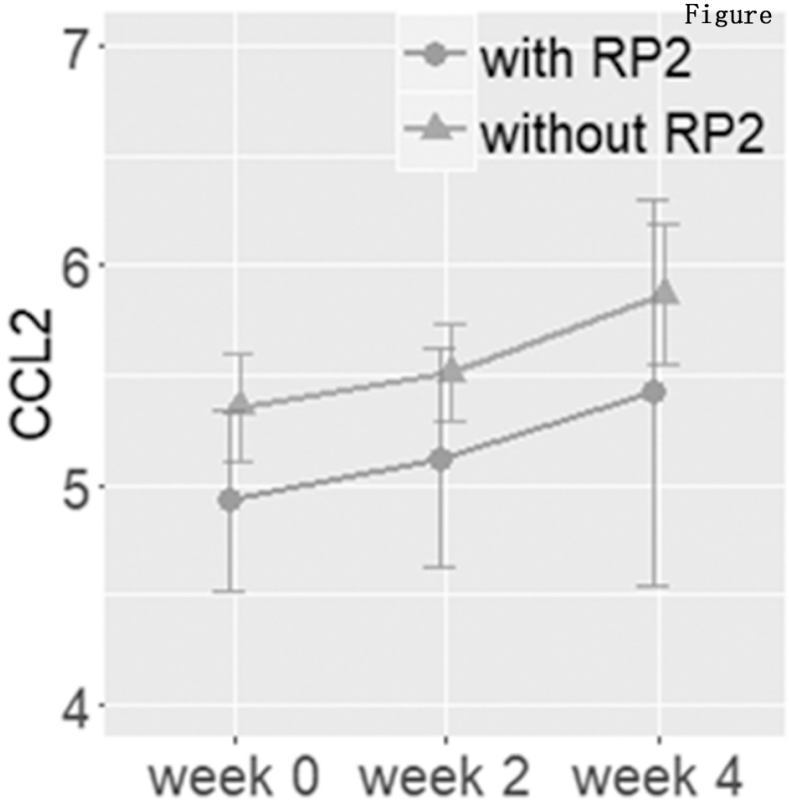


Figure 2

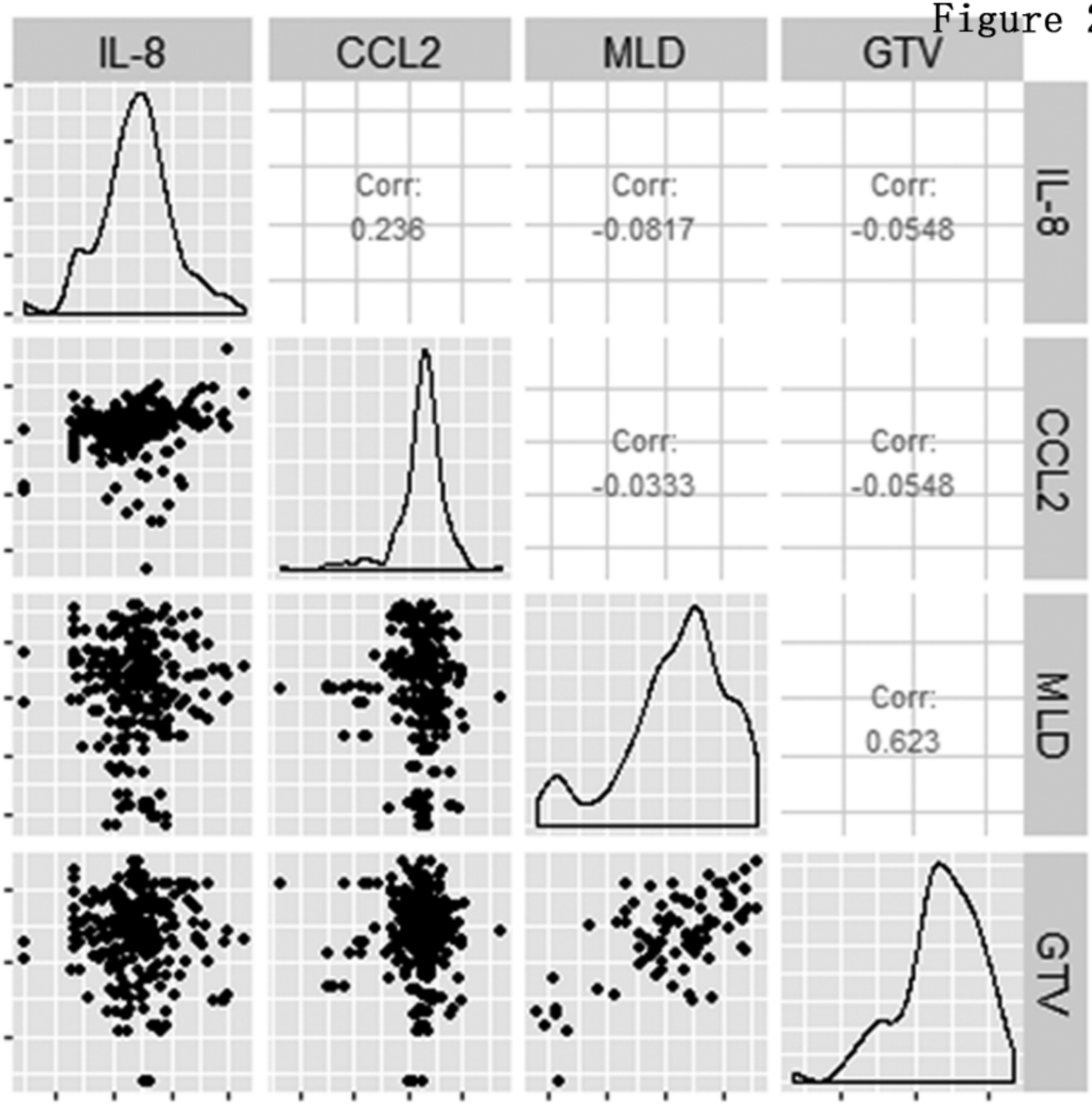


Figure 3

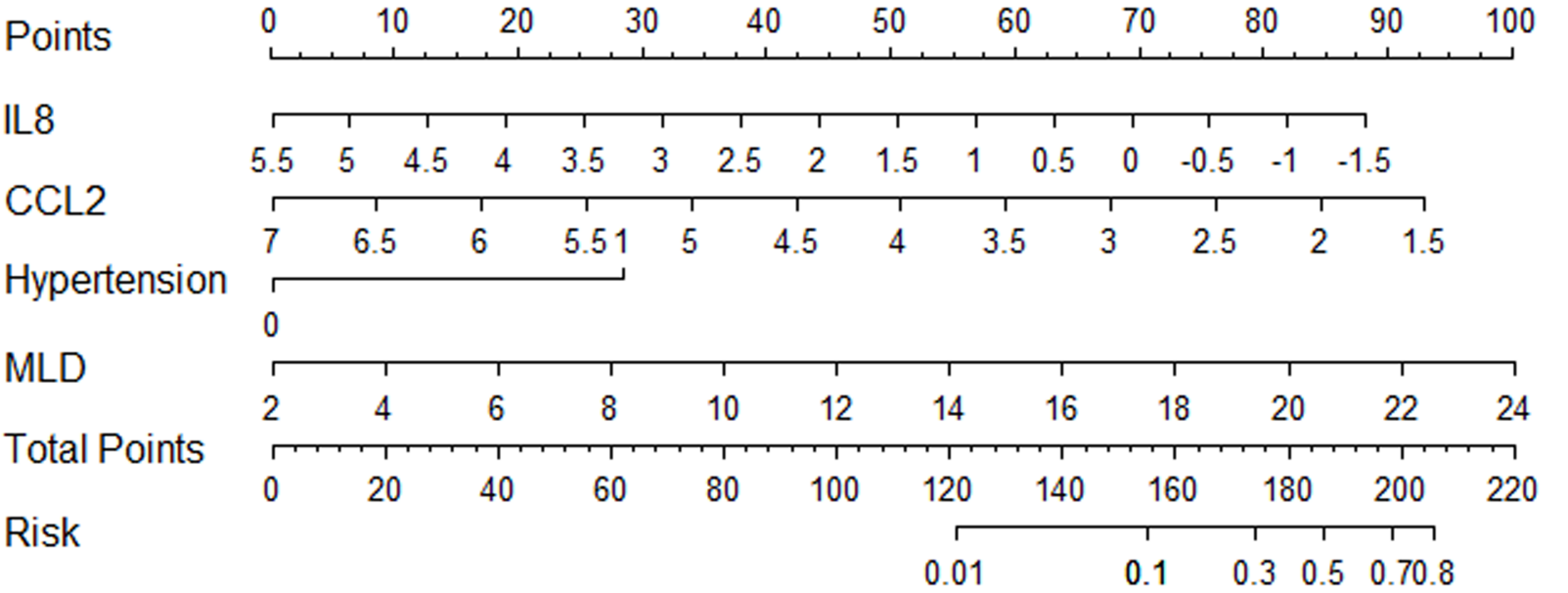


Figure 4