Supplementary Methods

PD-L1 immunohistochemistry (IHC)

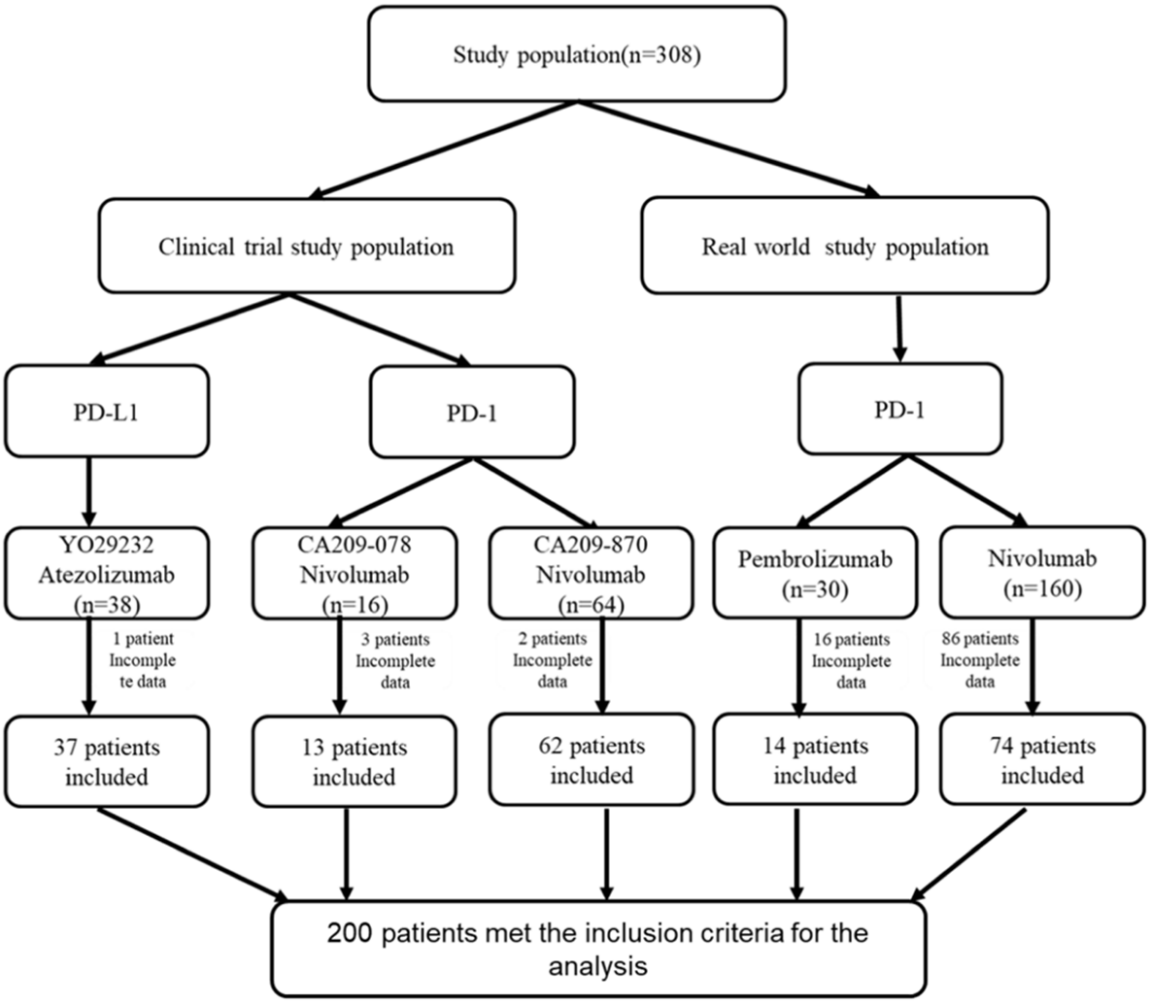
Our pathology department evaluated PD-1 ligands (PD-L1) in tumor biopsy specimens prior to the first immunotherapy by using a validated automated immunohistochemical (IHC) assay (Nivolumab: PD-L1 IHC 28-8 pharmDx, Dako; Pembrolizumab: PD-L1 IHC 22C3 pharmDx, Dako). Predefined expression levels were defined by tumor cell membrane staining (TPS) in a section containing at least 100 evaluable tumor cells. PD-L1 test data from 102 patients met the quality control requirements, including 97 patients using Nivolumab and 5 patients using Pembrolizumab.

Data extraction, nodule labeling and features

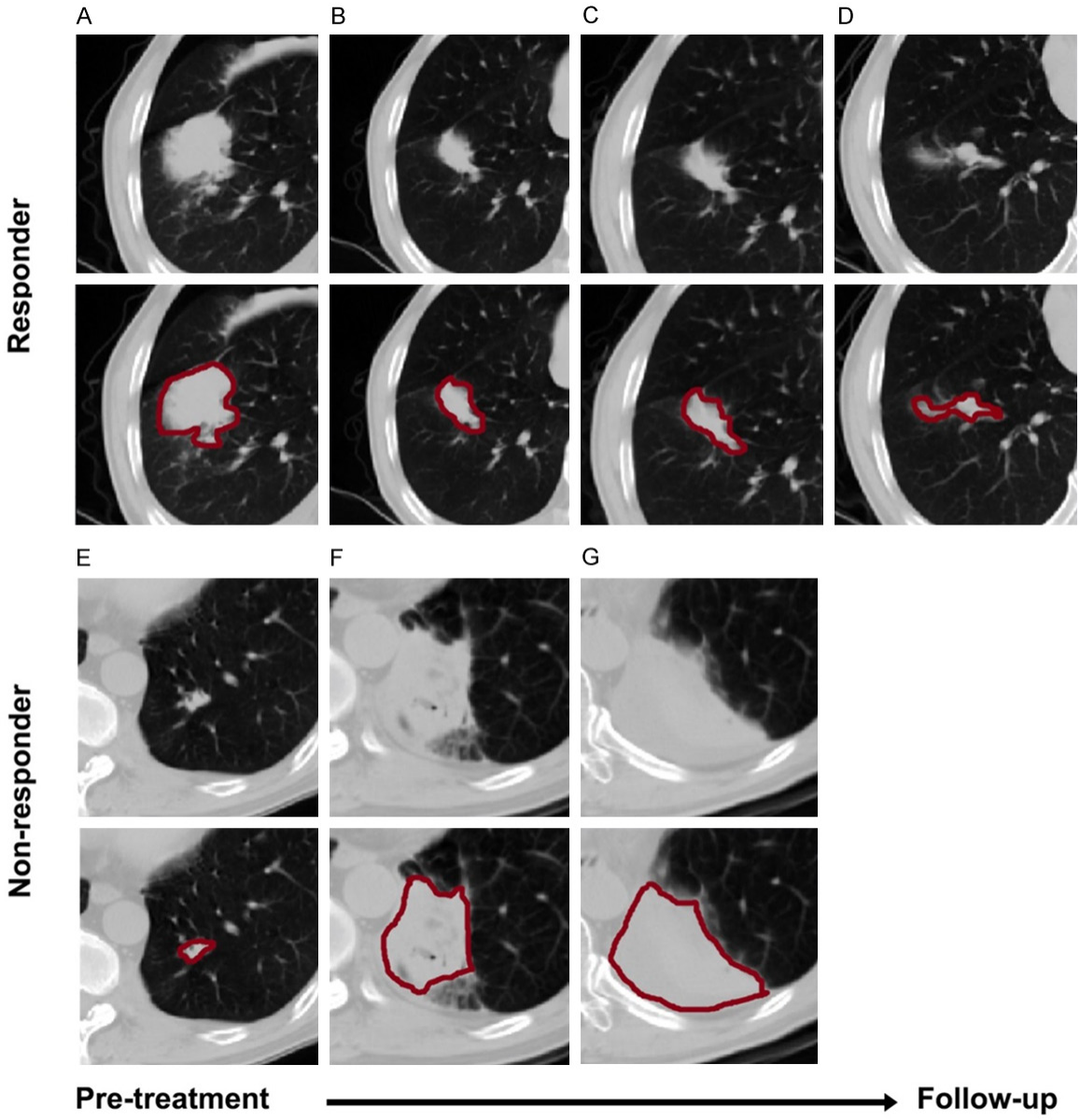
A 64-layer LightSpeed Volume CT (GE Healthcare, WI, USA) was used for chest CT. All imaging data were reconstructed using a medium-sharp reconstruction algorithm with a thickness of 3-5 mm. All nodules were manually labeled by two oncologists.

Using medical image processing and navigation software 3D Slicer (version 4.8.0, Brigham and Women's Hospital, MA, USA), a radiologist manually segmented the volume of interest (VOI) of the largest lesion in the lung at the voxel level (Lan Shen, with 10 years of work experience). Then, another oncologist Yi Yang (with 20 years of work experience) confirmed the VOI. An example is shown in Supplementary Figure 2. Large blood vessels and bronchioles were excluded from the volume of the nodule as much as possible. The images of lung computed tomography (CT) digital medical imaging and communication format were imported into the software for drawing, and then the images with VOI information were extracted in Neuroimaging Informatics Technology Initiative (NII) format for further analysis.

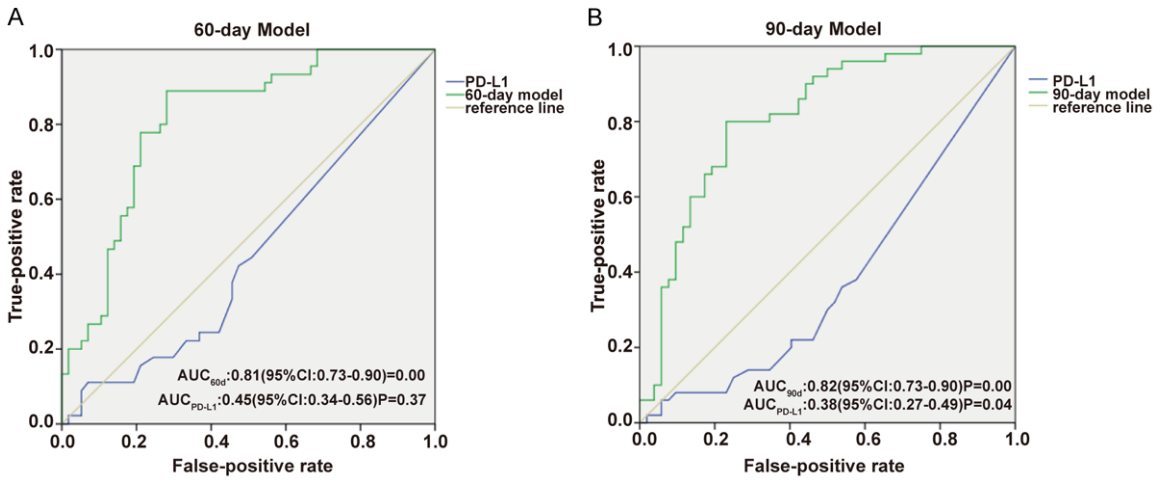
Due to limited CT scans, radiomics was used to represent the radiographic features, instead of end-to-end convolutional neural networks. The radiomics were extracted with PyRadiomics (Python 3.7.3, PyRadiomics 2.2.0). For each lesion, a radiomic feature of 107 dimensions was extracted from the CT scan. The radiomic features were used as serial inputs to the module if CT examinations were conducted more than once. The clinical and blood test information was categorized as serial and static. By filtering out the fields with more than 10% missing information, the serial blood test features were of 22 dimensions and the static features were of 18 dimensions, where categorical features were one-hot encoded. Numeric features were normalized to enable more stable and better convergence for model training.



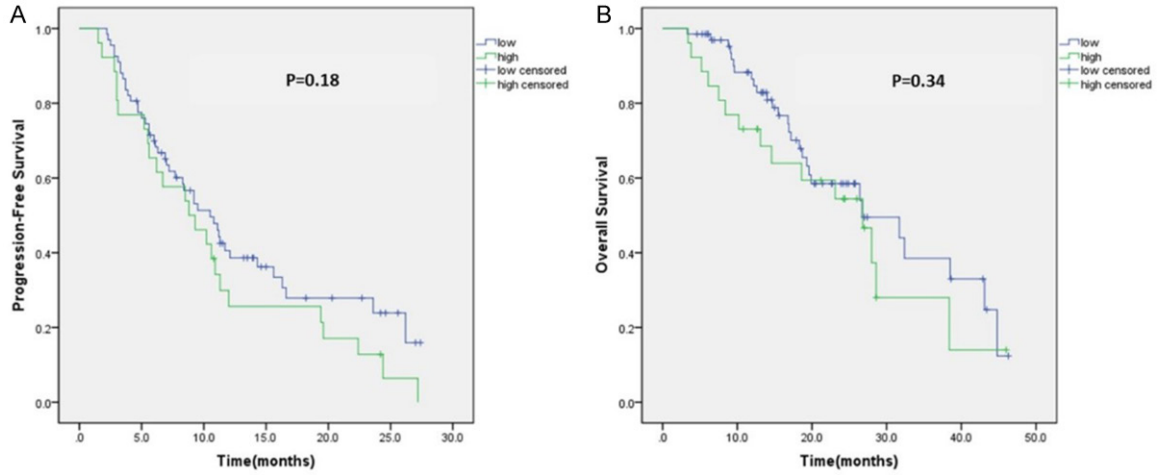
Supplementary Figure 1. The study flowchart.



Supplementary Figure 2. Serial CT scans of representative patients. A-D. CT images of a responder (patient 11) in the training cohort; E-G. CT images of a nonresponder (patient 95) in the validation cohort. The red lines indicate labeled pulmonary nodules. CT, Computed tomography.



Supplementary Figure 3. Model performance for response prediction in 102 patients. A. ROC AUC for the 60-day response model and PD-L1 expression. B. ROC AUC for the 90-day response model and PD-L1 expression. AUC, Area under the curve; ROC, receiver operating characteristic.



Supplementary Figure 4. Deep learning prediction of survival in 93 patients with confirmed stable disease (SD) at the first efficacy assessment after anti-PD-1/PD-L1 treatment. The 60-day prediction model stratified patients with SD into high- and low-risk nonresponse groups using a default cutoff. A. Progression-free survival in relation to risk stratification. B. Overall survival in relation to risk stratification. Tumor response was evaluated according to RECIST 1.1.