

Figure 1: Results on the PF dataset. **a**, Heatmap showing the effect from covariates, with darker color representing a stronger relationship between the cell type and covariate. **b**, Heatmap of the estimated regression coefficients, where red stands for a positive correlation. **c**, 95% credible interval for the significant β_{rj} chosen by the threshold controlling the FDR < 0.05. **d**, 95% credible interval for the significant β_{rju} of the tree structure chosen by the threshold controlling the FDR < 0.05. **e**, Cladogram of the identified cell types, with red dots representing the nodes of cell types have significantly higher abundance when PF exists. **f**, The underlying barplot represents the predicted proportion (α) of the certain cell type by the model, and the black dots are the real proportions of this cell type in each patient of the dataset.

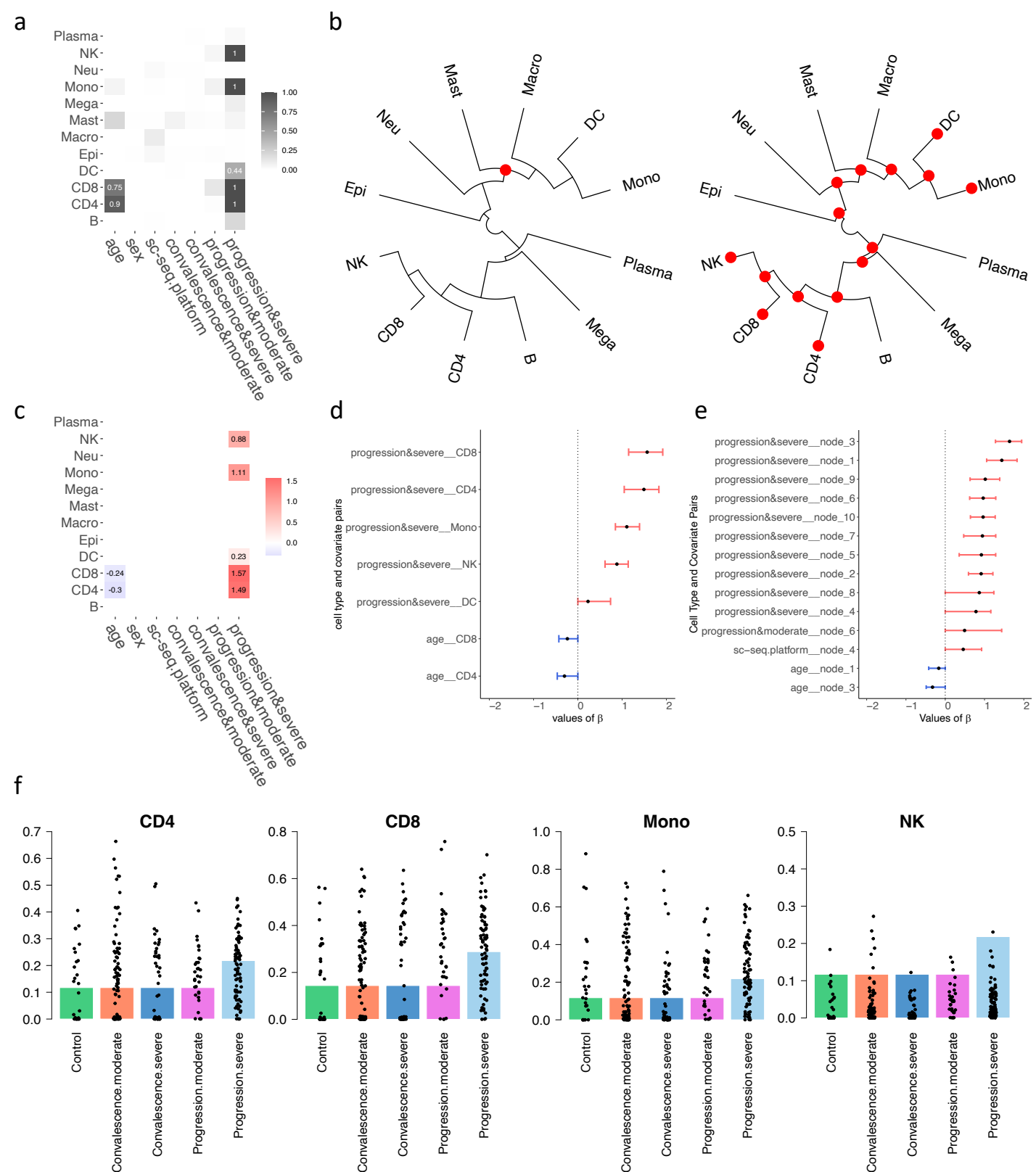


Figure 2: Results on the Covid-19 dataset. **a**, Heatmap showing the effect from covariates, with darker color representing a stronger relationship between the cell type and covariate. **b**, Cladogram of the identified cell types with sample time during “progression” and different severity (left is moderate and right is severe), where red dots represent the nodes of cell types have significantly higher abundance and blue represent a lower abundance. **c**, Heatmap of the estimated regression coefficients, where red stands for a positive correlation, and blue represents a negative correlation. **d**, 95% credible interval for the significant β_{rj} chosen by the threshold controlling the FDR < 0.05. **e**, 95% credible interval for the significant β_{rju} of the tree structure chosen by the threshold controlling the FDR < 0.05. **f**, The underlying barplot represents the predicted proportion (α) of the certain cell type by the model, and the black dots are the real proportions of this cell type in each patient of the dataset.

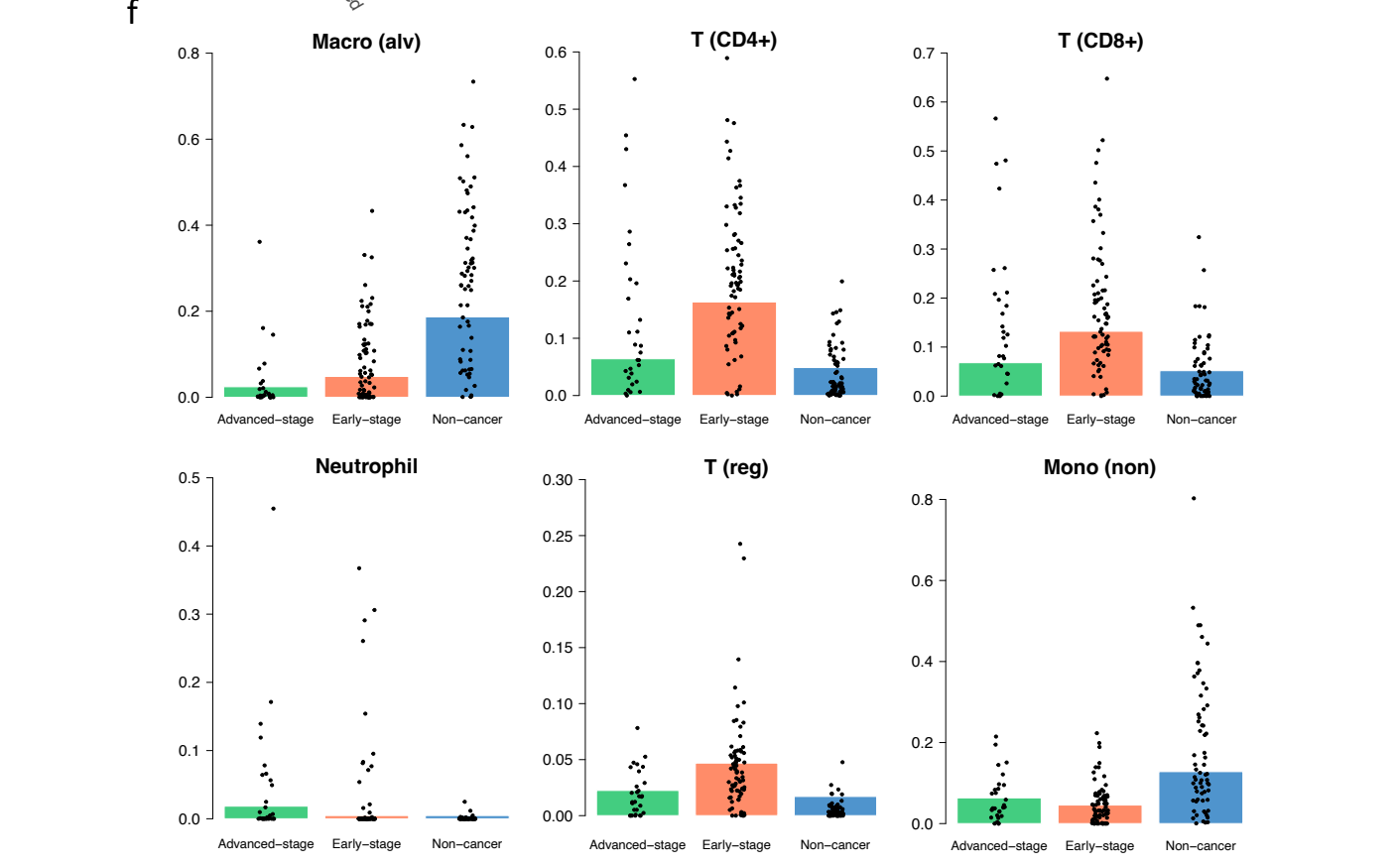
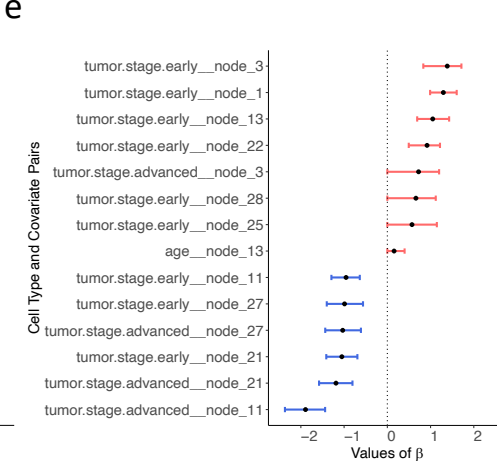
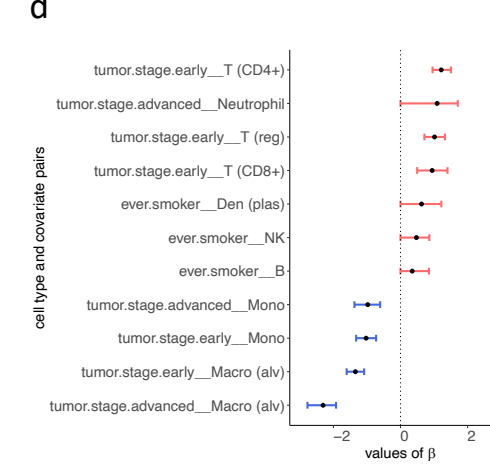
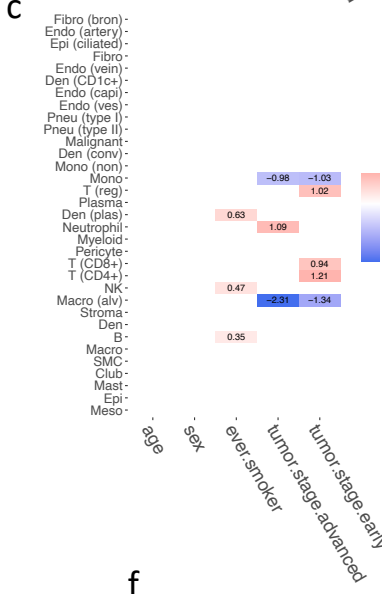
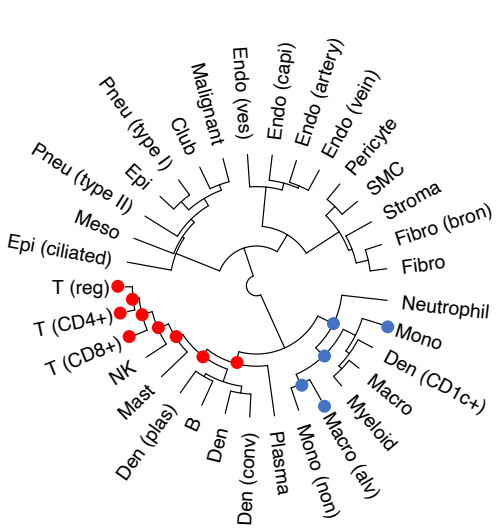
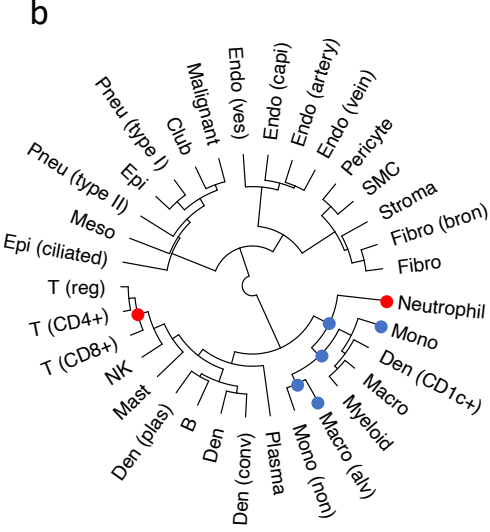
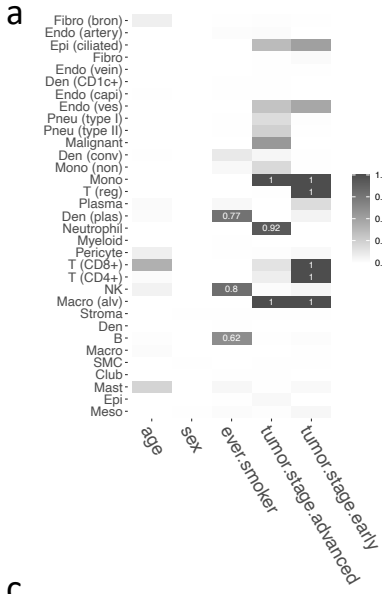


Figure 3: Results on the Lung Cancer dataset. **a**, Heatmap showing the effect from covariates, with darker color representing a stronger relationship between the cell type and covariate. **b**, Cladogram of the identified cell types where tumor stage is advanced (left) and early (right). Red dots represent the nodes of cell types have significantly higher abundance and blue represent a lower abundance. **c**, Heatmap of the estimated regression coefficients, where red stands for a positive correlation, and blue represents a negative correlation. **d**, 95% credible interval for the significant β_{rj} chosen by the threshold controlling the $FDR < 0.05$. **e**, 95% credible interval for the significant β_{rj_u} of the tree structure chosen by the threshold controlling the $FDR < 0.05$. **f**, The underlying barplot represents the predicted proportion (α) of the certain cell type by the model, and the black dots are the real proportions of this cell type in each patient of the dataset.

supplementary

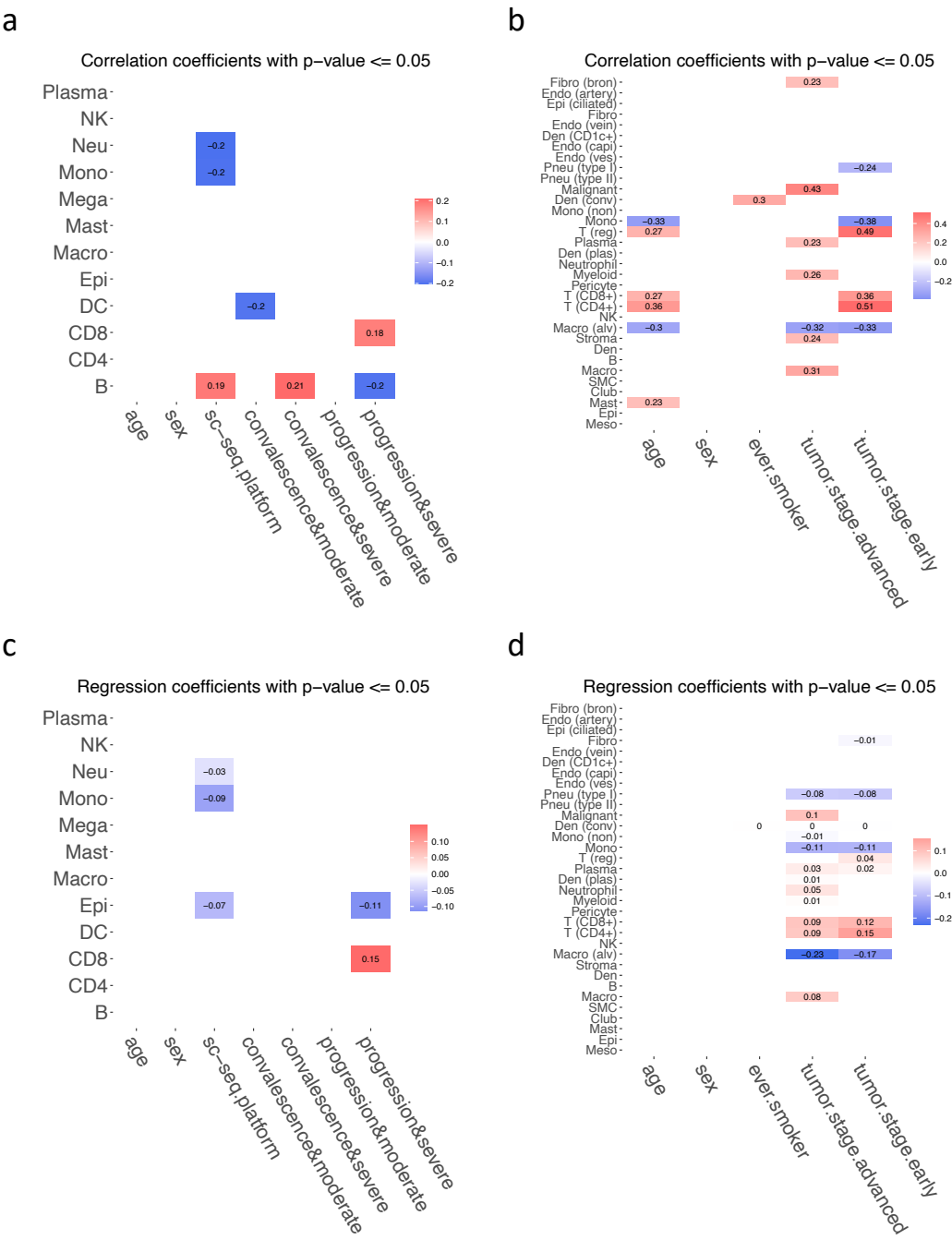


Figure 4: Results of two alternative methods on the Covid-19 and Lung Cancer dataset. a b are results by correlation test and **c d** are results by simple linear regression.