

Figure S1 Overall survival after lung cancer diagnosis, stratified by smoking status, among 986 lung cancer patients recruited from central and eastern Europe. Survival model was weighted by the smoking prevalence in the original cohort. P value represents the log-rank test for comparing the Kaplan-Meier curves.

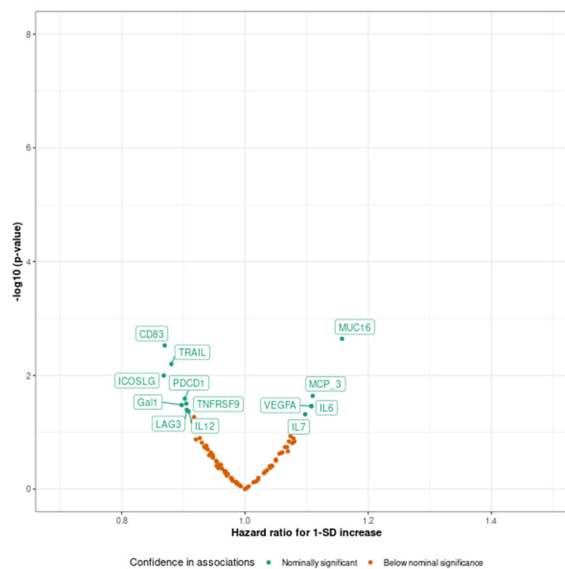


Figure S2 Associations between protein concentrations and 5-year disease progression (including recurrence, metastasis, and death whichever first) after lung cancer, among 986 lung cancer patients recruited from central and eastern Europe. Models were weighted by smoking prevalence from the full cohort and adjusted age at diagnosis, sex, smoking status (never, ever), education (elementary, high school, university and above), alcohol intake (never, ever), histology (adenocarcinoma and squamous cell carcinoma), and stage (continuous).

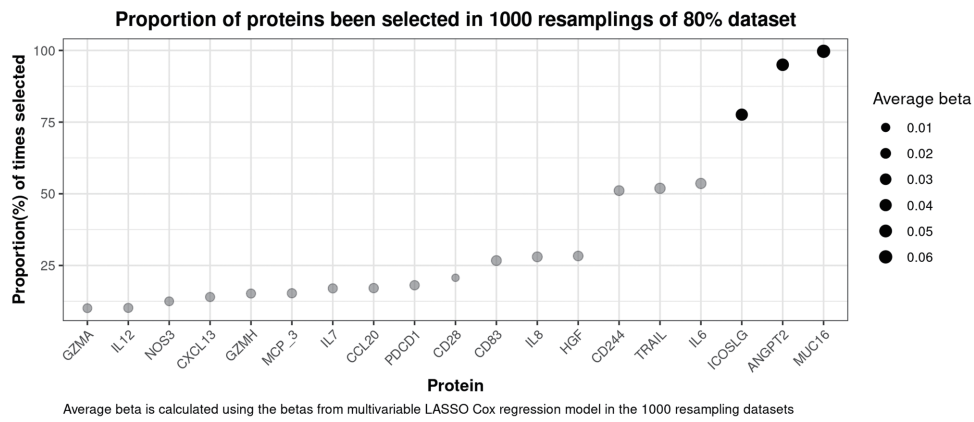


Figure S3 The proportion of proteins been selected for overall mortality for lung cancer patients by LASSO Cox model, with 1,000 bootstrapping.

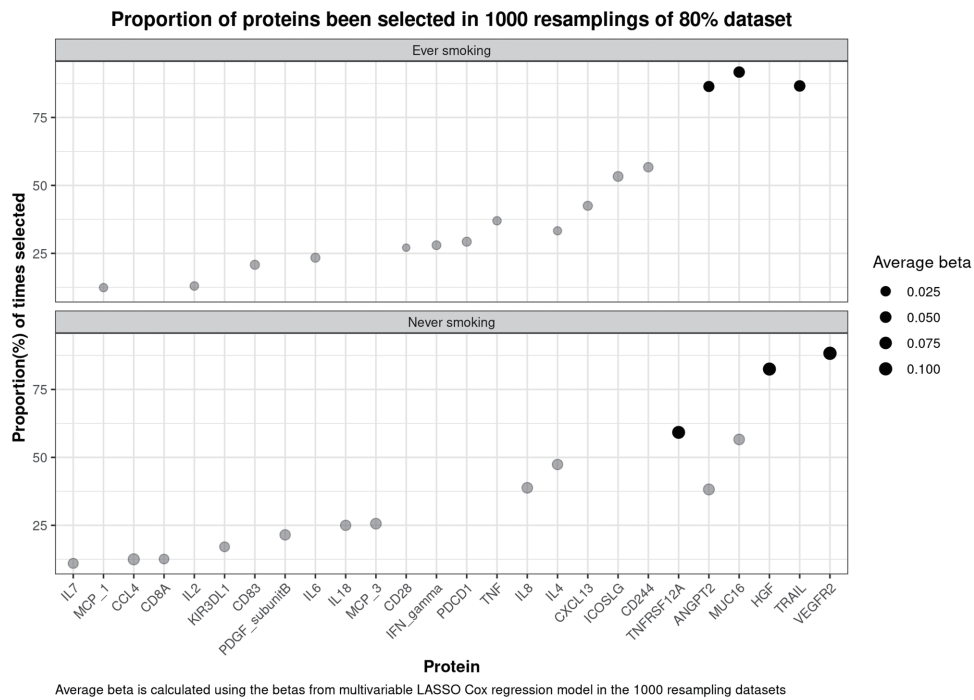


Figure S4 The proportion of proteins been selected for overall mortality for ever and never smoking lung cancer patients by LASSO Cox model, with 1,000 bootstrapping.

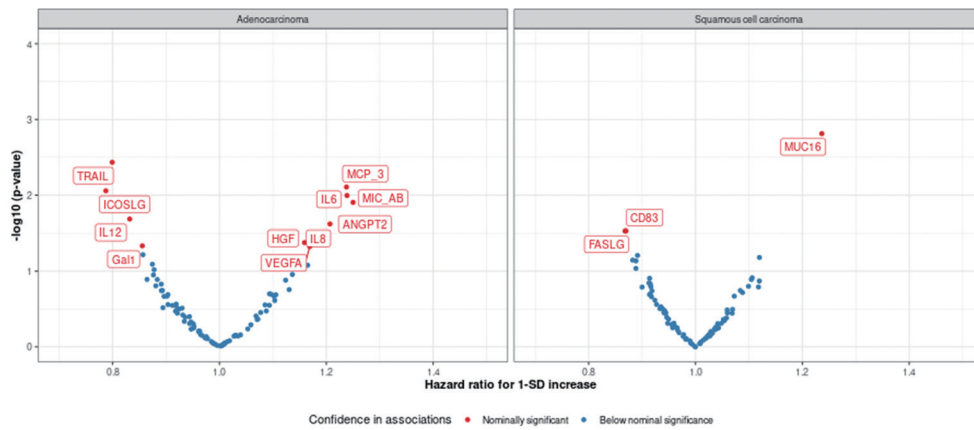


Figure S5 Associations between protein concentrations and overall mortality after lung cancer by histology. Models were weighted by smoking prevalence from the full cohort and adjusted age at diagnosis, sex, education (elementary, high school, university and above), alcohol intake (never, ever), smoking status (never, ever), and stage (continuous). There were no statistically significant associations after controlling for multiple testing.

Table S1 Hazard ratios and 95% CI of predictors in the integrated models for survival after lung cancer diagnosis

Factors	HR (95% CI)		
	All patients (N=986)	Ever smokers (N=742)	Never smoking (N=244)
ICOSLG	0.84 (0.76–0.93)		
MUC16	1.16 (1.06–1.26)	1.15 (1.04–1.27)	
ANGPT2	1.15 (1.05–1.26)	1.09 (0.98–1.20)	
TRAIL		0.85 (0.77–0.94)	
VEGFR2			0.69 (0.53–0.89)
HGF			1.23 (0.99–1.53)
TNFRSF12A			1.50 (1.12–2.00)
Age at diagnosis (per 5-year)	1.11 (1.05–1.17)	1.10 (1.03–1.17)	1.00 (0.89–1.13)
Sex			
Female	Ref	Ref	Ref
Male	1.25 (0.95–1.64)	1.24 (0.90–1.72)	1.08 (0.65–1.81)
Alcohol intake			
Never	Ref	Ref	Ref
Ever	1.23 (1–1.5)	1.32 (1.06–1.65)	1.08 (0.55–2.09)
TNM stage	1.31 (1.23–1.40)	1.30 (1.21–1.39)	1.46 (1.26–1.68)
Histology			
Adenocarcinoma	Ref	Ref	Ref
Squamous cell carcinoma	1.01 (0.81–1.25)	0.89 (0.72–1.11)	2.57 (1.53–4.31)
Smoking status			
Never	Ref		
Ever	1.13 (0.84–1.53)		

TNM stage was treated as continuous variable, and values from 1 to 5 represents stage IA, IB, IIA, IIB, and IIIA. HR, hazard ratio; CI, confidence interval.

Table S2 Assessment of the integrated models for survival after lung cancer diagnosis, and the utility of protein measurements beyond clinical factors

Factors	C-index (95% CI)		
	All patients (N=986)	Ever smokers (N=742)	Never smoking (N=244)
Apparent C-index for clinical model	0.66 (0.64 to 0.68)	0.63 (0.61 to 0.66)	0.70 (0.64 to 0.75)
Apparent C-index for integrated model	0.68 (0.65 to 0.70)	0.65 (0.62 to 0.68)	0.74 (0.69 to 0.80)
Bootstrap C-index for clinical model	0.65 (0.61 to 0.70)	0.63 (0.57 to 0.68)	0.68 (0.56 to 0.80)
Bootstrap C-index for integrated model	0.67 (0.62 to 0.71)	0.64 (0.58 to 0.70)	0.72 (0.59 to 0.82)
Bootstrap C-index difference, median (95% CI)	0.02 (–0.01 to 0.04), P=0.20	0.02 (–0.02 to 0.04), P=0.20	0.03 (–0.06 to 0.10), P=0.28

Bootstrap C-indices were estimated by fitting the models in a random 80% of patients and validating in the remaining 20% of patients, with 1,000 random 80%/20% splits of the dataset.