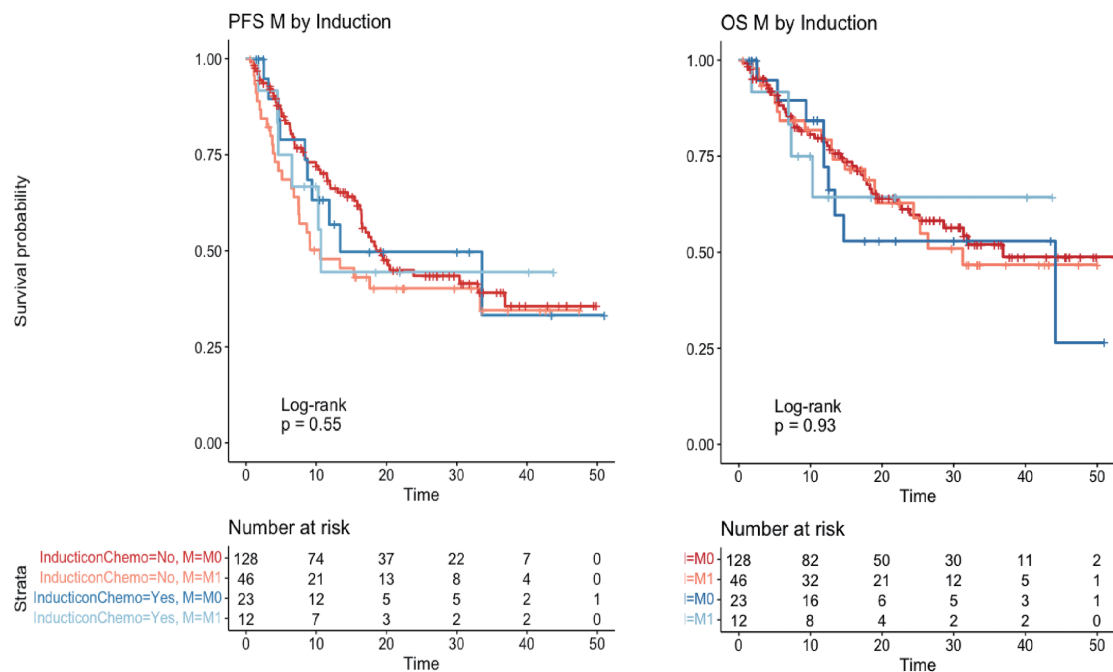
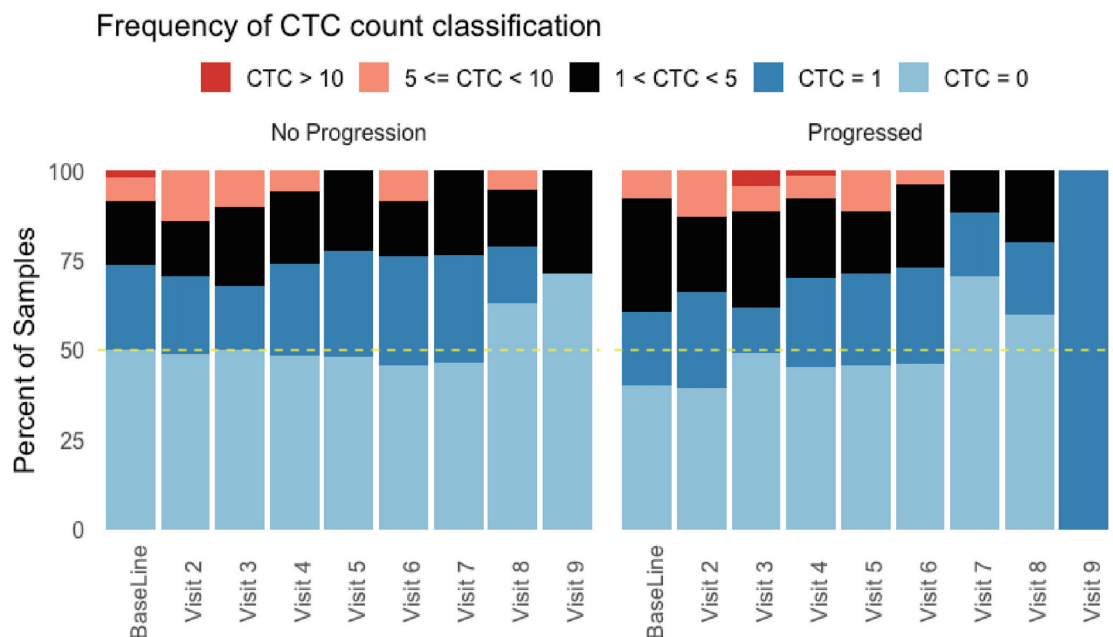


**Table S1** Genes included in the study

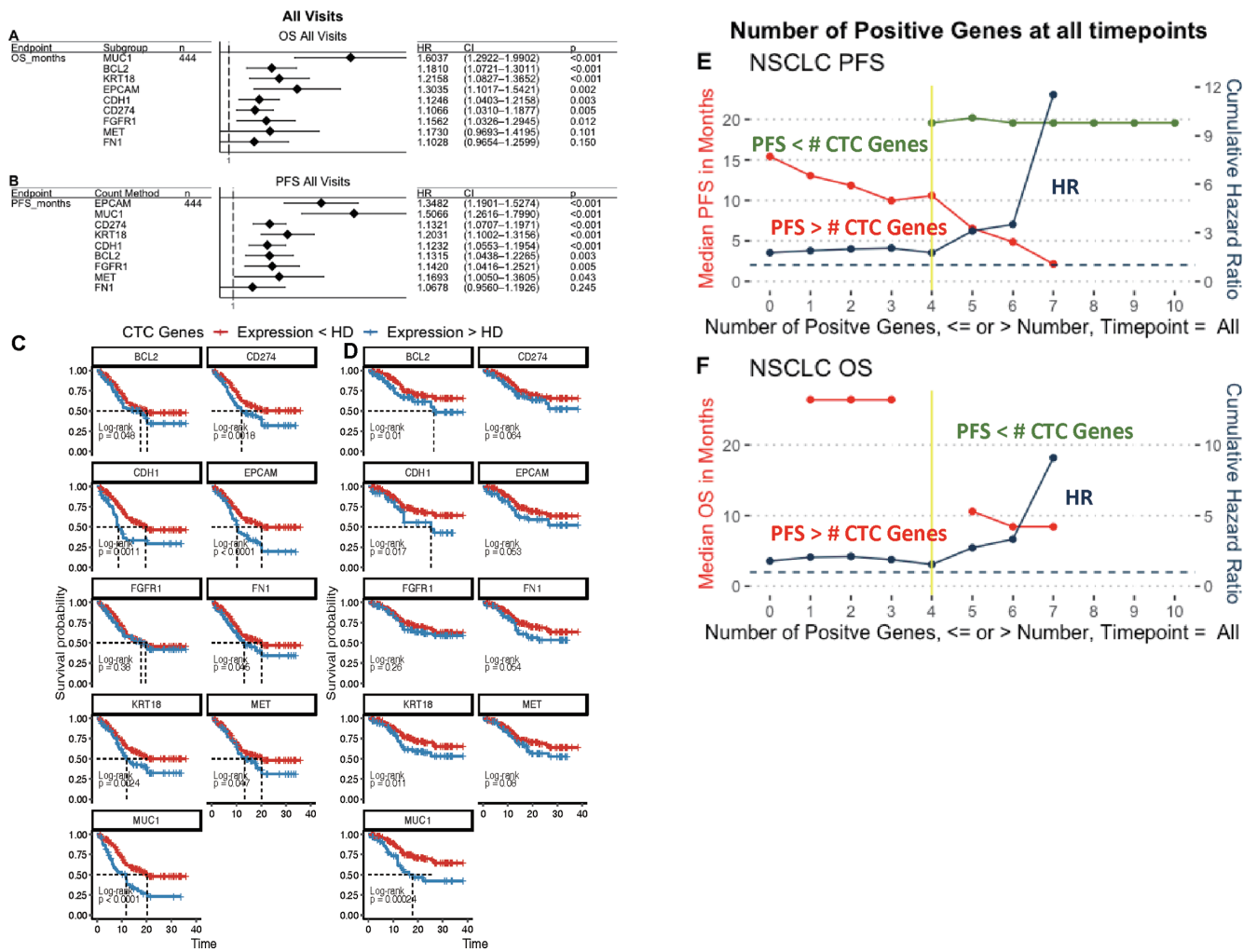
Housekeeping genes	
B2M:	Beta-2-Microglobulin (Beta Chain of MHC Class I Molecules) (qHsaCID0015347)
GAPDH:	Glyceraldehyde 3-phosphate dehydrogenase (qHsaCED0038674)
HPRT:	Hypoxanthine guanine phosphoribosyl transferase (qHsaCID0016375)
PTPRC/CD45:	Protein Tyrosine Phosphatase, Receptor Type C (white blood cell control) (qHsaCED0038908)
GYPA:	Glycophorin A (a red blood cell control gene) (qHsaCID0010750)
Epithelial-related genes	
CDH1:	Cadherin 1 (Epithelial (E)-Cadherin) (qHsaCID0015365)
EGFR:	Epidermal Growth Factor Receptor (qHsaCID0007564)
EPCAM:	Epithelial Cell Adhesion Molecule (qHsaCED0043827)
KRT7:	Keratin 7 (basic, low molecular weight cytokeratin) (qHsaCED0038533)
KRT18:	Keratin 18 (acidic, low molecular weight cytokeratin) (qHsaCED0035037)
MUC1:	Mucin 1 (Cell Surface Associated) (qHsaCED0019841)
Mesenchymal and EMT-related genes	
AXL:	AXL Tyrosine-protein kinase receptor (qHsaCID0008470)
FN1:	Fibronectin 1 (qHsaCID0012349)
SNAI2:	Snail Family Transcriptional Repressor 2 (SLUG) (qHsaCID0011342)
Other cancer-related genes	
ALDH1A1:	Aldehyde Dehydrogenase 1 Family Member A1 (qHsaCID0018574)
BCL2:	B-cell lymphoma 2 (qHsaCED0057245)
CD274/PD-L1:	Programmed Cell Death 1 Ligand 1 (qHsaCID0036468)
ERBB2/HER2:	Receptor Tyrosine Kinase 2 (Human Epidermal Growth Factor Receptor 2) (qHsaCED0045039)
FGFR1:	Fibroblast Growth Factor Receptor 1 (qHsaCED0042405)
MET:	Tyrosine-protein kinase Met/hepatocyte growth factor receptor (HGFR) (qHsaCED0002004)



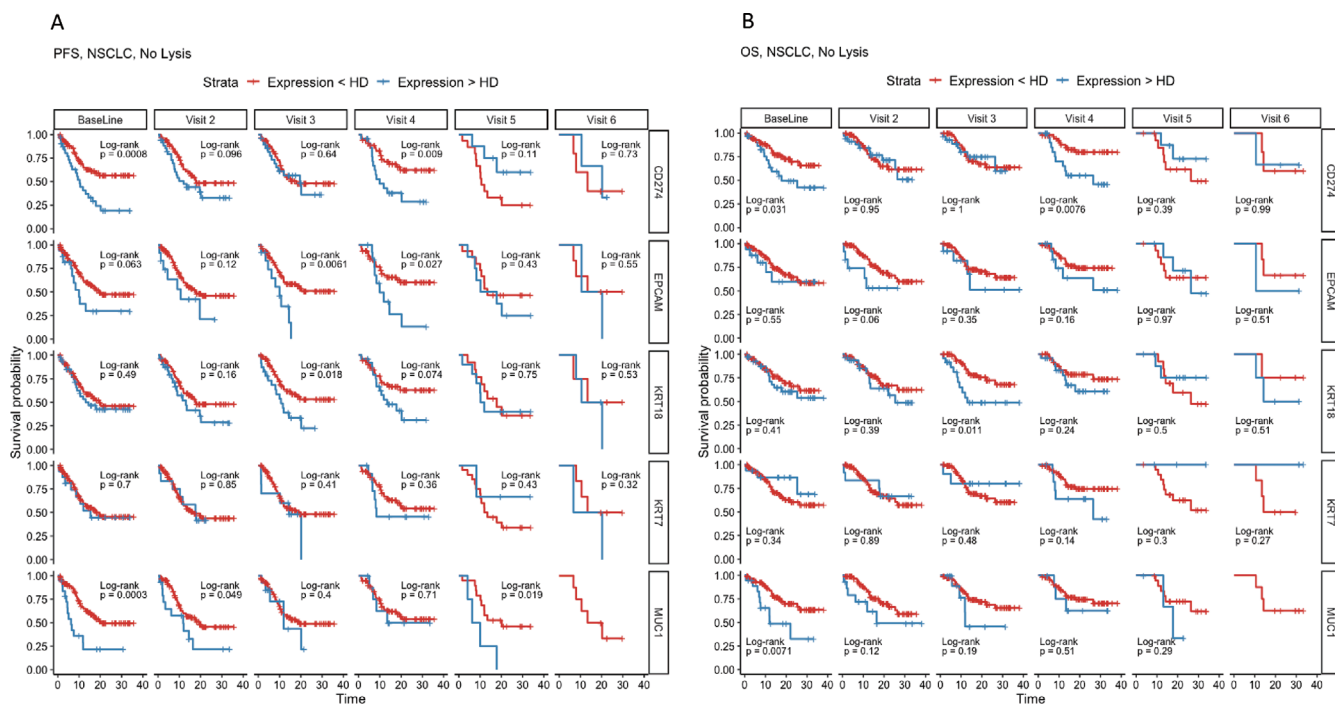
**Figure S1** Survival independent of stage at enrollment. There is no difference in survival between patients with metastatic disease and those without since some patients were enrolled after positive response to induction chemotherapy. M, metastatic status; OS, overall survival; PFS, progression-free survival.



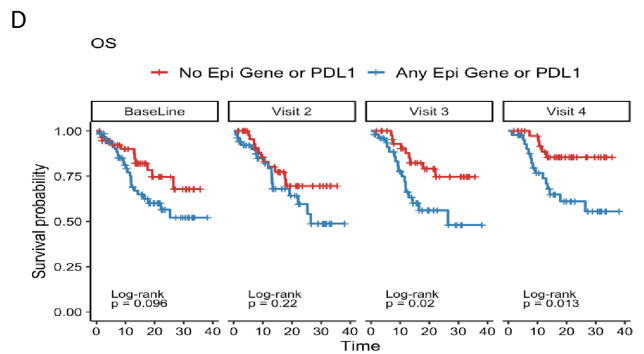
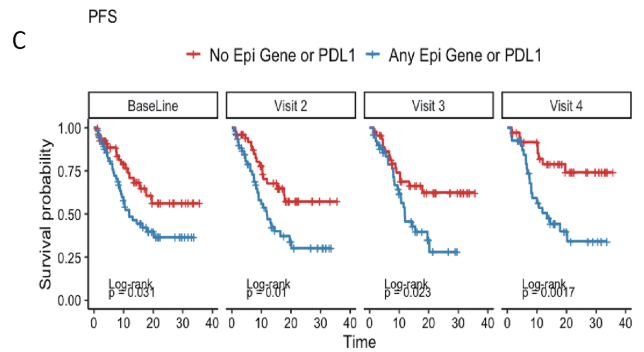
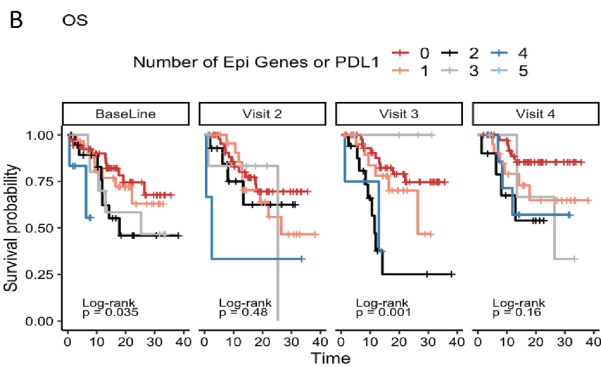
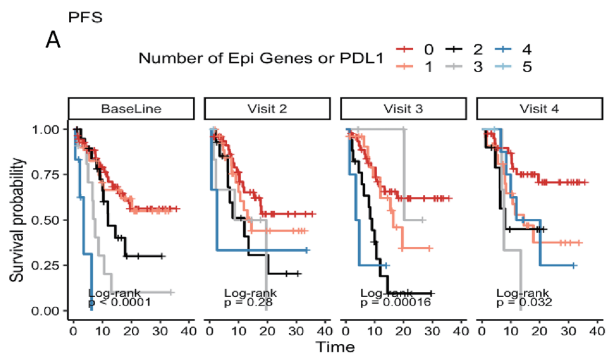
**Figure S2** Frequency of CTC count classification by response to therapy. CTC count distribution at each timepoint as the percentage of samples with 0, 1, 2-4, 5-10 or more than 10 CTC. Patients with progression at any point during the study after baseline are classified as “Progressed.” Data after visit 6 may be discounted since survivor bias and the small sample size limit interpretation. A lower percentage of patients who experienced progression had 0 CTC (light blue). For reference, the dashed yellow line indicates the percentage of patients with 0 CTC in the baseline sample from the cohort who did not later experience disease progression. CTC, circulating tumor cell.



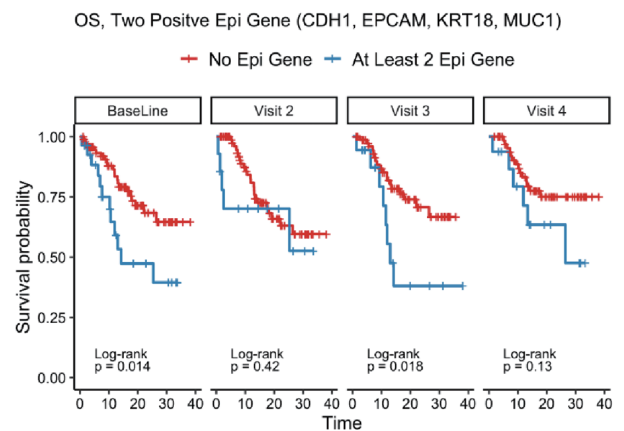
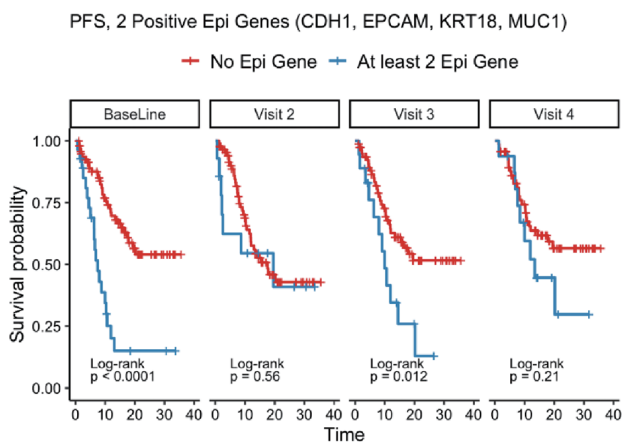
**Figure S3** Univariate gene expression by enriched CTCs in NSCLC patients across timepoints. Univariate Cox proportional hazard ratio models for each gene as a continuous variable showing the hazard ratio for each doubling (1-Ct change) in gene expression for (A) PFS and (B) OS. Gene expression was stratified using HD blood to define positive expression for (C) PFS and (D) OS. The survival based on the number of positive genes for (E) PFS and (F) OS. The median survival of patients with negative expression is shown with the green line, median survival for patients with positive CTC are shown with the red line and the hazard ratio for positive expression is blue. Points where median survival is not reached are not shown. PFS decreases as the number of positive genes in the CTC-enriched cells increases.



**Figure S4** Kaplan-Meier analysis of gene expression at each timepoint. Gene expression was stratified using HD blood to define positive expression for (A) PFS and (B) OS. Only genes significant by Kaplan-Meier analysis are shown.



**Figure S5** Univariate survival analysis of epithelial genes + PD-L1. Considering only the 4 primary epithelial genes (CDH1, EPCAM, KRT18, MUC1) and CD274 (PD-L1), survival decreases as the number of positive genes increases. Kaplan-Meier analysis for the number of positive genes at each timepoint. (A) PFS and (B) OS. Kaplan Meier analysis for any positive epithelial gene or PD-L1 (C) PFS and (D) OS. (Time in months).



**Figure S6** Univariate survival analysis of CTC epithelial gene expression: Survival stratified by detection of at least 2 epithelial genes (CDH1, EPCAM, KRT18, MUC1).

**Table S2** Time-dependent repeated measurements of gene expression in Cox model for OS after adjusting other clinical factors

	Parameter		Hazard ratio	95% hazard ratio confidence limits		P value
Model 1	BCL2		1.356	1.117	1.646	0.002**
	ECOG	1 vs. 0	1.078	0.475	2.446	0.858
	ECOG	2 vs. 0	4.282	1.487	12.327	0.007**
	Smoking	Yes vs. no	2.011	0.763	5.301	0.158
	Stage	3 vs. (1,2)	1.146	0.254	5.166	0.86
	Stage	4 vs. (1,2)	2.707	0.606	12.092	0.192
Model 2	CD274		1.159	1.012	1.327	0.033*
	ECOG	1 vs. 0	1.021	0.457	2.279	0.96
	ECOG	2 vs. 0	3.684	1.307	10.384	0.014*
	Smoking	Yes vs. no	1.538	0.598	3.951	0.372
	Stage	3 vs. (1,2)	1.805	0.411	7.918	0.434
	Stage	4 vs. (1,2)	3.583	0.809	15.867	0.093
Model 3	CDH1		1.096	0.93	1.293	0.274
	ECOG	1 vs. 0	0.961	0.43	2.147	0.923
	ECOG	2 vs. 0	3.767	1.332	10.65	0.012*
	Smoking	Yes vs. no	1.586	0.612	4.109	0.343
	Stage	3 vs. (1,2)	1.562	0.356	6.859	0.555
	Stage	4 vs. (1,2)	3.243	0.73	14.401	0.122
Model 4	EPCAM		1.409	1.071	1.853	0.014*
	ECOG	1 vs. 0	1	0.448	2.232	1.000
	ECOG	2 vs. 0	4.086	1.449	11.524	0.008**
	Smoking	Yes vs. no	1.657	0.634	4.33	0.303
	Stage	3 vs. (1,2)	1.849	0.423	8.079	0.414
	Stage	4 vs. (1,2)	3.459	0.781	15.32	0.102
Model 5	FN1		1.031	0.798	1.332	0.817
	ECOG	1 vs. 0	0.965	0.432	2.157	0.93
	ECOG	2 vs. 0	4.034	1.433	11.354	0.008**
	Smoking	Yes vs. no	1.682	0.654	4.329	0.281
	Stage	3 vs. (1,2)	1.697	0.383	7.523	0.486
	Stage	4 vs. (1,2)	3.465	0.766	15.671	0.107
Model 6	MUC1		2.18	1.48	3.212	<0.0001***
	ECOG	1 vs. 0	0.782	0.343	1.782	0.559
	ECOG	2 vs. 0	4.262	1.506	12.061	0.006**
	Smoking	Yes vs. no	1.656	0.637	4.31	0.301
	Stage	3 vs. (1,2)	1.377	0.311	6.094	0.674
	Stage	4 vs. (1,2)	3.022	0.677	13.48	0.147

\*, P&lt;0.05; \*\*, P&lt;0.01; \*\*\*, P&lt;0.001.

**Table S3** Time-dependent repeated measurements of gene expression in Cox model for PFS after adjusting other clinical factors

	Parameter		Hazard ratio	95% hazard ratio confidence limits		P value
Model 1	BCL2		1.219	1.039	1.431	0.015*
	ECOG	1 vs. 0	0.699	0.381	1.282	0.247
	ECOG	2 vs. 0	2.113	0.837	5.336	0.113
	Smoking	Yes vs. no	1.275	0.636	2.558	0.494
	Stage	3 vs. (1,2)	1.07	0.363	3.149	0.903
	Stage	4 vs. (1,2)	2.56	0.853	7.678	0.094
	Model 2	CD274		1.191	1.062	1.335
ECOG		1 vs. 0	0.711	0.388	1.302	0.269
ECOG		2 vs. 0	2.007	0.799	5.041	0.138
Smoking		Yes vs. no	1.121	0.563	2.229	0.745
Stage		3 vs. (1,2)	1.381	0.475	4.016	0.553
Stage		4 vs. (1,2)	3.47	1.162	10.368	0.026*
Model 3		CDH1		1.023	0.899	1.165
	ECOG	1 vs. 0	0.696	0.38	1.277	0.242
	ECOG	2 vs. 0	2.061	0.82	5.18	0.124
	Smoking	Yes vs. no	1.243	0.617	2.502	0.543
	Stage	3 vs. (1,2)	1.268	0.434	3.701	0.664
	Stage	4 vs. (1,2)	2.973	0.997	8.867	0.051
	Model 4	EPCAM		1.466	1.162	1.848
ECOG		1 vs. 0	0.756	0.41	1.393	0.369
ECOG		2 vs. 0	2.231	0.887	5.611	0.088
Smoking		Yes vs. no	1.151	0.571	2.321	0.694
Stage		3 vs. (1,2)	1.407	0.485	4.08	0.529
Stage		4 vs. (1,2)	3.064	1.036	9.06	0.043*
Model 5		FN1		0.834	0.614	1.134
	ECOG	1 vs. 0	0.656	0.356	1.209	0.177
	ECOG	2 vs. 0	2.046	0.821	5.1	0.125
	Smoking	Yes vs. no	1.277	0.64	2.545	0.488
	Stage	3 vs. (1,2)	1.254	0.429	3.668	0.679
	Stage	4 vs. (1,2)	2.97	0.994	8.876	0.051
	Model 6	MUC1		1.822	1.28	2.594
ECOG		1 vs. 0	0.615	0.331	1.139	0.122
ECOG		2 vs. 0	2.128	0.847	5.349	0.108
Smoking		Yes vs. no	1.18	0.586	2.378	0.643
Stage		3 vs. (1,2)	1.209	0.416	3.515	0.727
Stage		4 vs. (1,2)	2.747	0.922	8.181	0.07

\*, P&lt;0.05; \*\*, P&lt;0.01; \*\*\*, P&lt;0.001.