Appendix 1

1. Investigators

Sara Witting Christensen Wen[1,2], MD, Rikke Fredslund Andersen[3], MSc., PhD, Torben Frøstrup Hansen[1,2], associate Professor, MD, PhD, Christa Haugaard Nyhus[1], MD, Henrik Hager[2,4] associate professor, MD, PhD, Ole Hilberg[2,5], professor, MD, DMSc, Anders Jakobsen[1,2], professor, MD, DMSc

[1] Department of Oncology, Vejle Hospital, University Hospital of Southern Denmark, Beriderbakken 4, 7100 Vejle, Denmark
[2] Institute of Regional Health Research, University of Southern Denmark, 5230 Odense, Denmark
[3] Department of Clinical Biochemistry, Vejle Hospital, University Hospital of Southern Denmark, Beriderbakken 4, 7100 Vejle, Denmark
[4] Department of Pathology, Vejle Hospital, University Hospital of Southern Denmark, Beriderbakken 4, 7100 Vejle, Denmark
[5] Department of Medicine, Vejle Hospital, University Hospital of Southern Denmark, Beriderbakken 4, 7100 Vejle, Denmark

2. Methods

Methylated HOXA9 was analyzed with an in-house ddPCR assay. (Primers and probe from LGC Biosearch technologies, Aarhus, Denmark).

Primer sequence

Forward	GAGTATTTCGATTTTAGTTCGTGT
Reverse	CGCGTACACTAAATTCCAC

Probe sequence

Probe	FAM-TTAGTTTAAGGCGACGGTGTT-BHQ-1

Positive control:
Universal Methylated Human DNA Standard, 2 µl (DNA concentration 250 ng/µl, Zymo Research, Irvine, California, USA), healthy donor lymphocyte DNA 199 µl (DNA concentration approximately 20 ng/µl) and water 800 µl.
For each reaction, 20 µl of this mixture was added and bisulfite converted in parallel with the patient samples.

PCR conditions for the methylation specific droplet digital PCR assay

Steps	Temperature	Time
Step 1	95°C	10 minutes
Step 2: 44 cycles	95°C 56°C	15 seconds 1 minute
Step 3	98°C	10 minutes

Machine: Veriti Thermal Cycler (Applied Biosystems, Thermo Fisher Scientific, Foster City, California, USA).
Ramp rate: 1.5°C/second.

Cell free DNA yield:
We have measured total cell-free DNA by qPCR in the form of the β2 microglobulin gene in the present patient cohort. We found a median of 2825 copies/ml (mean 4920 copies/ml, range 279-106656 copies/ml) across the 228 baseline samples.

3. Univariate analyses

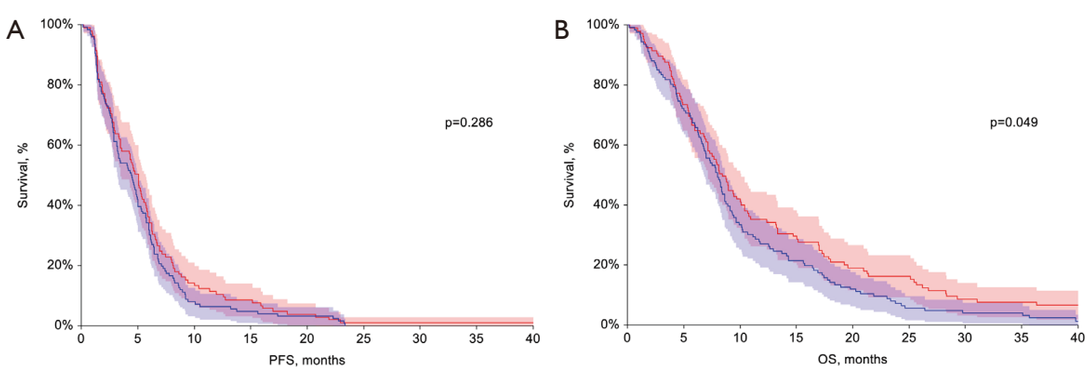

Figure S1 Prognostic impact of sex. Kaplan-Meier plots illustrating PFS and OS probability as a function of sex (A, B). Red line: Female. Blue line: Male. Colored areas represent 95% CIs. P-values for Log rank tests are shown.

http://dx.doi.org/10.21037/tlcr-20-826

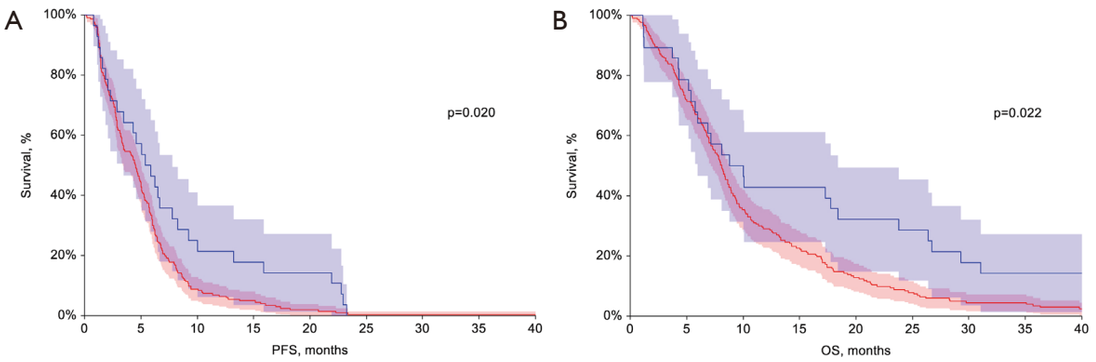

Figure S2 Prognostic impact of treatment. Kaplan-Meier plots illustrating PFS and OS probability as a function of treatment category (A, B). Red line: Platinum and vinorelbine. Blue line: Other treatments. The category 'Other' covers vinorelbine monotherapy and tyrosine kinase inhibitors. Colored areas represent 95% CIs. P-values for Log rank tests are shown.

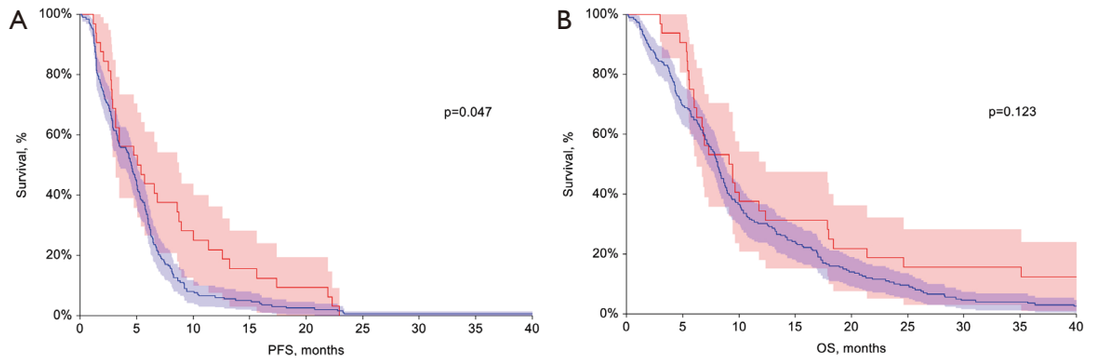

Figure S3 Prognostic impact of stage. Kaplan-Meier plots illustrating PFS and OS probability as a function of stage (A, B). Red line: Stage 3. Blue line: Stage 4. Colored areas represent 95% CIs. P-values for Log rank tests are shown.

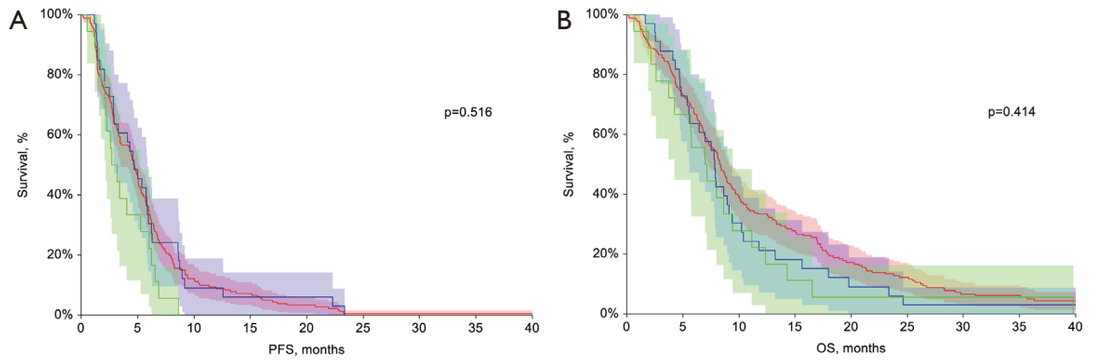

Figure S4 Prognostic impact of histologic type. Kaplan-Meier plots illustrating PFS and OS probability as a function of stage (A, B). Red line: Adenocarcinoma. Blue line: Squamous cell carcinoma. Green line: Other. The category 'Other' covers poorly differentiated non-small cell carcinoma and tumors with mixed histology treated as NSCLC. Colored areas represent 95% CIs. P-values for Log rank tests are shown.

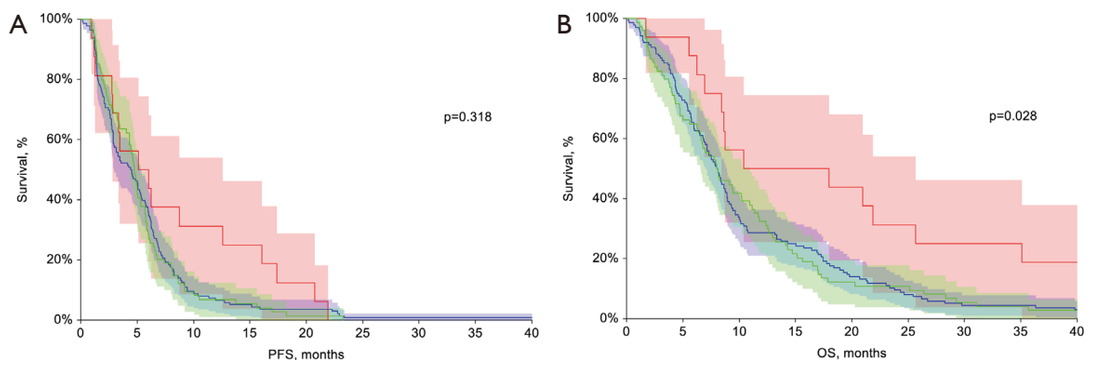

Figure S5 Prognostic impact of smoking status. Kaplan-Meier plots illustrating PFS and OS probability as a function of smoking status (A, B). Red line: Never smokers. Blue line: Previous smokers. Green line: Active smokers. Colored areas represent 95% CIs. P-values for Log rank tests are shown.

http://dx.doi.org/10.21037/tlcr-20-826

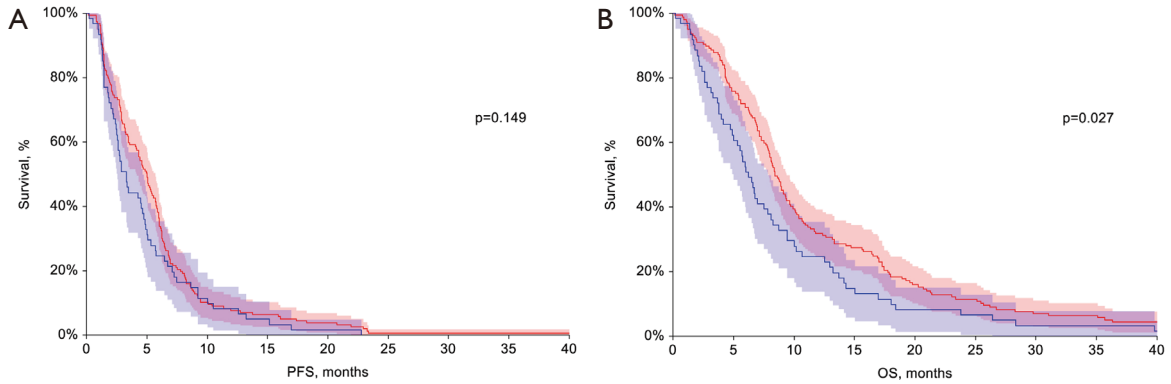

Figure S6 Prognostic impact of LDH status. Kaplan-Meier plots illustrating PFS and OS probability as a function of LDH status at baseline (A, B). Red line: LDH < 250 U/l. Blue line: LDH ≥ 250 U/l. Colored areas represent 95% CIs. P-values for Log rank tests are shown.

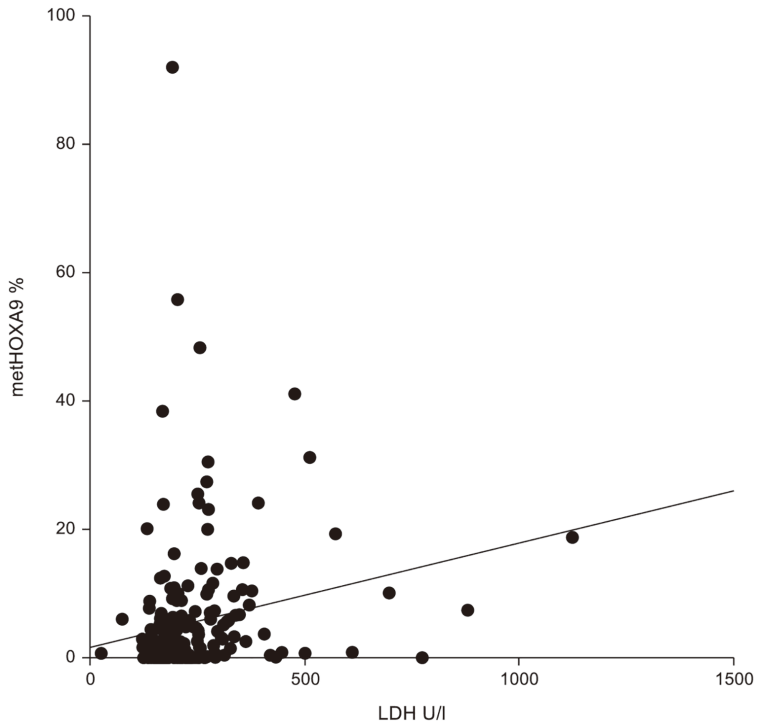

Figure S7 LDH in relation to metHOXA9. Scatter plot of LDH in U/l against metHOXA9 in percent showing a weak correlation depicted by a straight line, r=0.31 (A).

http://dx.doi.org/10.21037/tlcr-20-826

4. Multivariate analyses

Table S1 Multiple Cox regression analysis, all covariates

Covariate	HR	95% CI lower	95% CI upper	P-value
metHOXA9 status, 1 cycle of treatment				
Undetectable (reference)	1			
Detectable	3.619	2.085	6.280	<0.001*
Sex				
Female (reference)	1			
Male	1.237	0.912	1.677	0.172
Smoking status				
Never (reference)	1			
Previous	1.584	0.808	3.103	0.180
Active	1.722	0.859	3.450	0.125
Histology				
Adenocarcinoma (reference)	1			
Squamous cell carcinoma	1.166	0.730	1.862	0.520
Other	1.052	0.582	1.904	0.866
Treatment				
Platinum and vinorelbine (reference)	1			
Other	0.637	0.400	1.014	0.057
Stage				
Stage 3	1			
Stage 4	0.984	0.622	1.556	0.945
Age (numeric)	1.001	0.983	1.020	0.886
LDH baseline (numeric)	1.002	1.001	1.003	0.006*
Time-varying coefficient, metHOXA9	0.998	0.997	0.999	0.003*

Multiple Cox regression analysis performed on n=192 patients and 189 events with complete data for all covariates. Table presents hazard ratios (HR), 95% CIs, and p-values. *Statistically significant covariates.

The Cox regression model as reported in the manuscript (Table 2) was developed by the backward selection method. The full model (Table S1) was narrowed down by taking out one covariate with a high p-value and comparing the larger model with the smaller (nested) model by the likelihood-ratio test. If there was no significant difference between the two models, i.e. the missing covariate did not contribute significantly to the model, the smaller model was kept for further model development. There was no interaction between LDH and metHOXA9 neither as continuous nor binary variables.

The time varying coefficient (metHOXA9 status interacting with time) was included from the beginning. We tested the full model as described above (Table S1) but without the time varying coefficient with the proportional hazards test, and metHOXA9 status violated the proportional hazards assumption. The solution was to include a time varying coefficient in the model.

Missing data were treated as such and not included in the model or in other statistical analyses.

http://dx.doi.org/10.21037/tlcr-20-826