

## Genome-wide DNA methylation profiling in early stage I lung adenocarcinoma reveals predictive aberrant methylation in the promoter region of the long noncoding RNA PLUT: an exploratory study

Soo-Zin Kim-Wanner, Yassen Assenov, Mridul B. Nair, Dieter Weichenhan, Axel Benner, Natalia Becker, Katharina Landwehr, Ruprecht Kuner, Holger Sültmann, Manel Esteller, Ina Koch, Michael Lindner, Michael Meister, Michael Thomas, Matthias Bieg, Ursula Klingmüller, Matthias Schlesner, Arne Warth, Benedikt Brors, Erhard Seifried, Halvard Bönig, Christoph Plass, Angela Risch, Thomas Muley

### Angaben zur Veröffentlichung / Publication details:

Kim-Wanner, Soo-Zin, Yassen Assenov, Mridul B. Nair, Dieter Weichenhan, Axel Benner, Natalia Becker, Katharina Landwehr, et al. 2020. "Genome-wide DNA methylation profiling in early stage I lung adenocarcinoma reveals predictive aberrant methylation in the promoter region of the long noncoding RNA PLUT: an exploratory study." *Journal of Thoracic Oncology* 15 (8): 1338–50. <https://doi.org/10.1016/j.jtho.2020.03.023>.



# Genome-Wide DNA Methylation Profiling in Early Stage I Lung Adenocarcinoma Reveals Predictive Aberrant Methylation in the Promoter Region of the Long Noncoding RNA *PLUT*: An Exploratory Study

Soo-Zin Kim-Wanner, MD,<sup>a,b</sup> Yassen Assenov, PhD,<sup>a</sup> Mridul B. Nair, PhD,<sup>a</sup> Dieter Weichenhan, PhD,<sup>a</sup> Axel Benner, PhD,<sup>c</sup> Natalia Becker, PhD,<sup>c</sup> Katharina Landwehr, MSc,<sup>b</sup> Ruprecht Kuner, PhD,<sup>d,e</sup> Holger Sültmann, PhD,<sup>d,f</sup> Manel Esteller, MD, PhD,<sup>g,h,i,j</sup> Ina Koch, PhD,<sup>k</sup> Michael Lindner, MD,<sup>k</sup> Michael Meister, PhD,<sup>f,l</sup> Michael Thomas, MD,<sup>f,l</sup> Matthias Bieg, PhD,<sup>m</sup> Ursula Klingmüller, PhD,<sup>f,n</sup> Matthias Schlesner, PhD,<sup>f,m,o</sup> Arne Warth, MD, PhD,<sup>p</sup> Benedikt Brors, PhD,<sup>q</sup> Erhard Seifried, MD, PhD,<sup>b</sup> Halvard Bönig, MD, PhD,<sup>b</sup> Christoph Plass, PhD,<sup>a,f</sup> Angela Risch, PhD,<sup>a,f,r,s,\*</sup> Thomas Muley, PhD<sup>f,l</sup>

<sup>a</sup>Division of Epigenomics, German Cancer Research Center (Deutsches Krebsforschungszentrum [DKFZ]), Heidelberg, Germany

<sup>b</sup>German Red Cross Blood Donor Service, Institute of Transfusion Medicine and Immune Hematology of the Goethe University Medical School, Frankfurt, Germany

<sup>c</sup>Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>d</sup>Division of Cancer Genome Research, German Cancer Research Center (DKFZ), German Cancer Consortium (Deutsches Konsortium für Translationale Krebsforschung) and National Center for Tumor Diseases, Heidelberg, Germany

<sup>e</sup>Translational Oncology at the University Medical Center of Johannes Gutenberg University, Mainz, Germany

<sup>f</sup>Translational Lung Research Centre Heidelberg, Member of the German Centre for Lung Research (Deutsches Zentrum für Lungenforschung), Heidelberg, Germany

<sup>g</sup>Josep Carreras Leukaemia Research Institute, Badalona, Barcelona, Catalonia, Spain

<sup>h</sup>Institut Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain

<sup>i</sup>Centro de Investigación Biomédica en Red Cancer, Madrid, Spain

<sup>j</sup>Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona, Barcelona, Catalonia, Spain

<sup>k</sup>Center of Thoracic Surgery, Asklepios Fachkliniken München-Gauting, Ludwig Maximilians University, Gauting; Comprehensive Pneumology Centre Munich, German Centre for Lung Research (Deutsches Zentrum für Lungenforschung), Munich, Germany

<sup>l</sup>Translational Research Unit, Thoraxklinik at University Hospital Heidelberg, Heidelberg, Germany

<sup>m</sup>Division of Theoretical Bioinformatics, and Heidelberg Center for Personalized Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>n</sup>Division Systems Biology of Signal Transduction, German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>o</sup>Bioinformatics and Omics Data Analytics, German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>p</sup>Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany

<sup>q</sup>Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany

## \*Corresponding author.

**Disclosure:** Dr. Bönig reports receiving grants from Bayer, Chugai, EryDel, Miltenyi Biotec, Polyphor, Sandoz-Hexal (a Novartis Company), Stage (a Celgene Company), Terumo BCT, and uniQure; personal fees from Celgene, Chugai Pharmaceutical, Fresenius Medical Care, kiadis Pharma, Miltenyi Biotec, Novartis, Sandoz-Hexal, Terumo BCT, Genzyme, Novartis, Stage, and Terumo BCT; and Royalties and License Fees from Medac and Stock in Siemens Healthineers outside of the submitted work. Dr. Muley reports receiving nonfinancial support from the National Center for Tumor Disease, Heidelberg, during the conduct of the study; and grants and nonfinancial support from Roche Diagnostics GmbH outside of the submitted work. Dr. Meister reports receiving grants from Heidelberg Center for Personalized Oncology (DKFZ-HIPO) (project H004) during the conduct of the study. Dr. Plass reports receiving grants from the German Cancer Research Center (DKFZ), during the conduct of the study. Dr. Thomas reports receiving grants, personal fees, and

nonfinancial support from AstraZeneca, Bristol-Myers Squibb, Roche Holdings AG, and Takeda Pharmaceutical; and personal fees and nonfinancial support from Boehringer Ingelheim, Celgene, Chugai Pharmaceutical, Eli Lilly, Merck Sharp & Dohme, Novartis, and Pfizer outside the submitted work. The remaining authors declare no conflict of interest.

Address for correspondence: Angela Risch, PhD, Department of Biosciences, University of Salzburg, Billrothstr. 11, Salzburg 5020, Austria. E-mail: [angela.risch@sbg.ac.at](mailto:angela.risch@sbg.ac.at)

© 2020 International Association for the Study of Lung Cancer. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ISSN: 1556-0864

<https://doi.org/10.1016/j.jtho.2020.03.023>

<sup>r</sup>University of Salzburg, Department of Biosciences, Allergy-Cancer-BioNano Research Centre, Salzburg, Austria

<sup>s</sup>Cancer Cluster Salzburg, Universität Salzburg, Salzburg, Austria

Received 19 March 2019; revised 27 March 2020; accepted 29 March 2020

Available online - 7 April 2020

## ABSTRACT

**Introduction:** Surgical procedure is the treatment of choice in early stage I lung adenocarcinoma. However, a considerable number of patients experience recurrence within the first 2 years after complete resection. Suitable prognostic biomarkers that identify patients at high risk of recurrence (who may probably benefit from adjuvant treatment) are still not available. This study aimed at identifying methylation markers for early recurrence that may become important tools for the development of new treatment modalities.

**Methods:** Genome-wide DNA methylation profiling was performed on 30 stage I lung adenocarcinomas, comparing 14 patients with early metastatic recurrence with 16 patients with a long-term relapse-free survival period using methylated-CpG-immunoprecipitation followed by high-throughput next-generation sequencing. The differentially methylated regions between the two subgroups were validated for their prognostic value in two independent cohorts using the MassCLEAVE assay, a high-resolution quantitative methylation analysis.

**Results:** Unsupervised clustering of patients in the discovery cohort on the basis of differentially methylated regions identified patients with shorter relapse-free survival (hazard ratio: 2.23; 95% confidence interval: 0.66–7.53;  $p = 0.03$ ). In two validation cohorts, promoter hypermethylation of the long noncoding RNA *PLUT* was significantly associated with shorter relapse-free survival (hazard ratio: 0.54; 95% confidence interval: 0.31–0.93;  $p < 0.026$ ) and could be reported as an independent prognostic factor in the multivariate Cox regression analysis.

**Conclusions:** Promoter hypermethylation of the long noncoding RNA *PLUT* is predictive in patients with early stage I adenocarcinoma at high risk for early recurrence. Further studies are needed to validate its role in carcinogenesis and its use as a biomarker to facilitate patient selection and risk stratification.

© 2020 International Association for the Study of Lung Cancer. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

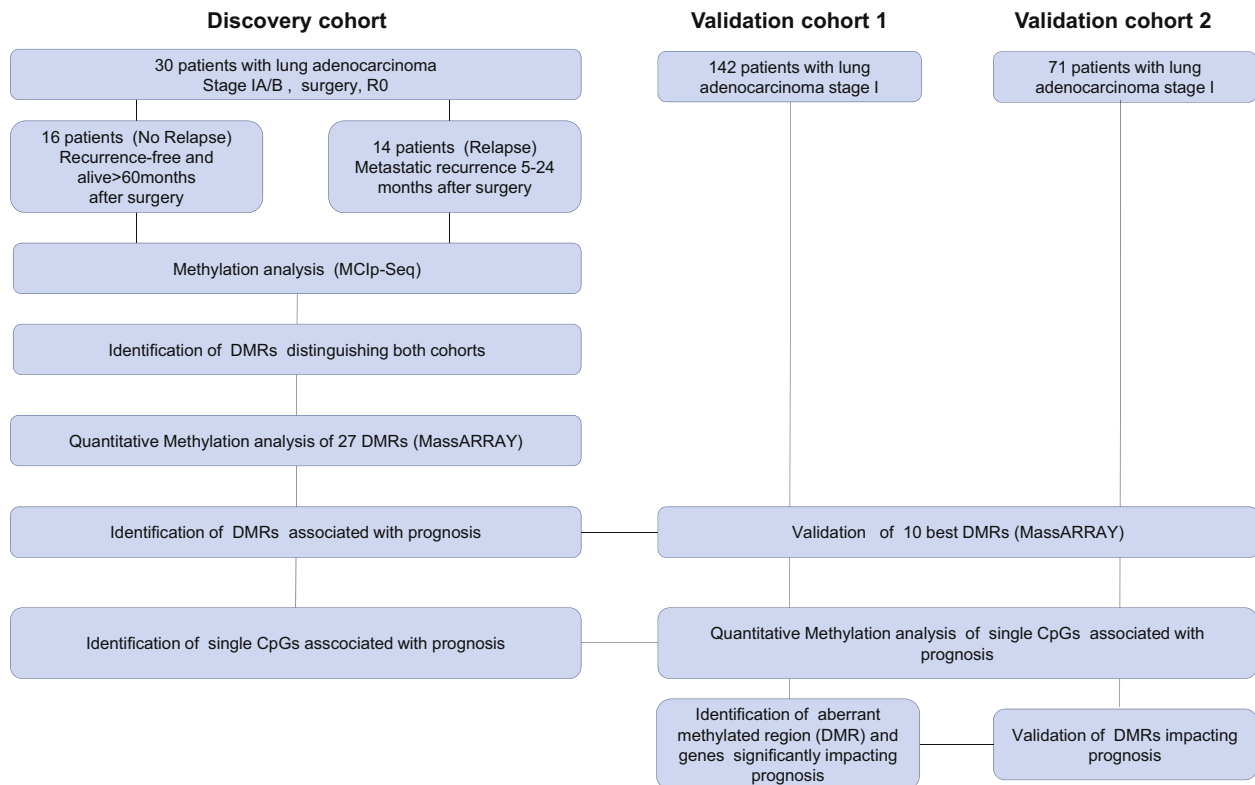
**Keywords:** PLUT; Methylation profiling; Lung adenocarcinoma; Prognostic marker; lncRNA

## Introduction

Lung cancer is the leading cause of cancer death worldwide, comprising a mortality of more than 1.5

million per annum and a poor prognosis with a 5-year overall survival (OS) of 18%.<sup>1,2</sup> Despite tremendous developments in targeted therapies for advanced stages, the treatment paradigms and prognosis have not changed in early stage lung adenocarcinoma for the past decades; recurrence rate remains high at 30% to 50% and the prognosis is still guarded.<sup>3–6</sup> Adjuvant chemotherapy is currently recommended for patients with resected stage II and III.<sup>7–12</sup> However, a benefit for adjuvant chemotherapy after radical resection in stage Ia lung adenocarcinoma could not be reported and remains controversial for stage Ib.<sup>13,14</sup> The heterogeneity of the disease, which evolves through a multitude of disrupted cell control mechanisms that ultimately converge into carcinogenesis, may explain why there are currently no known driver genes predisposing for early relapse.<sup>15,16</sup> Furthermore, outside of nodal involvement and tumor size, a separate (but no less urgent) problem is the lack of established biomarkers<sup>13,17</sup> to identify subpopulations of patients that are at high risk of recurrence, and therefore, might benefit from adjuvant treatment modalities.

In the search for new biomarkers, aberrant DNA methylation is gaining importance. Previous studies describe aberrant DNA hypermethylation in promoter regions as frequent molecular events; they were identified using the candidate gene approach or using the Infinium HumanMethylation BeadChip<sup>16,18–24</sup> with predetermined CpGs. Studies mapping unbiased genome-wide DNA hypomethylation and hypermethylation in lung adenocarcinoma, which would be reflective of the complex machinery of epigenetic regulation, are largely lacking.<sup>16,23–25</sup> Especially, differential epigenetic signatures and aberrantly methylated regions in intergenic regions remain to be identified; we predict that this could discriminate between subtypes or identify deregulated genomic regions, and, in turn, predict response to therapy and prognosis. Therefore, this study intended to identify differential methylation patterns by analyzing all CpGs of the whole genome using methylated-CpG-immunoprecipitation (MCIp) after next-generation sequencing. By this approach, a distinct methylation signature associated with early relapse in stage I lung adenocarcinoma can be detected after potentially curative treatment. We obtained a strong predictive DNA methylation marker that was validated in two independent cohorts that may provide information for further clinical studies on adjuvant treatment modalities and therapeutic targets.



**Figure 1.** Overview of the experimental design. In the discovery cohort, genome-wide methylation profiling was performed on 30 stage I lung adenocarcinomas comparing 14 patients with early metastatic recurrence to 16 patients with long-term relapse-free survival using MCIp and high-throughput next-generation sequencing. DMRs were validated using the MassCLEAVE Assay (MassARRAY). The 10 DMRs with the highest methylation differences between the two subgroups were validated in two independent validation cohorts. DMR, differentially methylated region; MCIp, methylated-CpG-immunoprecipitation.

## Materials and Methods

An overview detailing the experimental design is provided in [Figure 1](#).

### Patient Samples

For the discovery cohort and validation cohort 1, lung cancer and adjacent normal lung tissues were provided by Lung Biobank Heidelberg, a member of the tissue bank of the National Center for Tumor Diseases (Heidelberg, Germany), and the biobank platform of the German Centre for Lung Research, and the validation cohort 2 tissues were provided by the Asklepios Biobank for Lung Diseases, a member of the platform biobank of the Comprehensive Pneumology Center, German Center for Lung Research (Deutsches Zentrum für Lungenforschung [DZL]) (Munich, Germany), after approval by the respective local institutional review boards. Written and oral informed consent was obtained for all samples from patients with early stage I to IIb lung adenocarcinoma. Patients were diagnosed and treated at Thoraxklinik Heidelberg and Asklepios Clinic Munich-Gauting where clinical follow-up was pursued and documented. Tissue samples were snap-frozen in liquid nitrogen within 30

minutes after resection and stored at  $-80^{\circ}\text{C}$ . In all samples, the viable tumor content was at least 50% on the basis of the pathologic examination of hematoxylin and eosin-stained slides. To limit variance among the samples owing to heterogeneity between individuals, adjacent normal lung tissue samples ( $>2$  cm distant to tumor site) were obtained from each patient of the discovery cohort for pairwise comparison.

**Discovery Cohort.** Defined selection criteria were applied to stratify two different risk cohorts for further comparative methylation analysis. A total of 16 patients with stage Ia or Ib lung adenocarcinoma were identified for the “no relapse” cohort, fulfilling the criteria of relapse-free survival (RFS) of more than 60 months. Patients who had received adjuvant treatment were excluded from this cohort. Patients were eligible to enter the study as part of the “relapse” cohort if recurrence had occurred between 5 and 24 months after an operation. The rationale for choosing this arbitrary timeline was to exclude (R0) surgical failures ( $>5$  mo) and patients with risks of secondary malignancies ( $<24$  mo). A total of 14 patients with distant metastasis with or

Table 1. Patient Characteristics

Characteristics	Discovery Cohort				Validation Cohort 1 (HB)		Validation Cohort 2 (DZL)	
	No Relapse <sup>a</sup>		Relapse <sup>b</sup>					
	n	%	n	%	n	%	n	%
<b>Total</b>	16	53	14	47	142		71	
<b>Gender</b>								
M	8	27	12	40	81	57	31	43.7
F	8	27	2	6.7	61	43	40	56.3
<b>Age</b>								
Median [yrs]	64		62		66		65	
Range [yrs]	38-74		43-75		38-85		43-85	
<b>Stage TNM</b>	6th (7th)		6th (7th)		6th (7th)		7th	
Ia	5 (5)	17	2 (2)	6.7	30 (30)	21	34	47.9
Ib	11 (7)	37	12 (5)	40	112 (88)	79	37	52.1
IIa	0 (3)		0 (6)		0 (16)		0	
IIb	0 (1)		0 (1)		0 (8)		0	
<b>Tumor diameter</b>								
Median (range) [cm]	3.8 (1.4-11)		4.05 (1.9-16)		3.3 (1-17)		2.75 (0.6-5)	
<b>Histological subtype</b>								
Acinary		9			41		30	
Lepidic		3			18		3	
Micro papillary		0			2		0	
Papillary		3			12		18	
Solid		7			23		16	
Fetal		0			1		0	
Mixed Types		7			40		4	
Adenocarcinoma w/o further information		2			5		0	
<b>Smoking history</b>		20						
Smoker	6	20	4	13	49	35	43	60.6
Former smoker	7	23	8	6.7	22	15	18	25.4
Non-smoker	2	6.7	2	6.7	17	12	10	14.1
NA	1	3.3	0	0	1	0.7	0	0
<b>Adjuvant chemotherapy</b>	0	0	4	13	23	16	3	4.23
<b>Relapse free duration [months]</b>	0		9.1 (5.2-16)		41.3 (0.4-107)		34.9 (0-88.9)	
<b>Death</b>	0	0	13	93	38	27	9	12.7
<b>Median observation time (range) [months]</b>	72 (60-107)		35 (7.3-37.8)		49 (0.4-108.5)		36.2 (0-88.9)	

<sup>a</sup>Patients without relapse after R0-resection of the tumor and alive for more than 60 months.

<sup>b</sup>Patients with relapse after 5 to 24 months after R0-resection of the tumor.

DZL, Deutsches Zentrum für Lungenforschung; F, female; HB, Heidelberg Biobank; M, male.

without local recurrence were enrolled for the relapse cohort. Diagnosis and tumor resection were carried out from 2002 to 2010. The clinical characteristics of the patients are summarized in [Table 1](#).

**Validation Cohort 1 and Validation Cohort 2.** For validation cohort 1 (Heidelberg Biobank), 142 samples of patients with lung adenocarcinoma stage Ia or Ib (TNM sixth edition) were obtained from the Lung biobank, Heidelberg. Diagnosis and tumor resections were carried out from 2002 to 2011. Reassessment of stages according to TNM seventh edition revealed 128 stage Ia or Ib and 24 stage IIa or IIb. All patients were nodal negative. Validation cohort 2 (DZL) included 71 samples

of patients with lung adenocarcinoma stage Ia or Ib (TNM seventh edition) from the Asklepios Biobank Munich-Gauting from 2009 to 2015. Ethical approvals of the respective local institutional review boards were in place. Patient characteristics are summarized in [Table 1](#).

### DNA and RNA Isolation

DNA and RNA of the tumor and adjacent lung tissues were extracted using the QIAamp DNA mini kit and RNeasy mini kit (Qiagen, Hilden, Germany), respectively. Quality control activities were performed using gel electrophoresis (DNA) and Agilent Bioanalyzer 2100 (RNA) (Agilent, CA). For reverse transcriptase polymerase chain reaction (PCR) analysis, RNA was reverse



transcribed into complementary DNA using the SuperScript II (Invitrogen, Carlsbad, CA). Reverse transcriptase PCR was performed using complementary DNA equivalent to 10 ng RNA by adding a mixture of 5.5  $\mu$ L Probes 480 Master Enzyme Mastermix (Roche, Mannheim, Germany), 0.11  $\mu$ L of 20  $\mu$ M primer pairs, 0.11  $\mu$ L Universal Probe Library probe (UPL, Roche: *BTG2*: UPL1; *FYN*: UPL29; *PDX1*: UPL78; *RPTOR*: UPL23), and 0.28  $\mu$ L H<sub>2</sub>O on a LightCycler 480 PCR system (Roche, Mannheim, Germany). Crossing point values were adjusted to the housekeeping gene *GAPDH* and relative expression was calculated as given by Livak and Schmittgen.<sup>26</sup> Primer information is given in [Supplementary Figure 1A](#).

### MCIP and High-Throughput Next-Generation Sequencing

Global methylation analysis was performed using MCIP followed by high-throughput next-generation sequencing, as described earlier.<sup>27,28</sup> First, 60  $\mu$ g of methyl-CpG-binding domain Fc-protein was coupled to 40  $\mu$ L of protein A-coated magnetic beads (Diagenode, Liege, Belgium) and bound to 3  $\mu$ g of sonicated genomic DNA (Bioruptor NextGen, Diagenode) with a median fragment length of 145 base pairs (bp) as confirmed by Agilent's BioAnalyzer high-sensitivity DNA platform. Methylated DNA was then eluted by the variable stringency of the magnetically labeled protein using the SX-8G IP-Star robot (Diagenode) by increasing the salt concentration (300–1000 mM). Highly methylated DNA (1000 mM eluate) was desalted with MinElute columns (Qiagen) and enrichment was quantified by real-time PCR using the imprinted gene *SNRPN*. MCIP-enriched methylated DNA fragments were submitted to the DKFZ Genomics and Proteomics Core Facility for library preparation and next-generation sequencing. Afterward, fragmented DNA was end-repaired and ligated to Illumina-paired end adaptors using NEBnext DNA Library Prep Master Mix Set (New England Biolabs) in accordance with the manufacturer's instructions. Adapter ligated libraries were directly amplified by 14 cycles of PCR with the standard Illumina index primers and distributions were validated using the Agilent Bioanalyzer before it was quantified by a Qubit fluorometer (Invitrogen). The libraries were sequenced on the Illumina HiSeq 2000 sequencer (50 bp, single read 50 bp) using standard Illumina protocols.

### Data Analysis and Candidate Gene Selection

Sequencing reads were aligned to the hg19 genome assembly of the human reference genome using the Burrows-Wheeler Alignment tool.<sup>29</sup> Aligned reads were further processed by merging lane-level data and removing duplicates. The remaining uniquely mapped

reads were converted to Sequence Alignment Map or Binary Alignment Map formats using SAMtools. Read counts of each sample were normalized for total read length and the number of sequencing reads (reads per kilobase per million mapped reads; RPKMs). Peak calling was performed using the software HOMER (v4.4). A differentially methylated region (DMR) of one peak was defined if at least a twofold difference in tag densities between the tumor and adjacent normal tissue of one patient sample set was detected. DMRs enriched in the relapse subcohort compared with the no relapse subcohort were calculated using Fisher's exact test and *p* values were corrected by multiple testing (Benjamini-Hochberg). The second list of DMRs was obtained by comparison of the mean RPKMs of each cohort (no relapse versus relapse) using the Wilcoxon ranked sum test (*p* < 0.05) after multiple testing correction (Benjamini-Hochberg). For candidate marker selection and further quantitative methylation analysis, we focused on the nonrepetitive overlapping DMRs fulfilling both criteria.

### Quantitative Methylation Analysis

Quantitative methylation values for candidate gene regions were technically validated using matrix-assisted laser desorption and ionization time-of-flight mass spectrometry (MALDI-TOF MS), as described before.<sup>30</sup> The 10 best-validated amplicons were further assessed for clinical validation in the independent sample cohort (Heidelberg Biobank). Genomic DNA was bisulfite-converted using the EZ DNA methylation kit (Zymo Research, Orange, CA) in accordance with the manufacturer's protocol. Defined genomic regions covering the candidate gene regions of bisulfite-converted DNA were PCR-amplified, followed by in vitro transcription and base-specific cleavage. Methylation standards (0%, 20%, 40%, 60%, 80%, 100%) were used to control for the dynamic range of measurements. DNA methylation status was quantitatively assessed by MALDI-TOF MS using the Sequenom (Sequenom, San Diego, CA) (primer information is provided in [Supplementary Table 1A](#)). Besides the methylation value of single CpG units, the mean methylation value of overall CpG units per amplicon was used for further analysis. The samples of validation cohort 1 were divided into two groups (high and low methylation) for each DMR using the median methylation value of the whole cohort as the cutoff. The mean methylation values of the amplicons were used for survival analysis (Mantel-Cox test). The methylation status of the predictive amplicons was tested in the independent validation cohort (DZL) and correlated with clinical outcomes to validate the prognostic significance.

## Reporter Gene Assay

The reporter gene assay was performed as previously described.<sup>31</sup> The promoter region of the long noncoding RNA (lncRNA) PLUT was amplified from genomic DNA using phusion polymerase (Thermo Scientific) and inserted in the luciferase vector pGL4.10 (Promega). Luciferase activity was measured 48 hours after transfection into HEK-293 cells and normalized to the activity of a cotransfected cytomegalovirus promoter-driven  $\beta$ -galactosidase vector. Three independent experiments were performed in duplicates. Error bars show the SD. Transfections were performed with metafectene according to the manufacturer's instructions.

## Statistical Analysis

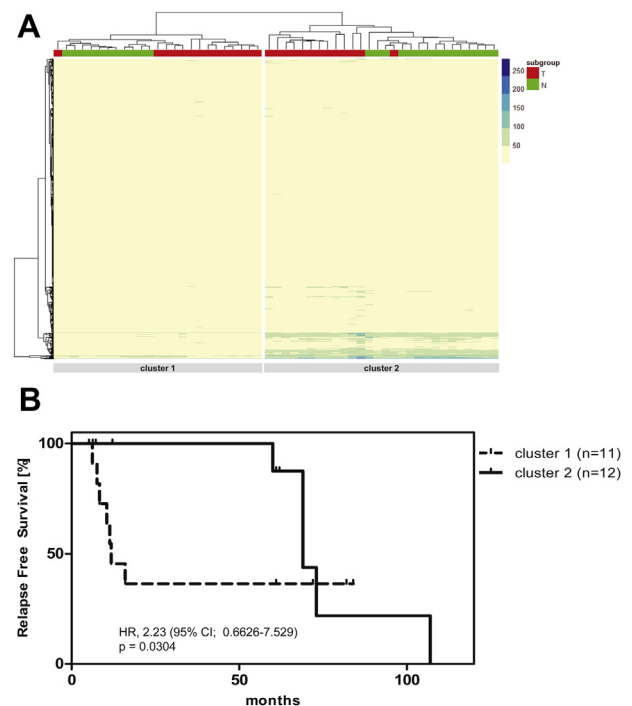
The primary end point of the statistical analysis was RFS, defined as the time from surgical procedure to tumor recurrence or death from any cause. The secondary end points were OS, defined as the time from surgical procedure to death from any cause, and disease-free survival (DFS). Events for DFS are the same as for RFS but, in addition, include any signs or symptoms of the disease, and time to event was also measured (whatever occurs first). Patients without an observed event were censored at the last available follow-up time point.

Distributions of survival times were estimated by using the method of Kaplan and Meier. The distribution of follow-up times was estimated by the reverse Kaplan-Meier estimate.

To investigate the influence of the DMRs on the survival end points (RFS, DFS, OS), multivariable Cox regression models were applied. A backward variable selection was used to recursively reduce the model size and keep only DMRs with *p* values less than 0.5 using the pooled residual chi-square as the stopping rule.

For illustration purposes, Kaplan-Meier curves for cohorts of patients with low- and high-methylation levels were computed for each DMR. Groups of low- and high-methylation levels were defined using the median methylation level (0.3) of validation cohort 1 as the cutoff.

Prediction error curves were calculated using the time-dependent Brier score, the quadratic difference between the 0 to 1 response status, and the predicted survival probability at time point *t*.<sup>32</sup> The 0.632+ bootstrap method was applied to reduce the bias in estimating the predictive accuracy. The integrated Brier score (IBS) over time was used as a statistic for comparison of models, including the model for the average methylation of the whole amplicon, the model using all single CpGs, and the model of CpGs after backward selection. The empty model, that is, the Kaplan-Meier estimate was used as a reference.



**Figure 2.** Methylation analysis and identification of DMRs. (A) Unsupervised hierarchical clustering of the RPKMs (0 to >250) of all DMRs derived by pairwise comparison between peak regions of tumor and adjacent normal tissue samples separates two clusters. Each cluster harbored tumor samples and their adjacent normal samples, although the tumor and normal samples separated in both clusters. (B) Kaplan-Meier estimates of both clusters revealed significant ( $p = 0.0304$ ) differences in relapse-free survival between both clusters. (C) Schematic overview of two peak regions derived from the MClp-seq analysis with differences in the “relapse” and “no relapse” cohorts. Amplicons (MA-amplicon) for quantitative methylation analysis using the MassCLEAVE assay (MassARRAY) and the results for each CpG unit are reported in a heatmap and a scattergram for the relapse cohort (pink) and the no relapse cohort (blue). CI, confidence interval; DMR, differentially methylated region; HR, hazard ratio; MClp-seq, methylated-CpG-immunoprecipitation sequencing; RPKMs, reads per kilobase per million mapped reads.

All statistical tests were two-sided. Values of *p* less than 0.05 were considered statistically significant. All calculations were done using the statistical software environment R, version 3.1.3 (R packages: pec\_ 2.4.7, prodlm\_1.5.1, rms\_4.3-1, survival\_2.38-1, and glmnet\_2.0-2).

## Results

### Distinct Genome-Wide Methylation Patterns Between the Early and Late Relapse Groups in the Discovery Cohort

Genomic DNA of 30 paired samples of bronchial adenocarcinoma and adjacent normal lung tissue were analyzed for genome-wide methylation using MClp. Analysis by real-time PCR revealed that methylation

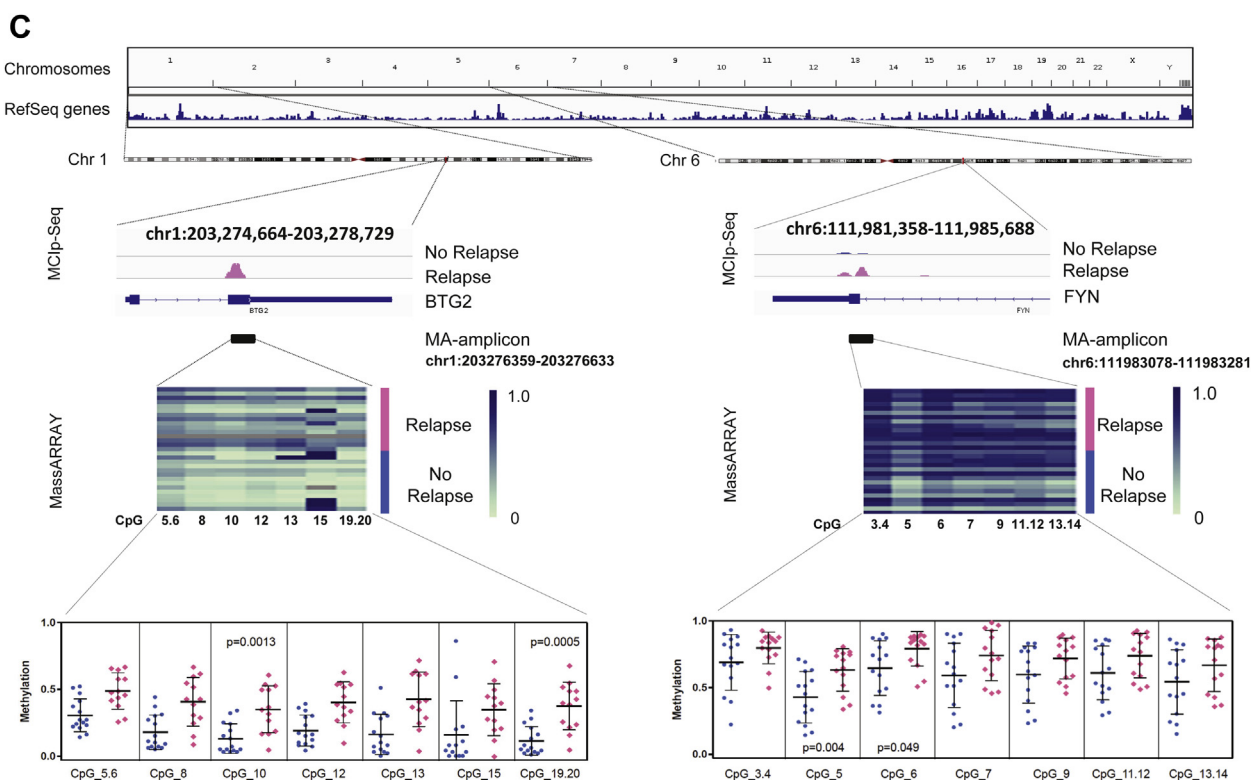


Figure 2. (continued).

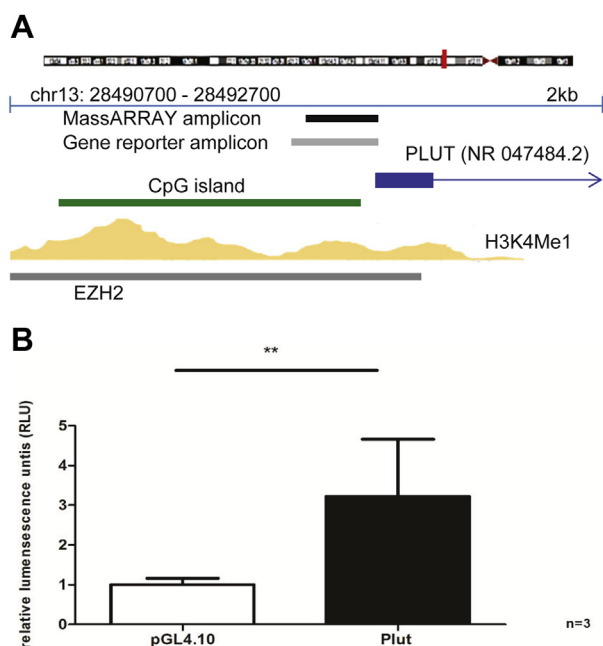
enrichment failed for six samples. The recovered DNA of the remaining 54 samples was sequenced using the Illumina HiSeq 2000 platform. Six libraries required resequencing to reach a saturation greater than 95%. One tumor sample was excluded from the analysis owing to suspected sample mix-up. The total number of sequence reads, mapped reads, and unique reads for each sample are represented in [Supplementary Table 2](#). For the final analysis, sequence reads of 24 paired sample sets and additional three tumor samples and two normal lung samples were available. In total 572,354 peaks could be identified in 200 bp genomic windows after normalization. On comparing the peaks in each paired sample for regions found to be differentially methylated in at least one tumor when compared with its lung counterpart, 238,331 DMRs between normal and tumor tissue were observed. Unsupervised hierarchical clustering of the RPKMs of all 238,331 DMRs separated the samples into two clusters. Clustering was performed using the package “pheatmap” in R Bioconductor (distance = average, k-means = 1000). The normal tissue samples clustered separately to their respective paired tumor tissue sample in each cluster ([Fig. 2A](#)), as expected. Kaplan-Meier estimates of the two clusters with distinct methylation patterns revealed significant differences in RFS (hazard ratio [HR]: 2.23; 95% confidence interval [CI]: 0.66–7.53;  $p = 0.03$ ) ([Fig. 2B](#)). A total

of 99 DMRs were significantly different between the relapse cohort and the no relapse cohort ( $p < 0.05$ ; Fisher’s exact test; Benjamini-Hochberg). For further candidate marker selection, we focused (after multiple testing corrections) on the nonrepetitive DMRs that were also significantly different in RPKMs between the two subcohorts ( $p < 0.05$ ; Wilcoxon rank sum test; Benjamini-Hochberg). A total of 27 DMRs fulfilled the selection criteria. The distribution of the DMRs among the genomic features is shown in [Supplementary Figure 2](#).

### Quantitative Methylation Analysis Reveals DMRs Between the Two Subcohorts

For technical validation of the results obtained with MCIp sequencing, the same genomic DNA of the discovery cohort was assessed for quantitative methylation in the 27 candidate marker regions by MALDI-TOF MS. For all selected regions, we confirmed differences between normal and tumor tissue and at least a trend between the relapse and no relapse cohorts. Significant differences between the subgroups were detected for the DMRs annotated to the closest genes *BTG2* and *OSTM1-AS* ( $t$  test,  $p < 0.05$ ). In some cases, an individual CpG unit within an analyzed amplicon revealed significantly stronger methylation differences between the two cohorts, compared with the average methylation of all





**Figure 3.** Analysis of the DMR upstream of the lncRNA *PLUT*. (A) Schematic overview of the DMR 22 bp upstream of the lncRNA. The DMR (MassARRAY amplicon) was quantitatively analyzed for methylation using the MassCLEAVE Assay and is annotated to a CpG island region enriched for H3K4Me1 mark (ChIP-seq ENCODE) for the transcription factor EZH2 (TF ChIP-seq ENCODE). A total of three CpGs (cg01193690, cg10034364, and cg15992563) within the amplicon are present on the Illumina Human Methylation 450kBeadChip. (B) The region (gene reporter amplicon) was cloned in the pGL4.10 luciferase vector and transfected in HEK-293 T cells. The transfection of the empty pGL4.10 vector served as a control. Luciferase promoter assay was performed in three independent experiments. bp, base pair; ChIP-seq, chromatin immunoprecipitation sequencing; DMR, differentially methylated region; ENCODE, encyclopedia of DNA elements; lncRNA, long noncoding RNA.

CpGs of the whole amplicon (Fig. 2C). None of the top 10 DMRs with the highest differences were located in promoter regions of coding regions. Two DMRs were located in promoter regions of lncRNAs: *PLUT* and *LOC648987* (uncharacterized RNA gene). The aberrant methylated region 22 bp upstream of the lncRNA *PLUT* (NR\_047484.2) is located in CpG islands within silico binding of H3K4me3 and revealed promoter activity in the reporter gene assay (Fig. 3). Peak regions of the 10 best DMRs derived from MCIP sequencing and amplicon information with annotations to genomic features are summarized in Supplementary Tables 1A and 1B.

### DMRs Correlate With Clinical Outcome in Patients With Stage I Lung Adenocarcinoma

To discover markers prognostic of risk of relapse in early stage lung adenocarcinoma, 10 candidate DMRs

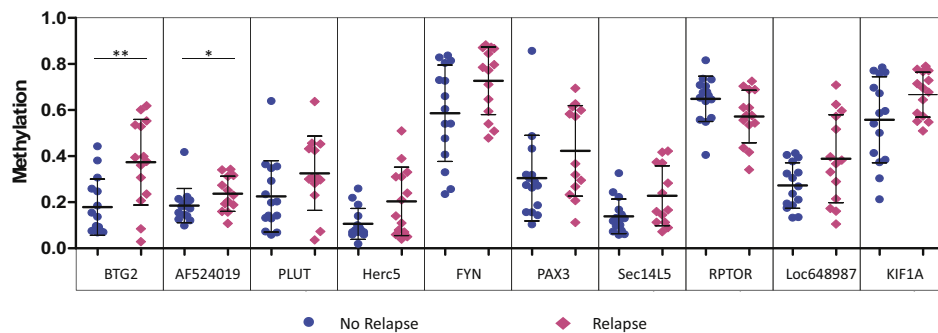
annotated to *HERC5*, *Loc648987*, *PAX3*, *SEC14L5*, *AF524019*, *BTG2*, *KIF1A*, *RPTOR*, *FYN*, and *PLUT* obtained from the quantitative methylation analysis (MALDI-TOF MS) (Fig. 4) were further investigated in 142 samples of validation cohort 1. The influence of the 10 candidate DMRs on clinical outcome was evaluated by multivariate Cox regression analysis. After backward selection, the average methylation of the amplicons *PLUT* ( $p = 0.01$ ) and *FYN* ( $p = 0.03$ ) revealed a significant prognostic impact on RFS. Figure 5A shows the corresponding Kaplan-Meier curves for *PLUT* using a median split (HR: 0.54; 95% CI: 0.31–0.93;  $p = 0.026$ ).

After backward selection methylation of the single CpGs of *PLUT*\_CpG17 ( $p < 0.0002$ ) and *FYN*\_CpG13.14 ( $p < 0.01$ ), *FYN*\_CpG6 ( $p = 0.01$ ) and *FYN*\_CpG9 ( $p = 0.01$ ) could be identified as significantly predictive of RFS (data not shown).

The predictive accuracy was estimated using the 0.632+ bootstrap method. Prediction error curves are shown in Supplementary Figure 3. The IBS over 59 months of follow-up revealed improvement of the model using selected CpGs (IBS = 0.180) compared with the Kaplan-Meier reference (IBS = 0.185) and using the model with the average methylation of the amplicon (IBS = 0.198). Regarding DFS and OS, average methylation of the DMR *KIF1A* (DFS:  $p = 0.02$ ; OS:  $p = 0.01$ ), and again, *PLUT* (DFS:  $p = 0.004$ ; OS:  $p = 0.005$ ) revealed a statistically significant prognostic impact (data not shown).

### Aberrant Methylation of the Promoter Region of lncRNA *PLUT* Predicts Clinical Outcome in an Independent Validation Cohort

The prognostic model including the DMRs *PLUT* and *FYN* was further tested in a second independent validation cohort with stage Ia or Ib (patient characteristics are shown in Table 1). Validation cohort 2 included 71 patients with a median follow-up of 36.2 months (95% CI: 31.1–47.1). The methylation levels of the same 10 DMRs that were tested in validation cohort 1, including *PLUT* and *FYN*, were analyzed by MALDI-TOF MS using the same amplicons and primers as for the discovery cohort and validation cohort 1. After backward selection, methylation of *PLUT* could be validated with respect to RFS (HR: 0.32; 95% CI: 0.12–0.86;  $p = 0.024$ ) (Fig. 5B). The DMR *FYN* was not significantly associated with RFS in validation cohort 2. The prediction error estimation revealed an IBS over 87.7 months of follow-up within the same range for the reference model (IBS = 0.151), the average methylation model (IBS = 0.166), and the model including the selected CpGs (IBS = 0.160).

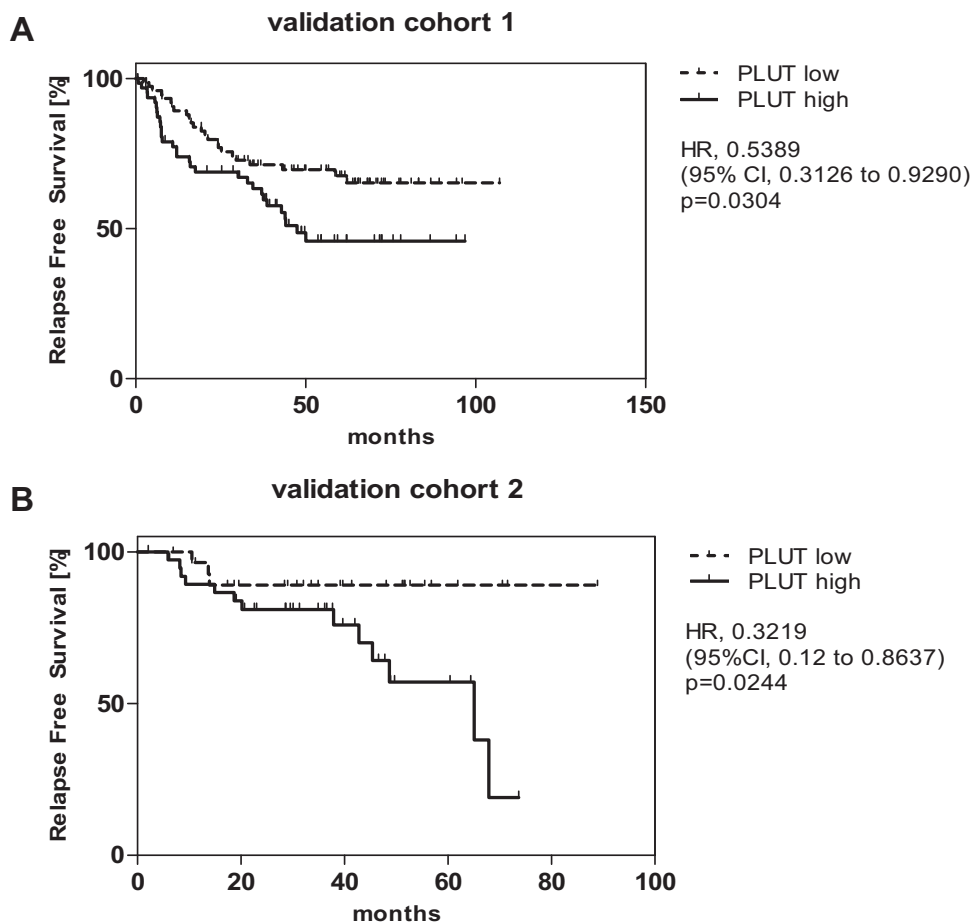


**Figure 4.** Quantitative methylation analysis of DMRs in the discovery cohort (n = 29) using MassARRAY. Quantitative methylation was analyzed in 29 samples of the discovery cohort using the MassCLEAVE Assay. The average methylation of the amplicons annotated to the genes *BTG2*, *HERC5*, *FYN*, *PAX3*, and *Sec14L5* significantly differed between the “relapse” and “no relapse” cohorts. DMR, differentially methylated region.

#### Aberrant Methylation of *BTG2*, *FYN*, *RPTOR*, and *PLUT* Correlates With Expression in the Discovery Cohort

To investigate whether the four detected DMRs with the highest ability to predict prognosis were involved in

the transcriptional regulation of their annotated genes *BTG2*, *FYN*, *RPTOR*, and *PDX1*, mRNA expression was determined in the two subgroups (no relapse versus relapse) of the discovery cohort and six additional samples. Expression levels for *BTG2*, *FYN*, and *RPTOR*



**Figure 5.** Kaplan-Meier survival estimates on the basis of promoter methylation of lncRNA *PLUT*. Relapse-free survival curves with respect to methylation of the promoter region of *PLUT* are reported for (A) validation cohort 1 and (B) validation cohort 2. Subgroups of low- and high-methylation levels were generated using the median methylation level in validation cohort 1 as a cutoff. CI, confidence interval; HR, hazard ratio; lncRNA, long noncoding RNA.

**Table 2.** Multivariate Regression Analysis Reveals Significant Correlation With RFS for *PLUT*, Sex, and Stage

	HR	95% CI	P value
<i>PLUT</i> methylation	5.53	1.33-22.92	0.018
Sex	2.19	1.22-3.95	0.008
Age (y)	1	0.97-1.03	0.99
P-stage (TNM seventh edition)	2.25	1.23-4.12	0.008

CI, confidence interval; HR, hazard ratio; lncRNA, long noncoding RNA; RFS, relapse-free survival.

significantly differed between the two groups. *PLUT* expression analysis could not be performed owing to lack of sufficient material and technical reasons. Nevertheless, expression of *PDX1* as a target gene of *PLUT* revealed a trend without reaching significance, but owing to assay performance, quantitative PCR could not be performed for all samples (22 of 36) (Supplementary Fig. 1B).

### Multivariate Regression Analysis Identifies Aberrant Promoter Methylation of *PLUT* as an Independent Prognostic Marker

Because clinical parameters are known to influence the outcome, we performed a multivariate analysis in the 142 patients of validation cohort 1 to test whether the DMR *PLUT* may serve as an independent prognostic marker for early recurrence. Using Cox regression, besides methylation status (binary) of *PLUT*, age, stage, and sex were included, and hypermethylation of *PLUT* could be detected as an independent unfavorable prognostic factor (HR: 5.53; 95% CI: 1.3–22.9) (Table 2). Furthermore, a significant difference in outcome regarding sex was apparent in both validation cohorts. As illustrated by the Kaplan-Meier curves, women revealed significantly better survival than men (Supplementary Fig. 4A). Men with promoter hypermethylation of *PLUT* had the worst outcome with an estimated DFS of 33% after 60 months compared with 76% for women with hypomethylation of the promoter region of *PLUT* ( $p = 0.001$ ) (Supplementary Fig. 4B), the other subcohorts revealing intermediate outcomes.

## Discussion

Clinical management in lung cancer will be greatly aided by the discovery of reliable prognostic biomarkers to guide treatment decisions. Stratification of patients by the need, or lack of it, for (neo)adjuvant treatment is highly warranted. We report that the methylation status of the promoter of lncRNA *PLUT* distinguishes between patients at risk and not at risk for early relapse.

This study represents the most comprehensive analysis of genome-wide DNA methylation in primary lung adenocarcinoma. Previous DNA methylation studies

were limited to analyses of promoter hypermethylation of CpG island methylation.<sup>18,19,33–35</sup> Other DNA methylation profiles were approached with a lower resolution in small cohorts and not selective for stage I cancers.<sup>22,34</sup> Suggested surrogate markers like *p16*, *APC*, *CDH13*, and *RASSF1A* promoter hypermethylation were found to be associated with early recurrence<sup>36</sup> but not validated in independent cohorts. Also, a signature on the basis of promoter hypermethylation of the five genes *HIST1H4F*, *PCDHGB6*, *NPBWR1*, *ALX1*, and *HOXA9* was recently reported to determine two subgroups associated with survival<sup>33</sup>; besides lung adenocarcinoma, the study also included SCLC with a different molecular pathomechanism.

Whereas it is well accepted that promoter hypermethylation and silencing of tumor suppressor genes are involved in carcinogenesis, DNA methylation is involved in different epigenetic processes genome-wide.<sup>37</sup> Altered transcription and gene regulation may also be influenced by DNA methylation of regions other than promoter regions of coding genes.<sup>19</sup> Furthermore, the spread of DNA methylation and its interaction with chromatin remodeling also play important roles.<sup>38,39</sup> The methylation patterns unraveled in this study are in line with these observations, in that, just 10% of DMRs differentiating the two subgroups are annotated in transcription start site regions (Supplementary Fig. 2), as opposed to more than 40% of DMRs mapping to intergenic regions. None of the previously suggested candidate genes in promoter regions were among the best 100 DMRs, underscoring the importance of sequencing-based genome-wide methylation analyses, whereas the Illumina Human-Methylation 450k BeadChip just covers 1.7% of all CpGs in the whole genome and is focused on promoter and CpG island regions.

In this study, promoter CpG island hypermethylation of the lncRNA *PLUT* was found to correlate significantly with early recurrence. This region has a high density of transcription factor binding sites and is enriched for H3K4me1 histone marks in the encyclopedia of DNA elements data (ENCODE) set. Owing to the shortness of tissue material, the expression of the lncRNA *PLUT* could not be investigated and correlated to the methylation status of its promoter region in this study. However, aberrant methylation of this genomic region is present in all disease stages of patients with lung adenocarcinoma when looking at publicly available The Cancer Genome Atlas Program (TCGA) Infinium HumanMethylation450 BeadChip data, implying a role in carcinogenesis (Supplementary Fig. 5). However, its possible role and downstream molecular mechanism in lung adenocarcinoma remains to be determined.

Methylation analysis at individual CpG sites within the DMRs revealed variations in methylation and

detected single CpGs discriminating the two subgroups with a higher statistical significance than the whole amplicon. The functional regulation of single CpG sites is not yet well-known and needs further investigation, though the importance of single CpG sites was previously commented on.<sup>18,40</sup> However, functional studies and causal explanations of aberrant single CpG methylation in lung cancer are rare.

lncRNAs are believed to play critical regulatory roles as tumor suppressors or oncogenic drivers.<sup>41,42</sup> In lung cancer, MALAT1, HOTAIR, HNF1A-AS1, and linc00673 are believed to increase proliferation and invasiveness by regulating protoadhesion factors (e-cadherin), transcription factors, and reorganization of chromatin and by acting as oncogenic lncRNAs.<sup>43-47</sup> Conversely, the lncRNAs MEG3 and PANDAR are believed to inhibit tumor growth by regulating TP53 or BCL2 or by inhibition of epithelial-mesenchymal transition.<sup>45,48,49</sup>

lncRNA PLUT(O) (PDX1-associated lncRNA, up-regulator of transcription) was recently found to regulate the transcription of *PDX1* by promoting contacts between the *PDX1* promoter region and enhancer clusters in human pancreatic  $\beta$ -cells.<sup>50</sup> In this study, *PDX1* expression in the discovery cohort revealed differences; however, sample numbers were too small for correlation analysis with methylation data of the lncRNA PLUT.

PDX1 is a homeobox transcription factor and a key regulator in the embryonic development of the pancreas and pancreatic  $\beta$ -cell function and also in the differentiation of the gastric mucosa and pseudopyloric glands.<sup>51,52</sup> For pancreatic cancer, there is evidence that PDX1 acts as an oncogene and regulates cell proliferation and invasion<sup>45,46</sup>; its overexpression is associated with poor survival.<sup>39,53</sup> In addition to pancreatic cancer, overexpressed PDX1 is associated with several solid tumors, including breast, colon, prostate, and kidney carcinoma.<sup>54</sup> Its role in tumorigenesis remains unclear and may differ among tissues. In gastric cancer, the expression of *PDX1* is aberrantly reduced by promoter hypermethylation and histone deacetylation, and PDX1 is, therefore, considered a tumor suppressor.<sup>55,56</sup> We postulate that the transcription of *PLUT* among others is regulated by methylation in lung adenocarcinoma. Hypermethylation of its promoter region resulted in reduced expression, and subsequent translation of the gene *PDX1* was unhindered, resulting in a higher expression in patients with early relapse in our discovery cohort. To our knowledge, this is the first study revealing an association of the methylation status of the promoter region of *PLUT* with lung adenocarcinoma. Whether aberrant methylation of this gene region mediates tumorigenesis in lung adenocarcinoma mainly by transcriptional regulation of *PLUT* and subsequent translational modulation of PDX1 or whether other

downstream molecular mechanisms are involved is not known. The idea of using *PDX1* or other *PLUT*-induced genes as therapeutic molecular targets is certainly compelling.

However, the tremendous advances made in deciphering the molecular mechanisms involved in lung tumorigenesis have not yet translated into better prognosis for patients, especially those with early stage tumors for whom cure is most feasible. Although new sensitive screening methods like low-dose computed tomography may increase the numbers of patients diagnosed in early stage I disease,<sup>57</sup> the need for prognostic biomarkers to identify newly diagnosed patients with a high risk of recurrence will gain further importance. The results of ongoing studies with EGFR tyrosine kinase inhibitors (ALCHEMIST; NCT02201992, NCT02193282) or immunotherapy (PEARLS/KEYNOTE-09, NCT02504372) are eagerly awaited; however, these will only be applicable for a subgroup of patients. This study reveals that promoter hypermethylation of the lncRNA PLUT significantly correlates with early recurrence, especially in male patients. Whether the methylation status of this genomic region will serve as another puzzle piece for implementation in personalized medicine as a biomarker for adapted treatment modalities in early stage lung cancer to overcome bad prognosis has to be proved in further studies.

## Acknowledgments

This study was funded by the Heidelberg Center for Personalized Oncology (DKFZ-HIPO) (project H004) and partly funded by the German Center for Lung Research (DZL) 82DZL00402. The authors thank the Genomics and Proteomics Core Facility of the German Cancer Research Center for providing and performing the high-throughput next-generation sequencing; Oliver Mücke, Marion Bähr, Stefanie Herkt, Elisabeth Ehrend, Annekarin Meyer for their technical assistance; and Jörn Lausen for providing the plasmids.

## Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *Journal of Thoracic Oncology* at [www.jto.org](http://www.jto.org) and at <https://doi.org/10.1016/j.jtho.2020.03.023>.

## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin*. 2017;67:7-30.
2. Jemal A, Ward EM, Johnson CJ, et al. Annual report to the nation on the status of cancer, 1975-2014, featuring survival. *J Natl Cancer Inst*. 2017;109:djx030.



3. Office for National Statistics. Cancer survival by stage at diagnosis for England: adults diagnosed 2012, 2013 and 2014 and followed up to 2015; 2016. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalbystageatdiagnosisforenglandexperimentalstatistics/adultsdiagnosed20122013and2014andfollowedupto2015>. Accessed January 19, 2019.
4. Nesbitt JC, Putnam JB Jr, Walsh GL, Roth JA, Mountain CF. Survival in early-stage non-small cell lung cancer. *Ann Thorac Surg*. 1995;60:466-472.
5. Berger AH, Brooks AN, Wu X, et al. High-throughput phenotyping of lung cancer somatic mutations. *Cancer Cell*. 2016;30:214-228.
6. Imielinski M, Berger AH, Hammerman PS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012;150:1107-1120.
7. Keller SM, Adak S, Wagner H, et al. A randomized trial of postoperative adjuvant therapy in patients with completely resected stage II or IIIA non-small-cell lung cancer. Eastern Cooperative Oncology Group. *N Engl J Med*. 2000;343:1217-1222.
8. Artal CA, Calera UL, Hernando CJ. Adjuvant chemotherapy in non-small cell lung cancer: state-of-the-art. *Transl Lung Cancer Res*. 2015;4:191-197.
9. Douillard JY, Rosell R, De Lena M, et al. Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIA non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. *Lancet Oncol*. 2006;7:719-727.
10. Felip E, Rosell R, Maestre JA, et al. Preoperative chemotherapy plus surgery versus surgery plus adjuvant chemotherapy versus surgery alone in early-stage non-small-cell lung cancer. *J Clin Oncol*. 2010;28:3138-3145.
11. Gilligan D, Nicolson M, Smith I, et al. Preoperative chemotherapy in patients with resectable non-small cell lung cancer: results of the MRC LU22/NVALT 2/EORTC 08012 multicentre randomised trial and update of systematic review. *Lancet*. 2007;369:1929-1937.
12. Pisters KM, Vallieres E, Crowley JJ, et al. Surgery with or without preoperative paclitaxel and carboplatin in early-stage non-small-cell lung cancer: Southwest Oncology Group Trial S9900, an intergroup, randomized, phase III trial. *J Clin Oncol*. 2010;28:1843-1849.
13. Strauss GM, Herndon JE 2nd, Maddaus MA, et al. Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non-small-cell lung cancer: CALGB 9633 with the Cancer And Leukemia Group B, Radiation Therapy Oncology Group, and North Central Cancer Treatment Group Study Groups. *J Clin Oncol*. 2008;26:5043-5051.
14. Pignon JP, Tribodet H, Scagliotti GV, et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. *J Clin Oncol*. 2008;26:3552-3559.
15. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499:214-218.
16. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511:543-550.
17. Butts CA, Ding K, Seymour L, et al. Randomized phase III trial of vinorelbine plus cisplatin compared with observation in completely resected stage IB and II non-small-cell lung cancer: updated survival analysis of JBR-10. *J Clin Oncol*. 2010;28:29-34.
18. Buckingham L, Penfield FL, Kim A, et al. PTEN, RASSF1 and DAPK site-specific hypermethylation and outcome in surgically treated stage I and II nonsmall cell lung cancer patients. *Int J Cancer*. 2010;126:1630-1639.
19. Lokk K, Voorder T, Kolde R, et al. Methylation markers of early-stage non-small cell lung cancer. *PLoS One*. 2012;7:e39813.
20. Guo Y, Zhang R, Shen S, et al. DNA methylation of LRRC3B: a biomarker for survival of early-stage non-small cell lung cancer patients. *Cancer Epidemiol Biomarkers Prev*. 2018;27:1527-1535.
21. Han L, Xu G, Xu C, Liu B, Liu D. Potential prognostic biomarkers identified by DNA methylation profiling analysis for patients with lung adenocarcinoma. *Oncol Lett*. 2018;15:3552-3557.
22. Selamat SA, Chung BS, Girard L, et al. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res*. 2012;22:1197-1211.
23. Lenka G, Tsai MH, Lin HC, et al. Identification of methylation-driven, differentially expressed STXBP6 as a novel biomarker in lung adenocarcinoma. *Sci Rep*. 2017;7:42573.
24. Wei Y, Liang J, Zhang R, et al. Epigenetic modifications in KDM lysine demethylases associate with survival of early-stage NSCLC. *Clin Epigenetics*. 2018;10:41.
25. Carvalho RH, Haberle V, Hou J, et al. Genome-wide DNA methylation profiling of non-small cell lung carcinomas. *Epigenetics Chromatin*. 2012;5:9.
26. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta C(T)) method. *Methods*. 2001;25:402-408.
27. Sonnet M, Claus R, Becker N, et al. Early aberrant DNA methylation events in a mouse model of acute myeloid leukemia. *Genome Med*. 2014;6:34.
28. Heilmann K, Toth R, Bossmann C, Klimo K, Plass C, Gerhauser C. Genome-wide screen for differentially methylated long noncoding RNAs identifies Esrp2 and lncRNA Esrp2-as regulated by enhancer DNA methylation with prognostic relevance for human breast cancer. *Oncogene*. 2017;36:6446-6461.
29. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589-595.
30. van den Boom D, Ehrich M. Mass spectrometric analysis of cytosine methylation by base-specific cleavage and primer extension methods. *Methods Mol Biol*. 2009;507:207-227.
31. Kuvardina ON, Herkt S, Meyer A, et al. Hematopoietic transcription factors and differential cofactor binding regulate PRKACB isoform expression. *Oncotarget*. 2017;8:71685-71698.
32. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*. 1999;18:2529-2545.



33. Sandoval J, Mendez-Gonzalez J, Nadal E, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. *J Clin Oncol*. 2013;31:4140-4147.
34. Sato T, Arai E, Kohno T, et al. DNA methylation profiles at precancerous stages associated with recurrence of lung adenocarcinoma. *PLoS One*. 2013;8:e59444.
35. Pfeifer GP, Rauch TA. DNA methylation patterns in lung carcinomas. *Semin Cancer Biol*. 2009;19:181-187.
36. Brock MV, Hooker CM, Ota-Machida E, et al. DNA methylation markers and early recurrence in stage I lung cancer. *N Engl J Med*. 2008;358:1118-1128.
37. Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer*. 2004;4:143-153.
38. Berman BP, Weisenberger DJ, Aman JF, et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet*. 2011;44:40-46.
39. Quint K, Stintzing S, Alinger B, et al. The expression pattern of PDX-1, SHH, patched and Gli-1 is associated with pathological and clinical features in human pancreatic cancer. *Pancreatol*. 2009;9:116-126.
40. Claus R, Lucas DM, Stilgenbauer S, et al. Quantitative DNA methylation analysis identifies a single CpG dinucleotide important for ZAP-70 expression and predictive of prognosis in chronic lymphocytic leukemia. *J Clin Oncol*. 2012;30:2483-2491.
41. Tsai MC, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res*. 2011;71:3-7.
42. Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. *Mol Cancer*. 2011;10:38.
43. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010;464:1071-1076.
44. Qiu M, Xu Y, Yang X, et al. CCAT2 is a lung adenocarcinoma-specific long non-coding RNA and promotes invasion of non-small cell lung cancer. *Tumour Biol*. 2014;35:5375-5380.
45. Han L, Zhang EB, Yin DD, et al. Low expression of long noncoding RNA PANDAR predicts a poor prognosis of non-small cell lung cancer and affects cell apoptosis by regulating Bcl-2. *Cell Death Dis*. 2015;6:e1665.
46. Ma C, Wu G, Zhu Q, et al. Long intergenic noncoding RNA 00673 promotes non-small-cell lung cancer metastasis by binding with EZH2 and causing epigenetic silencing of HOXA5. *Oncotarget*. 2017;8:32696-32705.
47. Wu Y, Liu H, Shi X, Yao Y, Yang W, Song Y. The long non-coding RNA HNF1A-AS1 regulates proliferation and metastasis in lung adenocarcinoma. *Oncotarget*. 2015;6:9160-9172.
48. Lu KH, Li W, Liu XH, et al. Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression. *BMC Cancer*. 2013;13:461.
49. Sun M, Liu XH, Wang KM, et al. Downregulation of BRAF activated non-coding RNA is associated with poor prognosis for non-small cell lung cancer and promotes metastasis by affecting epithelial-mesenchymal transition. *Mol Cancer*. 2014;13:68.
50. Akerman I, Tu Z, Beucher A, et al. Human pancreatic  $\beta$  cell lncRNAs control cell-specific regulatory networks. *Cell Metab*. 2017;25:400-411.
51. Larsson LI, Madsen OD, Serup P, Jonsson J, Edlund H. Pancreatic-duodenal homeobox 1 - role in gastric endocrine patterning. *Mech Dev*. 1996;60:175-184.
52. Offield MF, Jetton TL, Labosky PA, et al. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development*. 1996;122:983-995.
53. Koizumi M, Doi R, Toyoda E, et al. Increased PDX-1 expression is associated with outcome in patients with pancreatic cancer. *Surgery*. 2003;134:260-266.
54. Wang XP, Li ZJ, Magnusson J, Brunicaudi FC. Tissue MicroArray analyses of pancreatic duodenal homeobox-1 in human cancers. *World J Surg*. 2005;29:334-338.
55. Ma J, Chen M, Wang J, et al. Pancreatic duodenal homeobox-1 (PDX1) functions as a tumor suppressor in gastric cancer. *Carcinogenesis*. 2008;29:1327-1333.
56. Ma J, Wang JD, Zhang WJ, et al. Promoter hypermethylation and histone hypoacetylation contribute to pancreatic-duodenal homeobox 1 silencing in gastric cancer. *Carcinogenesis*. 2010;31:1552-1560.
57. Jemal A, Fedewa SA. Lung cancer screening with low-dose computed tomography in the United States-2010 to 2015. *JAMA Oncol*. 2017;3:1278-1281.