Supplementary methods

Analysis of genome wide DNA methylation levels by Reduced Representation Bisulfite sequencing.

A total 0.3 to 1 µg of DNA was used for RRBS library preparation using published protocols with minor modifications (1, 2). Briefly, genomic DNA was digested with MspI, end-repaired and A-tailed with the Klenow-fragment enzyme and ligated with Illumina TruSeq adapters (Illumina, San Diego, CA, USA). Fragments in a range of 40 to 280 bps insert size were gel-purified. Libraries were bisulfite converted using the EZ DNA Methylation™ Kit (Zymo Research Europe, Freiburg, Germany) and amplified using PfuTurboCx DNA polymerase (Agilent Technologies, Santa Clara, CA, USA). The libraries were sequenced on an Illumina HiScanSQ instrument with version 3 sequencing chemistry. Libraries were spiked with 45% PhiX DNA to counteract the imbalance in nucleotide representation.

Human genomic sequences (hg19) and other tracks (eg, RefSeq genes) were downloaded from the University of California Santa Cruz Genome Browser database (3). Adapter sequences were removed using Cutadapt Version 0.9.3 (4) and sequences were mapped to hg19 genome using Bismark Version 0.5 (5). Methylation calls from Bismark were extracted with a modified script that removed 3′-MspI sites. We calculated the conversion rate of all non-CpG cytosine positions from the Bismark methylation_extractor output. All samples had a good conversion rate of at least 99% of all non-CpG cytosine positions being converted to uracil (Supplementary Table S3).

Methylation data were analyzed in R/Bioconductor with help of the BiSeq package (6). Differentially methylated region detection was restricted to regions with a high

CpG-site density covered across all samples (CpG clusters); methylation raw data were smoothed via locally weighted binomial likelihood (2). Smoothed methylation levels (between 0 and 1) were calculated every 10 bps within each CpG cluster. Regions with methylation differences of at least 30% were called differentially methylated regions (DMRs; minimum 1 base pair).

To determine if regions of interest were over- or underrepresented in DMRs, we supposed that the number k of region centers within DMRs is binomially distributed with n = number of region centers in CpG clusters and p = sum of DMR widths divided by the sum of CpG cluster widths. We then carried out a two-sided binomial test that $k \neq p \times n$. This approach was used for Figures 2B, 5A (left and right).

Illumina methylation bead arrays

Bisulfite conversion of genomic DNA was performed using the EZ DNA Methylation Kit (Zymo Research Europe, Freiburg, Germany). A total of 500 ng of converted DNA was hybridized to Infinium HumanMethylation BeadChips (Illumina, CA, USA) and scanned using the iScan instrument (Illumina). Data preprocessing and methylation-level extraction was done using Genome Studio Version 2011.1 software, including Methylation Module Version 1.9.0 and Illumina Genome Viewer Module Version 1.9.0 (Illumina).

Gene expression analysis

We used the human Gene 1.0 ST Array (Affymetrix, CA, USA) using parental and highly metastatic A549 cell lines according to the manufacturer´s instructions. Arrays

were scanned at 1.56 µm resolution using the Affymetrix GeneChip Scanner 3000. Raw gene expression data was imported to the Affymetrix expression console and subject to robust multi-array average (RMA). Differential gene expression was calculated with RankProd software (7).

Single nucleotide polymorphism analysis

We used high-resolution single nucleotide polymorphism (SNP-) array (SNP Array, Affymetrix, 6.0) method to screen for copy-number alterations in the genome of parental, as well as highly metastatic NSCLC cell lines using A549 and HTB56 similar to published protocols (8, 9). SNP array raw data are available at NCBI *Gene Expression Omnibus* (GSE44549).

Exome Capture and High-throughput Sequencing

The SureSelect XT Human All Exon V5+UTR kit (Agilent) for Illumina sequencing was used following the manufacturer's standard protocol. Exome enriched Libraries were subject to 2 x 100 cycles paired-end sequencing on a HiScanSQ instrument (Illumina).

Exome-Seq Processing-Read Mapping, Variant Calling and Effect Determination

The Burrows-Wheeler alignment algorithm (BWA) [PMID:19451168] was applied to map raw reads from the Illumina HiSeq2000 to the hg19 genome. Then, we removed duplicate reads and used the Genome Analysis Toolkit (GATK) [PMID:21478889] to apply local realignment, base quality score recalibration and SNP and INDEL

discovery across all 4 samples simultaneously. Standard hard filtering parameters (as suggested in the Best Practice Variant Detection using GTAK v4 documentation) were applied to remove low quality mutations. Finally, detected variations were transferred to the Ingenuity© Variant Analysis (CA, USA) online software.

Raw reads from the Illumina HiScanSQ were mapped to the hg19 genome using default parameters The high quality filtered data were aligned with human reference genome and variants were called against (LifeScope suite). Then, variations (vcf. file) were transferred to Ingenuity© Variant Analysis (CA, USA) online software Exome Seq. raw data are available at NCBI *Gene Expression Omnibus* (GSE XX pending).

Global DNA methylation analysis

Genomic DNA was extracted from parental and highly metastatic A549 cell lines with DNAzol (Invitrogen, Carlsbad, CA, USA). Global methylation levels were determined by capillary electrophoresis as described (10).

Functional in vitro assays – Analysis of double strand break by H2AX staining

Determination of H2AX-positive cells was done according to the manufacturer instructions (Cell Signaling, Frankfurt/Main, Germany). Briefly, cells (1x10^6) were resuspended in PBS (pH 7.4) and fixed by adding formaldehyde to a final concentration of 4%. After incubation of 15 min at room temperature Triton-X100 was added to a final concentration of 0.1%. After incubation of 30 min at room temperature cells were washed twice with PBS (pH 7.4) + 0.5% BSA. Staining was

done for 60 min at room temperature in the dark with H2AX-PE antibody or IgG control (Cell signaling H2AX-PE#5763, IgG-control # 5742). After washing with PBS (pH 7.4) + 0.5% BSA cells were analyzed by flow cytometry. H2AX-positive cells were considered as apoptotic cell population.

Supplementary references

1. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008;454:766-70.

2. Schoofs T, Rohde C, Hebestreit K, Klein HU, Gollner S, Schulze I, et al. DNA methylation changes are a late event in acute promyelocytic leukemia and coincide with loss of transcription factor binding. Blood. 2013;121:178-87.

3. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, et al. The UCSC Genome Browser database: update 2011. Nucleic acids research. 2011;39:D876-82.

4. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads; 2011.

5. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571-2.

6. Hebestreit K, Dugas M, Klein HU. Detection of significantly differentially methylated regions in targeted bisulfite sequencing data. Bioinformatics. 2013;29:1647-53.

7. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics. 2006;22:2825-7.

8. Yamamoto G, Nannya Y, Kato M, Sanada M, Levine RL, Kawamata N, et al. Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. American journal of human genetics. 2007;81:114-26.

9. Thoennissen NH, Krug UO, Lee DH, Kawamata N, Iwanski GB, Lasho T, et al. Prevalence and prognostic impact of allelic imbalances associated with leukemic transformation of Philadelphia chromosome-negative myeloproliferative neoplasms. Blood. 2010;115:2882-90.

10. Rius M, Stresemann C, Keller D, Brom M, Schirrmacher E, Keppler D, et al. Human concentrative nucleoside transporter 1-mediated uptake of 5-azacytidine enhances DNA demethylation. Molecular cancer therapeutics. 2009;8:225-31.