Destabilization of the human epigenome: consequences of foreign DNA insertions

Aim: We previously reported changes of DNA methylation and transcription patterns in mammalian cells that carry integrated foreign DNA. Experiments were now designed to assess the epigenetic consequences of inserting a 5.6 kbp plasmid into the human genome. **Methods:** Differential transcription and CpG methylation patterns were compared between transgenomic and nontransgenomic cell clones by using gene chip microarray systems. **Results:** In 4.7% of the 28,869 gene segments analyzed, transcriptional activities were up- or downregulated in the transgenomic cell clones. Genome-wide profiling revealed differential methylation in 3791 of >480,000 CpG's examined in transgenomic versus nontransgenomic clones. **Conclusion:** The data document genome-wide effects of foreign DNA insertions on the epigenetic stability of human cells. Many fields in experimental biology and medicine employ transgenomic or otherwise genome-manipulated cells or organisms without considering the epigenetic consequences for the recipient genomes.

Keywords: destabilization • differential gene expression • differential methylation • DNA methylation • epigenetics • foreign DNA • transgenomes

Studies on mammalian cells with integrated viral (adenovirus type 12, Ad12) DNA [1–4] revealed that cellular genomes as foreign DNA recipients exhibited genome-wide alterations of DNA methylation patterns [5]. Similar changes were documented in cells carrying bacterial plasmid or bacteriophage λ DNA as foreign genome inserts [5,6]. As a corollary, transcriptional profiles in Ad12 DNA or λ DNA transgenomic hamster cell genomes were significantly altered [7]. It has been suggested earlier that these epigenomic destabilizations of transgenomic cells or organisms play a role in (viral) oncogenesis and possibly in evolution, since foreign DNA insertions in evolutionary times might have generated cells with novel functional profiles [8–10].

Additional evidence from several experimental systems also supports the notion of epigenetic destabilization due to foreign DNA integration. In Epstein–Barr virus (EBV)-transformed human lymphoblastoid

cell lines, in EBV-infected cells [11,12] or in human macrophages carrying integrated HIV-1 proviral genomes [13], cellular DNA methylation and transcription patterns have been significantly altered. In induced pluripotent human stem cells with multiple integrated foreign genes, genome-wide alterations of CpG methylation profiles have been described [14]. A detailed analysis of the methylation boundary [15] upstream of the human *FMRI* (fragile X mental retardation 1) gene in EBV- or telomerase gene-transformed human cells has revealed stability of the boundary itself, but extensive loss of CpG methylation far upstream of the boundary [16]. The same upstream *FMRI* region is CpG hypermethylated in nontransgenomic human cells.

In the present report, we have documented widespread alterations in cellular transcription and methylation patterns in human cells that carry a 5.6 kbp bacterial plasmid as integrated transgenome. SNP

Stefanie Weber^{†,1}, Andrea Hofmann^{†,2}, Stefan Herms^{2,3}, Per Hoffmann^{2,3} & Walter Doerfler^{*,1,4}

¹Institute of Clinical & Molecular Virology, University of Erlangen–Nürnberg Medical School, D-91054 Erlangen, Germany

²Institute of Human Genetics, Life & Brain Center, Bonn University, D-53127 Bonn, Germany

³Division of Medical Genetics, University Hospital Basel, CH-4055 Basel, Switzerland

⁴Institute of Genetics, University of Cologne, D-50674 Cologne, Germany

*Author for correspondence:

Tel.: +49 9131 852 6002

walter.doerfler@

viro.med.uni-erlangen.de

[†]Authors contributed equally

Future
Medicine

part of
fsg

array karyotyping failed to detect major chromosomal alterations in transgenomic as compared with nontransgenomic cells.

Methods

Cells & cell culture

The human colon tumor cell line HCT116 [17] was cultivated under standard conditions at 37°C using Dulbecco modified medium supplemented with 10% fetal bovine serum in a 5% CO₂, 95% air atmosphere. Individual cell clones of nontransgenomic or of plasmid-transgenomic cells were isolated by the serial dilution method and recloned twice in 96-well plastic dishes.

Bacterial plasmid & transfection of HCT116 cells

The structure of the plasmid pC1–5.6 is shown in Figure 1A. Plasmid pC1–5.6 was constructed in the following way: From plasmid pEGFP-C1 (Clontech Laboratories Inc. Mountain View, CA, USA), the *EGFP* gene and the CMV promoter were removed and a 2298 bp *FMR1* segment devoid of coding sequences was inserted. The pC1–5.6 plasmid was purified from bacterial strain DH5α by standard procedures [18]. HCT116 cells (ATCC Cat. No. CCL-247) were transfected with plasmid pC1–5.6 by using the nucleofection protocol Cell Line Nucleofector® Kit V which had been optimized by Amaxa Biosystems Inc. (Cologne, Germany), specifically for this cell line. By following the manufacturer's instructions, transfection efficiency was reported to be about 80%. Upon transfection, kanamycin resistant cells were selected and continuously maintained in the presence of 150 µg of Geneticin (G418) per ml culture medium. Successfully transfected HCT116 clones were identified by analyzing the total intracellular DNA from the transgenomic clones by the polymerase chain reaction (PCR) using the following primer pairs (Figure 1B & C): Kanamycin: Kan.pC1–5.6 forward 5'-ATGATTGAACAAGATGGATT-3'; Kan.pC1–5.6 reverse 5'-TCAGAAGAACTCGTCAACAA-3'. Plasmid genome integration was ascertained by PCR using the primer pair: Alu – pC1–5.6: Alu A3: 5'-GACTGCACTCCAGCCTGGGC-3'; and pC1–5.6: 5'-GAATAGACCGAGATAGGGTTGAGTGTTG-3' [19].

Extraction of DNA & RNA

DNA from cultured cells was purified by phenol/chloroform extraction and ethanol precipitation [1]. Total RNA from cultured cells was isolated by the RNeasy® Mini Kit from Qiagen and was treated with DNase by using the 'RNase-Free DNase Set' again from Qiagen and the manufacturer's protocol.

Molecular karyotyping by SNP microarrays

Illumina's CoreExome BeadChip was used for molecular karyotyping. Genotyping was performed according to the manufacturer's manual on a fully automated iScan system. Briefly, the genomic DNA was amplified overnight, then enzymatically fragmented and subsequently hybridized to the array in another overnight step. After hybridization allele specificity was achieved by a single base extension using chemically modified deoxynucleoside triphosphates (dNTPs). After staining, the arrays were read out by using an Illumina iScan system. Analyses were performed by using Genome Studio (version 2011.1) and copy number variation (CNV) partition (version 3.1.6).

Affymetrix gene chip microarray assays

Sample preparations for microarray hybridization were carried out as described in the Ambion WT Expression Kit Protocol (Thermo Scientific) and the Affymetrix WT Terminal Labeling and Hybridization User Manual (Affymetrix, Inc., CA, USA). In brief, 300 ng of total RNA were used to generate double-stranded cDNA, followed by an *in vitro* transcription reaction. Twelve micrograms of RNA were purified and reverse transcribed into deoxyuridine triphosphate-(d-UTP)-containing sense-strand-(ss)-cDNA. Purified ss-cDNA was fragmented by using a combination of uracil DNA glycosylase and apurinic/apyrimidinic endonuclease 1 (APE 1), followed by terminal labeling with biotin. An amount of 3.8 µg of fragmented and biotin-labeled ss-cDNA was hybridized to Affymetrix Human Gene 1.0 ST arrays for 16 h at 45°C in a rotating chamber. Hybridized arrays were washed and stained in an Affymetrix Fluidics Station FS450, and the fluorescent signals were measured with an Affymetrix GeneChip Scanner 3000–7G. According to the Affymetrix manual, the GeneChip® Human Gene 1.0 ST Array for Gene Expression Analysis interrogates 28,869 annotated human genes with 764,885 distinct probes, in other words, an average of 26 probes per gene with 25-mer oligonucleotides. Sample processing was performed at an Affymetrix Service Provider and Core Facility, KFB – Center of Excellence for Fluorescent Bioanalytics, Regensburg, Germany. The laboratory that performed the activity determinations (KFB, Regensburg) had not been informed about the nature of these analyses or the derivation of the RNA samples.

Methylation assays

DNA was sodium-bisulfite converted using the Eazy DNA Methylation kit (Zymo Research, Freiburg,

Figure 1. The pC1-5.6 plasmid used as transgenome. (A) Genetic map of the plasmid. For details see text. **(B)** Results of PCR analyses on the DNA from HCT116 cell clones as indicated for the presence of the transgenome. Primers placed inside the plasmid were described under Methods. Lanes 01–03 – DNA from nontransgenic control clones; lanes 04–08 – DNA from plasmid pC1-5.6-transgenic cell clones. **(C)** PCR results on DNA from cell clones as indicated using one primer inside the plasmid, the second one in an Alu sequence. The positioning of one primer in Alu sequences of the neighboring cellular DNA and of the second primer inside the plasmid sequence facilitated detection of the integrated plasmid genomes. All primers used were listed under section Methods.

Germany). Samples were processed on the Human Methylation 450 K Bead Chips from Illumina (San Diego, CA, USA) according to the manufacturer's manual. The laboratory process was already described under molecular karyotyping.

Expression & methylation data analyses

Expression assay quality control was monitored through hybridization controls. Summarized expression probe set signals were calculated by the RMA algorithm [20] using Bioconductor in R (statistical programming language). Multiple probes on different exons were summarized into accumulative expression values representing all transcripts from the same gene. Genes with a fold change above twofold and a Student's *t*-test false discovery rate (FDR) adjusted *p*-value < 0.05 were considered as significantly regulated.

For the analysis of the methylation data, a subset quantile normalization approach was applied [21]. This processing pipeline included signal corrections for the adjustment of color balance and background level corrections as well as the Infinium I/ Infinium II shift correction between sample normalization. By using the 'Genome Studio' software, technical quality parameters such as hybridization, extension, bisulfite conversion and specificity were evaluated. Signal distributions of β -values, representing a continuous measurement for methylation levels were inspected by density plots. Sample variation was visualized by unsupervised hierarchical cluster and principal component analysis. The data were analyzed by using the Bioconductor in R program. Cross-reactive probes and polymorphic CpG's as well as CpG's with SNP's in the CpG 50-mer probe defined by [22] were removed

leaving 361,983 CpG's for analysis. Differentially methylated CpG loci were identified using a *t*-test; p-values were corrected for multiple testing using Benjamin-Hochberg corrections [23]. A difference in mean β -values of 0.2 and FDR-adjusted p-values of < 0.05 were chosen as thresholds to call differential methylation. Pathway analysis of differentially expressed genes and differentially methylated CpG's was performed using Ingenuity Pathway Analysis.

Results & discussion

Design of experiments & structure of pC1–5.6 plasmid transgenome

Experiments with a human cell line were designed to assess the consequences of the integration of a plasmid transgenome on transcription and methylation patterns in the recipient genomes. Five independently isolated and propagated nontransgenic single cell clones of the human colon tumor cell line HCT116 [17] were analyzed as controls. These cells showed nearly identical expression (mean $r^2 = 0.99$, Figure 2) and very comparable methylation profiles (mean $r^2 = 0.97$, see below). This finding documented

the stability and similarity of cellular methylation and transcription profiles in individual cell clones. These results established the basis for comparisons of these profiles to those of seven plasmid-transgenic cell clones.

The 5648 bp bacterial plasmid pC1–5.6 (Figure 1A) served as transgenome that contained the kanamycin/neomycin resistance gene under early SV40 promoter control as the sole functional eukaryotic promoter in the construct. The plasmid also carried the bacterial kanamycin promoter and the nontranslated first exon with parts of the 5'-UTR and the CGG repeat of the human fragile X mental retardation gene 1 (*FMR1*) [24]. Total length of the *FMR1* fragment was 2298 bps. Persistence and integrated state of this plasmid in transgenic cell clones were ascertained by kanamycin resistance of the cell clones and by PCR (Figure 1B & C). Primers selected for PCR were listed under the section Methods. The positioning of one primer in Alu sequences of the neighboring cellular DNA sequences and of the second primer inside the plasmid sequence facilitated documentation of the integrated state of the plasmid genomes (Figure 1C).

Figure 2. Transcriptional activities in five cell clones of nontransgenic HCT116 cells. (A–E) Scatter plots of double determinations from five different single HCT116 cell clones. X and Y axes display the same clones analyzed in double determinations. See text for details.

The positive PCR products in Figure 1C which were derived from the DNA of transgenomic clones 2, 3, 5, 6 and 8 were all different in size and in part heterogeneous in lengths. They span the plasmid sequence and the Alu primer-targeted sequences of the abutting

integration sites (Figure 1C). The profiles indicate that the insertion sites of plasmid pC1–5.6 are different in each of these clones. Moreover, there might be more than one integration site per clone, because some of the PCR products indicate size heterogeneity.

Table 1. Differentially expressed genes.					
Probe_ID	LogFC	p-value	Gene accession	Gene symbol	Symbol_affy
8009380	4.55	1.1E-07	NR_003706	SNORA38B	
8066262	4.43	1.7E-09	NR_003018	SNORA71D	
8034512	4.43	6.8E-08	NR_002751	SNORD41	
8135943	4.33	1.7E-06	–	–	RNA5SP242
7938291	3.81	6.8E-08	NR_002580	SNORA3	
7896746	3.72	1.9E-06	–	–	FAM156A
7938329	3.67	3.7E-07	NR_002962	SNORA23	
8081233	3.47	1.7E-05	–	–	–
7967030	3.37	4.0E-08	NR_003925	RNU4-1	
7999404	3.27	7.2E-06	–	–	–
8136159	3.16	7.2E-06	–	–	–
7948894	3.15	4.5E-05	NR_002716	RNU2-1	
8165672	3.14	1.0E-04	L23320	RFC1	
8130578	3.14	5.8E-07	NR_002960	SNORA20	
8008132	3.11	4.0E-08	NM_005175	ATP5G1	
7961151	3.10	4.2E-06	NM_007360	KLRK1	
7914216	3.03	3.9E-09	NR_003035	SNORA16A	
7954434	-2.02	5.5E-07	–	–	–
7999362	-2.09	3.7E-06	–	–	RNA5SP403
7995320	-2.11	6.8E-08	–	–	–
7979698	-2.12	3.7E-07	NM_015994	ATP6V1D	
8031867	-2.17	4.9E-05	–	–	–
8160782	-2.18	3.0E-07	–	–	–
7942875	-2.20	1.1E-07	–	–	SLC12A3
8155497	-2.26	5.4E-08	NR_027421	FAM27C	
7970696	-2.31	5.8E-06	NM_182488	USP12	
8126512	-2.37	4.0E-08	NM_015950	MRPL2	
7995350	-2.51	9.9E-08	–	–	
8120057	-2.53	5.4E-07	–	–	
8081878	-2.56	4.3E-06	–	–	
7996260	-2.92	6.1E-06	–	–	
8165705	-3.07	1.7E-07	–	–	
8001457	-3.15	7.7E-08	NM_001025195	CES1	
8027884	-3.49	4.4E-09	–	–	RPS28
7915592	-3.74	1.3E-07	–	–	RNU5D-1

Shown are all differentially expressed genes between five nontransgenomic versus seven transgenomic cell clones with a FC > 3 and FC < -2; and a p < 0.05.
Affy: Affymetrix; FC: Fold change

Molecular karyotyping

To exclude major chromosomal abnormalities introduced through transfection of the pC1–5.6 plasmid into human cells, we analyzed all DNA samples from transgenic and nontransgenic cells by using SNP microarrays. Hybridization to the arrays allows measurements of chromosomal abnormalities through changes in the signal intensity of the probes (log-R ratio) and the distribution of heterozygous genotypes (b-allele frequency). By molecular karyotyping, shared copy number variations (CNV's) including duplications, deletions and copy-neutral loss of heterozygosity were detected across all samples (Supplementary Figure 1; for supplementary figures, see online at: www.futuremedicine.com/doi/full/10.2217/epi.15.40). However, comparisons between nontransgenic and transgenic cell clones revealed no significant clonal aberrations and high congruency across all samples.

Analysis of transcription activities in nontransgenic versus transgenic human cell clones

The results of comparisons of transcriptional activities between nontransgenic and transgenic cell clones can be critically interpreted only when independently isolated clones of nontransgenic cells have identical or very similar transcription patterns across the genome. We have isolated five clones of nontransgenic HCT116 cells and expanded and propagated them under identical conditions. From each of these cell populations total RNA was isolated by the Qiagen RNeasy® Mini Kit. By using an Affymetrix Expression Array system with fluorescent hybridization probes, transcriptional activities were measured in 28,869 genes covered by 764,885 distinct probes (see section Methods). Expression profiles of all five independently isolated and propagated nontransgenic cell clones analyzed in double determinations had a pair-wise Pearson's correlation coefficient above 0.98 for all transcripts measured on the microarrays (Figure 2). Moreover, the mean correlation coefficient was 0.99 across all samples with a minimum correlation of 0.97. We conclude that individual HCT116 cell clones exhibit identical, or very similar, transcription activities across the analyzed parts of the genome. This result renders meaningful the comparison of transcriptional activities between nontransgenic and plasmid-transgenic cell clones.

Differential gene expression was next assessed between mean values of two biological replicates from five nontransgenic control clones against seven plasmid-transgenic cell clones by a fold-change (FC)/p-value filter. Genes were called differentially expressed when the FC was above ± 2 and the FDR-adjusted p-value < 0.05. Among the 28,869 human

genes, 1343 genes were differentially expressed (Table 1), 907 were upregulated and 436 were downregulated in transgenic cell clones (Figure 3). Genes with the highest changes are summarized in Table 1. Gene set enrichment analysis was performed using Ingenuity Pathway Analysis for all identified differentially expressed genes (n = 1343 genes). A significant enrichment was found in 43 canonical pathways which included *EIF2* signalling, Regulation of *eIF4* and *p70S6K* signalling, Glutathione-mediated Detoxification, *FAK* signalling, Insulin Receptor signalling and *ErbB4* signalling (Table 2). The predominant upregulation effect was found in *snoRNA* (small nucleolar RNA) genes raising important questions about far-reaching effects on the biochemical modifications of many classes of RNA (ribosomal, transfer and small nuclear) and their consequences for cell function [25]. It is also worth noting that several signalling genes show differential expression. The data indicate that the insertion of foreign DNA, in other words, of bacterial plasmid pC1–5.6, into the recipient human cell genome led to alterations of transcriptional activities in 4.7% of the 28,869 analyzed genes.

Analysis of DNA methylation patterns in plasmid-transgenic versus nontransgenic human cell clones

Genome-wide methylation profiling was performed by assessing over 480,000 CpG sites in four nontransgenic clones and five pC1–5.6 transgenic clones using Illumina 450K Infinium arrays. Similarly to expression profiling, the mean correlation across all nontransgenic human cell clones was calculated as 0.97 ± 0.02 SD; thus rendering meaningful the comparison between nontransgenic and transgenic cell clones. Using both the stringent criteria of $\Delta\beta$ values ≥ 0.2 between nontransgenic and transgenic cell clones as well as a FDR-adjusted p < 0.05 in the comparison between both groups, 3791 CpG's were found differentially methylated. Of them, 1504 were hyper- and 2287 were hypomethylated in transgenic clones (Figure 4A & B). Most CpG's were located in gene bodies (29.91%), followed by transcription start sites (27.49%) as summarized in Figure 4C.

IPA pathway analysis using 2622 genes corresponding to the 3791 probes identified as differentially methylated revealed a significant enrichment in 109 canonical pathways including several important functional classes of genes associated with neuron signalling (Table 3). We conclude that in the pC1–5.6 plasmid transgenic cell lines, levels of CpG methylation are significantly altered as compared with methylation profiles in nontransgenic cells.

Figure 3. Alterations of transcriptional activities in pC1-5.6 transgenic HCT116 cell clones. (A) Volcano plot displays nonstandardized signal (log2 fold-change) on the X-axis against standardized signal (-log10 false discovery rate adjusted p-value) on the Y-axis for the comparison of five nontransgenic against seven transgenic cell clones of all 28,869 genes analyzed. Upregulated genes in transgenic cell clones were displayed in red and downregulated genes in blue (fold change ± 2 , adjusted p-values < 0.05 ; $n = 1343$ genes). **(B)** Hierarchical cluster analysis of differentially expressed genes. Parameters for cluster calculation: distance = ‘correlation’, linkage = ‘average’. Expression signals were z-transformed. Upregulated genes are indicated in carmine, downregulated genes in blue.

Conclusion

This study extends earlier work on changes of cellular transcription and methylation patterns in mammalian cells upon the chromosomal insertion of viral or plasmid DNAs [5-7,16]. We have now chosen a targeted approach and used human cells in culture with a small integrated bacterial plasmid (pC1-5.6), which carried the kanamycin gene under SV40 pro-

motor control. SNP karyotyping documented that, when compared with nontransgenic clones, the transgenic cells had not been subject to larger insertions, deletions or copy number variations in their exome.

Differential transcription and methylation was assessed by comparing transgenic cell clones to nontransgenic control cell clones. In 4.7% of the

Table 2. Top canonical pathways for differentially expressed genes.			
Ingenuity canonical pathways	-log(p-value)	Ratio	z-score
EIF2 signalling	7.94	0.162	3.130
Regulation of eIF4 and p70S6K signalling	2.82	0.11	2.121
Glutathione-mediated detoxification	2.65	0.238	NA
FAK signalling	2.56	0.13	NA
Insulin receptor signalling	2.45	0.108	NA
ErbB4 signalling	2.41	0.143	1.633

Shown are the six most highly enriched pathways derived from ingenuity pathway analysis (IPA) for all 1343 differentially expressed genes. The significance is expressed as a p-value calculated using the right-tailed Fisher's exact test; ratio is the overall ratio of differentially expressed genes from each pathway and the z-score predicts the activation state of the upstream regulator, using the gene expression patterns of the genes downstream of an upstream regulator. An absolute z-score of ≥ 2 is considered significant. An absolute z-score > 0 implies more upregulated genes in this pathway.
NA: Not applicable.

Figure 4. Epigenetic alterations in pC1–5.6 transgenic HCT116 cell clones. (A) Hierarchical cluster analysis of differentially methylated CpG's. Parameters for cluster calculation: distance = 'correlation,' linkage = 'average.' Upregulated genes are indicated in carmine, downregulated genes in blue. (B) Volcano plot displays differences in methylation on the X-axis against standardized methylation (-log10 false discovery rate adjusted p-value) on the Y-axis for the comparison of four nontransgenic against five pC1–5.6 transgenic cell clones of all 361,983 CpG's interrogated. Hypermethylated CpG's in transgenic cell clones are displayed in red and hypomethylated CpG's in blue ($\Delta\beta$ value ≥ 0.2 , adjusted p-value < 0.05 ; $n = 3,791$ CpG's). (C) Pie chart indicates structural characteristics of differentially methylated CpG sites. Methylation changes found in different genes have been assigned to specific parts of these genes: Body of gene (body); 5'UTR, 3'UTR, first exon are self-explanatory; NA not assigned; TSS transcription start site.

28,869 gene segments analyzed, the transcriptional activities were upregulated (907 genes) or downregulated (436 genes) in the cell clones that carried the bacterial plasmid pC1–5.6. A significant gene set enrichment was found in 43 canonical pathways. Genome-wide methylation profiling was performed for $> 480,000$ CpG sites in four nontransgenic and five pC1–5.6 transgenic clones. Stringent statistical analyses documented that in a comparison of methylation levels in transgenic versus nontransgenic cell clones 3791 CpG's were differentially

methylated, 1504 CpG's were hyper- and 2287 were hypomethylated in the transgenic clones. The different transgenic clones differ in their profiles. Viewed under the perspective of previously published investigations from our [5–7,16] and from other [11–14] laboratories, it is apparent that the insertion of foreign DNA into an established mammalian genome can lead to genome-wide destabilizations of cellular transcription and methylation patterns. We have not yet conducted a systematic investigation into the epigenetic trans effects of transgenome size, its

Table 3. IPA top canonical pathways for differentially methylated CpGs.			
Ingenuity canonical pathways	-log(p-value)	Ratio	z-score
Neuropathic pain signalling in dorsal horn neurons	6.4	0.27	1.347
Axonal guidance signalling	5.48	0.164	NA
CREB signalling in neurons	5.42	0.211	1.768
Glutamate receptor signalling	4.88	0.298	1.265
GABA receptor signalling	4.45	0.269	NA
Netrin signalling	4.43	0.333	NA
Details as described in the legend to Table 2.			

gene or CpG content or copy number of the integrated foreign DNA. From the sum total of all our previous investigations on this topic, we conclude that the integrations of Ad12, bacteriophage λ or plasmid DNA as well as the currently used pC1–5.6 plasmid have all led to alterations of methylation and transcriptional profiles in the recipient cells [5–7]. There is also evidence that cells carrying EBV genomes or the integrated gene for telomerase are subject to DNA methylation changes [16]. Hence, the integration of a number of quite different foreign genomes has elicited epigenetic destabilization which then does not appear to depend on one type of DNA sequence. These genome-wide alterations in DNA methylation patterns go far beyond local CpG methylation changes at the site of foreign DNA integration which were reported earlier [26,27].

The mechanism(s) underlying the observed epigenetic alterations are presently not understood. It is

conceivable that extent and location of alterations in genome activities and CpG methylation depend on the site(s) of foreign DNA insertion. In the transgenic cells analyzed here, the highest upregulations have been found in *snoRNA* (small nucleolar RNA) genes. A number of genes involved in signalling pathways have also been markedly affected. More general functional discussions on the affected genes will have to await analyses on the interdependency of foreign DNA integration site locations and affected genes. In this context it is of interest to note that massive reporter gene integration has been shown to lead to extensive variation in cellular gene expression levels [28]. Foreign DNA insertions into noncoding parts of the genome would possibly affect a multitude of genetic elements as identified by the ENCODE or FANTOM project [29,30]. In this way, a large number of cellular functions might be fundamentally altered.

Executive summary

- In studies on the integrated state of adenovirus type 12 (Ad12) DNA in Ad12-transformed hamster cells, we discovered that the CpG methylation profiles in some of their endogenous retrotransposon sequences and in several cellular genes were increased. This augmented methylation persisted in revertants of the transformed cells that had lost all Ad12 genomes. Moreover, alterations of DNA methylation and transcription profiles were documented in Ad12 DNA- and in bacteriophage λ DNA-transgenic cells.
- We previously hypothesized that epigenetic effects in mammalian genomes due to the insertion of foreign DNA are a general phenomenon. These alterations might play a role in (viral) oncogenesis and are possibly instrumental during evolution as a consequence of multiple retroviral DNA insertions into ancient genomes. Over evolutionary times, these alterations of transcription profiles might have led to novel phenotypes that were then selected for or against depending on environmental conditions during evolution.
- To examine the general significance of these observations, we designed a model system for proof of principle assessment. Human cells from cell line HCT116 were rendered transgenic by transfecting a 5.6 kbp bacterial plasmid and selecting cell clones with foreign plasmids stably integrated, most likely at different genomic sites.
- In five nontransgenic HCT116 control clones without the plasmid, transcription and methylation patterns proved similar, if not identical, among individual cell clones. This finding opened the possibility for comparisons of these patterns between nontransgenic and transgenic clones.
- Molecular karyotyping demonstrated shared copy number variations including duplications, deletions and copy-neutral loss of heterozygosity across all DNA samples. A comparison of nontransgenic and transgenic clones, however, revealed no significant clonal aberrations but high congruency.
- In 4.7% of the 28,869 gene segments analyzed, the transcriptional activities were upregulated (907 genes) or downregulated (436 genes) in plasmid-transgenic cell clones in comparison to control clones. A significant gene set enrichment was found in 43 canonical pathways. Frequent upregulations were noted in small nucleolar RNA genes that regulate RNA metabolism and in genes involved in signalling pathways.
- Genome-wide methylation profiling was performed for >480,000 CpG sites. In comparisons of methylation levels in five transgenic versus four nontransgenic cell clones, 3791 CpG's were differentially methylated, 1504 CpG's were hyper- and 2287 were hypo-methylated.
- Thus, the epigenetic effects in the wake of foreign DNA integration events can be considered a general effect also in human cells. We still lack insights into the role of transgenome size, gene or CG content or copy number. The mechanism(s) underlying the observed epigenetic alterations are unknown. Extent and location of alterations in genome activities and CpG methylation might depend on the site(s) of foreign DNA insertion.
- We note that genome manipulations in general – work with transgenic or knocked cells and organisms – have assumed a major role in molecular biology and medicine. The consequences of cellular genome manipulations for epigenetic stability have so far received unwarrantedly limited attention. Before drawing far-reaching conclusions from work with cells or organisms with manipulated genomes, critical considerations for and careful analyses of their epigenomic stability will prove prudent.

Future perspective

In published investigations, work on the epigenetic consequences of genome manipulations in mammalian (human) cells, including the insertion of foreign DNA, has so far received almost no attention. There are, however, numerous unresolved problems in this context, and one might expect increased research activities along these lines. By the same token, genome manipulations, the addition of foreign DNA to mammalian cells by various methods of transfection, and the selection of transgenic or knocked cells and organisms, have been everyday practice in a large number of laboratories. With further refinement of epigenetic technologies, it will be but a question of (possibly a very short) time that hitherto unsuspected complications in the evaluation of experiments with genome manipulated cells and organisms will come to the fore. It is the intent of this article, not only to document further these complexities of epigenetic mechanisms, but to alert researchers in all fields of biomedicine to possibly far-reaching problems with their interpretations.

Acknowledgements

The transcriptional profiles (primary data) were determined by KFB in Regensburg, Germany. We are indebted to the Institute of Clinical and Molecular Virology, University of Erlangen-Nürnberg Medical School for their continued support of W Doerfler's senior research group. S Weber performed most of the laboratory experiments and was in-

involved in the planning and interpretation of the project. A Hofmann performed all statistical analyses, was involved in the interpretation of data and wrote part of the manuscript. S Weber and A Hofmann have contributed equally to this work. S Herms performed molecular karyotyping. P Hoffmann carried out methylation profiling, interpreted data and wrote part of the manuscript. W Doerfler initiated and planned the project, was involved in the interpretation of data and wrote the manuscript. KFB in Regensburg did the transcription analyses.

Financial & competing interests disclosure

This research was made possible by grants to W Doerfler from the Thyssen Foundation, Köln (Az. 10.07.2.138.) and from the Deutsche Forschungsgemeinschaft, Bonn (DO 165/28-1). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open Access

This article is distributed under the terms of the Creative Commons Attribution License 4.0 which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited. To view a copy of the license, visit <http://creativecommons.org/licenses/by/4.0/>

References

Papers of special note have been highlighted as:

•• of considerable interest

- 1 Sutter D, Westphal M, Doerfler W. Patterns of integration of viral DNA sequences in the genomes of adenovirus type 12-transformed hamster cells. *Cell* 14, 569–585 (1978).
- 2 Doerfler W, Gahlmann R, Stabel S *et al.* On the mechanism of recombination between adenoviral and cellular DNAs: the structure of junction sites. *Curr. Top. Microbiol. Immunol.* 109, 193–228 (1983).
- 3 Hochstein N, Muiznieks I, Mangel L, Brondke H, Doerfler W. The epigenetic status of an adenovirus transgenome upon long-term cultivation in hamster cells. *J. Virol.* 81, 5349–5361 (2007).
- 4 Doerfler W. DNA methylation and gene activity. *Ann. Rev. Biochem.* 52, 93–124 (1983).
- First demonstration of causal relationship between promoter methylation and long-term gene inactivation.
- 5 Heller H, Kämmer C, Wilgenbus P, Doerfler W. Chromosomal insertion of foreign (adenovirus type 12, plasmid, or bacteriophage lambda) DNA is associated with enhanced methylation of cellular DNA segments. *Proc. Natl Acad. Sci. USA* 92, 5515–5519 (1995).
- The insertion of the adenovirus type 12 genome into the hamster cell genome leads to genome-wide alterations in cellular CpG DNA methylation patterns, both in some of the endogenous retrotransposon sequences and in cellular genes. These changes persist even after loss of the transgenomes.
- 6 Remus R, Kämmer C, Heller H, Schmitz B, Schell G, Doerfler W. Insertion of foreign DNA into an established mammalian genome can alter the methylation of cellular DNA sequences. *J. Virol.* 73, 1010–1022 (1999).
- 7 Müller K, Heller H, Doerfler W. Foreign DNA integration. Genome-wide perturbations of methylation and transcription in the recipient genomes. *J. Biol. Chem.* 276, 14271–14278 (2001).
- 8 Doerfler W. A new concept in (adenoviral) oncogenesis: integration of foreign DNA and its consequences. *BBA Rev. Cancer* 1288, F79–F99 (1996).
- 9 Doerfler W. Epigenetic consequences of foreign DNA integration: Global alterations of methylation and transcription patterns in recipient genomes. *Rev. Med. Virol.* 21, 336–346 (2011).
- 10 Doerfler W. The impact of foreign DNA integration on tumor biology and evolution via epigenetic alterations. *Epigenomics* 4, 41–49 (2012).
- Describes the possible epigenetic consequences of foreign DNA insertions into mammalian genomes for oncogenesis and in evolution.

- 11 Grafodatskaya D, Choufani S, Ferreira JC *et al.* EBV transformation and cell culturing destabilizes DNA methylation in human lymphoblastoid cell lines. *Genomics* 95, 73–83 (2010).
- 12 Birdwell CE, Queen KI, Kilgore PC *et al.* Genome-wide DNA methylation as an epigenetic consequence of Epstein–Barr virus infection of immortalized keratinocytes. *J. Virol.* 88, 11442–11458 (2014).
- 13 Soto-Girón MJ, García-Vallejo F. Changes in the topology of gene expression networks by human immunodeficiency virus type 1 (HIV-1) integration in macrophages. *Virus Res.* 163, 91–97 (2012).
- 14 Doi A, Park IH, Wen B *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 41, 1350–1353 (2009).
- 15 Naumann A, Hochstein N, Weber S, Fanning E, Doerfler W. A distinct DNA methylation boundary in the 5'-upstream sequence of the FMR1 promoter binds nuclear proteins and is lost in fragile X syndrome. *Am. J. Hum. Genet.* 85, 606–616 (2009).
- **A distinct boundary between an unmethylated and a hypermethylated region demarcates the promoter region from its upstream environment and protects it from *de novo* methylation.**
- 16 Naumann A, Kraus C, Hoogeveen A, Ramirez CM, Doerfler W. Stable DNA methylation boundaries and expanded trinucleotide repeats: Role of DNA insertions. *J. Mol. Biol.* 426, 2554–2566 (2014).
- **In transgenomic cells or in cells carrying episomal Epstein–Barr virus genomes, the usually hypermethylated region far upstream of the methylation boundary loses CpG methylation.**
- 17 Brattain MG, Fine WD, Khaled FM, Thompson J, Brattain DE. Heterogeneity of malignant cells from a human colonic carcinoma. *Cancer Res.* 41, 1751–1756 (1981).
- 18 Sambrook J, Russell DW. *Molecular Cloning, 3rd Edition*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, NY, USA (2001).
- 19 Minami M, Poussin K, Bre'Chot C, Paterlini P. A novel PCR technique using Alu-specific primers to identify unknown flanking sequences from the human genome. *Genomics* 29, 403–408 (1995).
- 20 Irizarry RA, Hobbs B, Collin F *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264 (2003).
- 21 Touleimat N, Tost J. Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* 4, 325–341 (2012).
- 22 Chen YA, Lemire M, Choufani *et al.* Discovery of cross-reactive probes and polymorphic CpG's in the Illumina Infinium Human Methylation450 microarray. *Epigenetics* 8, 203–209 (2013).
- 23 Benjamin Y, Hochberg Y. Controlling of false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 289, 289–300 (1995).
- 24 Verkerk AJMH, Pieretti M, Sutcliffe SS *et al.* Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65, 905–914 (1991).
- 25 McMahon M, Contreras A, Ruggero D. Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. *Wiley Interdiscip. Rev. RNA* 6, 173–189 (2015).
- 26 Jähner D, Jaenisch R. Retrovirus-induced *de novo* methylation of flanking host sequences correlates with gene inactivation. *Nature* 315, 594–597 (1985).
- 27 Lichtenberg U, Zock C, Doerfler W. Integration of foreign DNA into mammalian genome can be associated with hypo-methylation at site of insertion. *Virus Res.* 11, 335–342 (1988).
- 28 Akhtar W, de Jong J, Pindyurin AV *et al.* Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* 154, 914–927 (2013).
- 29 The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).
- 30 FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. *Nature* 507, 462–470 (2014).