


Article

A Global Gene Body Methylation Measure Correlates Independently with Overall Survival in Solid Cancer Types

Dietmar Pils ^{1,*} , Elisabeth Steindl ^{1,†}, Anna Bachmayr-Heyda ^{2,†,‡}, Sabine Dekan ³ and Stefanie Aust ²

¹ Division of General Surgery, Department of Surgery, Comprehensive Cancer Center (CCC) Vienna, Medical University of Vienna, 1090 Vienna, Austria; Elisabeth.Steindl@meduniwien.ac.at

² Department of Obstetrics and Gynecology, Comprehensive Cancer Center (CCC) Vienna, Medical University of Vienna, 1090 Vienna, Austria; Anna.Bachmayr@hotmail.com (A.B.-H.); Stefanie.Aust@meduniwien.ac.at (S.A.)

³ Department of Pathology, Medical University of Vienna, 1090 Vienna, Austria; Sabine.Dekan@meduniwien.ac.at

* Correspondence: dietmar.pils@univie.ac.at; Tel.: +43-1-40400-41690; Fax: +43-1-40400-66740

† These authors contributed equally.

‡ Current address: Research Cancer Immunology, Boehringer Ingelheim RCV GmbH & Co KG, Boehringer Gasse 5–11, 1120 Vienna, Austria.

Received: 30 June 2020; Accepted: 8 August 2020; Published: 12 August 2020



Abstract: Epigenetics, CpG methylation of CpG islands (CGI) and gene bodies (GBs), plays an important role in gene regulation and cancer biology, the former established as a transcription regulator. Genome wide CpG methylation, summarized over GBs and CGIs, was analyzed for impact on overall survival (OS) in cancer. The averaged GB and CGI methylation status of each gene was categorized into methylated and unmethylated (defined) or undefined. Differentially methylated GBs and genes associated with their GB methylation status were compared to the corresponding CGI methylation states and biologically annotated. No relevant correlations of GB and CGI methylation or GB methylation and gene expression were observed. Summarized GB methylation showed impact on OS in ovarian, breast, colorectal, and pancreatic cancer, and glioblastoma, but not in lung cancer. In ovarian, breast, and colorectal cancer more defined GBs correlated with unfavorable OS, in pancreatic cancer with favorable OS and in glioblastoma more methylated GBs correlated with unfavorable OS. The GB methylation of genes were similar over different samples and even over cancer types; nevertheless, the clustering of different cancers was possible. Gene expression differences associated with summarized GB methylation were cancer specific. A genome-wide dysregulation of gene-body methylation showed impact on the outcome in different cancers.

Keywords: gene body; DNA methylation; CGI; overall survival; epigenetics; CpG island

1. Introduction

Epigenetic changes are an increasingly important research field for understanding cancer biology and cancer treatment [1], despite the fact that epigenetics has not yet reached the status of a cancer hallmark [2,3]. DNA methylation on cytosines is one of the epigenetic changes which can be measured relatively easily, even genome wide, and whose impact on gene expression is—more or less—well known. In vertebrates usually most of cytosines in CpG-dinucleotides are methylated throughout the genome [4]. Only promoter regions, with or without CpG-islands (CGIs), of expressed genes are frequently hypo-methylated. One of the epigenetic changes associated with cancer (progression) is

the hyper-methylation of such CGIs in promoter regions of so-called tumor suppressor genes (TSG), leading to the silencing of the corresponding genes, with the well-known examples TP53 and BRCA1. Gene body CpG-methylation is not so well understood, but usually associated with higher expression of the corresponding gene, or vice versa, hypo-methylation of gene bodies (GBs) associated with lower expression [5]. Positive correlations between active transcription and gene body methylation have recently been confirmed on the active X chromosome [6,7]. It was also shown that intragenic DNA methylation, especially if located near exon–intron boundaries, is associated with alternative splicing events [8]. Archederra et al. discovered, in a clinically relevant hepatocellular carcinoma (HCC) mouse model, that the expression of oncogenes with hypermethylated CGIs either in their 5'-UTR or in the gene body is raised [9]. Many of these genes discovered in the mouse model—several of them well-known oncogenes—are also affected by concurrent CGI hypermethylation and gene expression upregulation in human HCC patients, especially in 56% of HCC patients, which belong to the “HCC proliferative-progenitor” subclass.

However, overall, “Intriguingly, the genes regulated by intragenic methylation in cancer cells are related to cell type-specific functions rather than tumor suppressors”, a citation from Lee et al. [8].

It was also recognized, that global hypomethylation of cancer genomes (often measured on Alu repeats [10]) is a frequent event and associated with a more aggressive tumor [11]. Here, we analyzed global CGI methylation, as well as global gene body methylation levels, in different solid cancer entities and its effect on overall survival.

2. Results

2.1. Reduced Representation Bisulfite Sequencing

DNA methylation of fresh frozen (FF) and formalin fixed paraffin embedded (FFPE) ovarian cancer tissues was measured with the enhanced reduced representation bisulfite sequencing (RRBS) methodology [12]. CpG call rates (more than five informative reads) of median 1.86 million (0.57–2.80 m) CpG-cytosines for FFPE and 2.58 million (1.53–3.44 m) for FF material were obtained. CGI coordinates were downloaded from the UCSC genome browser (HG38) and GB coordinates of protein coding genes extracted from the GENCODE V25 annotation. Beta values were calculated as ratios of methylated cytosines to total called cytosines (0 being unmethylated and 1 fully methylated) and the weighted ($\log_{10}(\text{calling depth} + 0.1)$) average of all CpG cytosines calculated for each CGI and every GB. Only averaged beta values of CGIs or GBs with at least eleven called CpGs were used for further analyses. These averaged beta values for each CGI and every gene body and sample were categorized into ten deciles, i.e., 0–10%, 10–20%, . . . , 90–100% and summed up.

In Figure S1 the correlations of averaged GB methylation values of matched pairs of tumor tissues analyzed from FF and FFPE material are shown, indicating a high consistency of results. In addition, correlations of all possible combinations within the group of FF and FFPE tissues and between FF and FFPE tissues of different patients were obtained (Figure S2A). Interestingly, there was a very high correlation between all samples, indicating a very stable GB methylation situation in ovarian cancer.

For survival analysis only values from FFPE tissues were selected as this cohort was the larger one and tissues used for DNA isolation were enriched for tumor cells using macro- and laser capture microdissection. Using the ten frequency deciles for each patient a robust survival analysis was performed with censored overall survival data as an outcome variable. In Figure 1 this procedure is outlined.

Four GB methylation deciles showed significant correlation with overall survival in ovarian cancer (red and green dots in Figure 1D), i.e., defined/extreme states (0–10%, means not methylated and 90–100%, means completely methylated) showed a negative impact on survival, whereas undefined/intermediate states (50–60% and 60–70%) showed a positive impact on survival. Using these slots a final “methylation definition factor” (MDF) was calculated as given in Figure 1E, i.e., the number of GBs with a defined/extreme methylation status (<10% and >90%) divided by the

number of GBs with undefined/intermediate methylation states (50–70%), dichotomized at the optimal cutoff and used for multiple Cox regression analysis including known clinicopathologic factors such as age, FIGO stage, and residual tumor after debulking surgery (Table 1). A higher MDF, corresponding to more defined/extreme methylated genes compared to undefined/intermediate methylated genes, indicated unfavorable OS. A similar analysis with CGI methylation frequency deciles revealed no significant impact on survival at all.

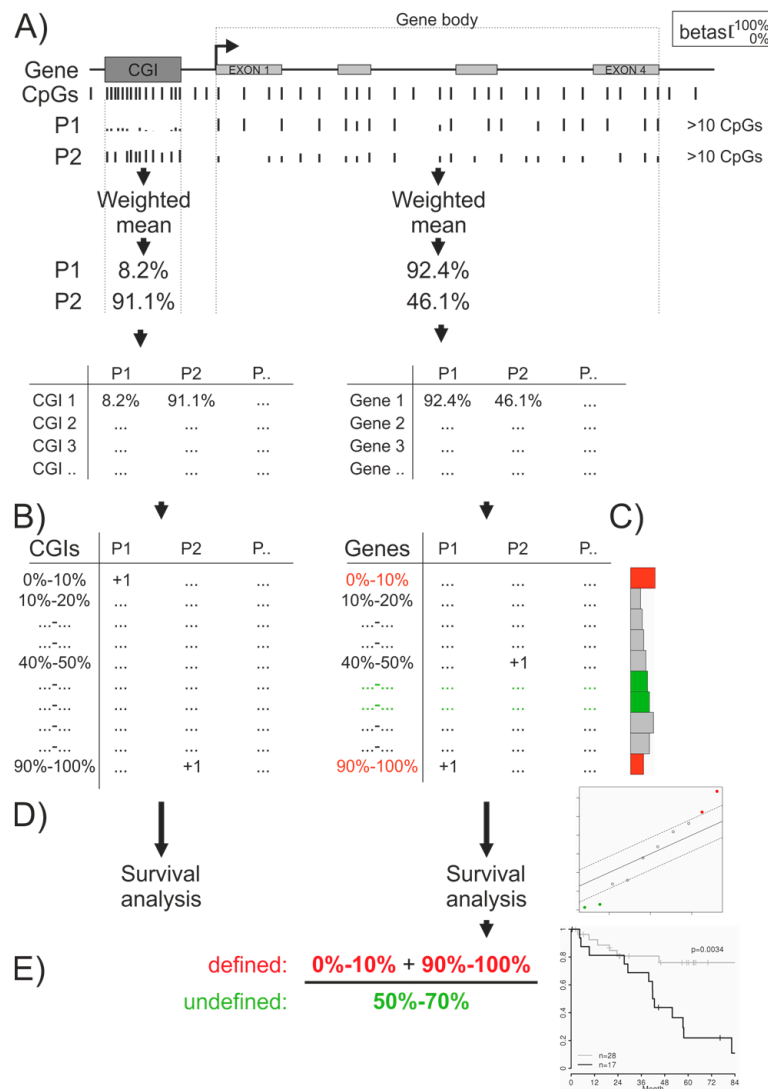


Figure 1. Overview of the generation of the global CpG island (CGI) and gene body (GB) methylation measures used for correlation with overall survival (OS). (A) Beta values from CpGs mapped to CGIs or GBs were averaged—from RRBS data with a weighted method—and (B) summarized in decile slots, (C) shown as a histogram exemplarily for the GBs of one patient. (D) Using these summarized decile numbers robust Cox regression analyses with censored OS data were performed (plot on the right showing the expected and observed scores for the ten deciles and the 5% false discovery rate (FDR) cutoff as dashed line; in red and green the significant deciles). (E) Using all negatively (red dots in the plot of (D)) and positively (green dots in the plot of (D)) with OS correlated deciles a methylation definition factor (MDF) was calculated as indicated and used for Cox regression analysis including known cancer type specific clinicopathologic parameters. An optimal cutoff for the MDF was determined from this model and a Kaplan–Meier estimate using the optimally dichotomized MDF is shown on the right. A plot of the Cox model, corrected for clinicopathologic factors, is shown in Figure 2 and details of the Cox regression models are given in Table 1.

Table 1. Overview of Characteristic Numbers (Samples, Death Event Numbers, Significant Deciles, Dichotomized Numbers, Hazard Ratios, and *p*-Values of Single and Multiple Cox Regression Analyses and Included Clinicopathologic Factors) of Cancer Types Analyzed in This Study (the grey shadow indicates subanalyses for ovarian cancer and is otherwise alternating for better readability).

Cancer Type	Characteristics and OS Analyses				Cox Regression (Overall Survival)			Clinical Params ⁶	Cit
	Patients ¹	Gene Bodies ²	SAM Sign. ³	MDF Predictor (Def/UnDef) ⁴	Cutpoint ⁵	Univariate	Multiple		
Ovarian (OvCa)	45 (19)	17,798	+ 0–10% + 90–100% –50–60% –60–70%	D: <10% and >90% U: 50–70%	28 vs. 17	HR 4.29 (1.62–11.3) <i>p</i> = 0.0034	HR 4.44 (1.63–12.08) <i>p</i> = 0.0035	Age Residual tumor (FIGO stage) (Grade)	RRBS
Ovarian—Exons	45 (19)	14,902 (Exons)	+ 0–10% + 90–100% –all other deciles	D: <10% and >90% U: 30–70%	27 vs. 18	HR 3.17 (1.24–8.11) <i>p</i> = 0.0162	HR 3.44 (1.31–9.06) <i>p</i> = 0.0124	Age Residual tumor (FIGO stage) (Grade)	RRBS
Ovarian—Intror	45 (19)	15,223 (Introns)	+ 0–10% + 80–90% + 90–100% –60–70% –70–80%	D: <10% and >80% U: 60–80%	28 vs. 17	HR 5.33 (1.96–14.52) <i>p</i> = 0.0011	HR 6.10 (2.01–18.51) <i>p</i> = 0.0014	Age Residual tumor (FIGO stage) (Grade)	RRBS
Breast (BC)	690 (87) 680 (84)	7814	+ 10–20%	D: <20% U: 30–70%	428 vs. 252	HR 1.79 (1.17–2.76) <i>p</i> = 0.0078	HR 1.78 (1.16–2.75) <i>p</i> = 0.0088	Stage (Histology)	[13]
Colorectal (CRC)	390 (87) 370 (80)	7816	+ 70–80% + 80–90% –20–30% –30–40% –40–50% –50–60% –0–10%	D: >70% U: 20–60%	192 vs. 178	HR 1.8 (1.13–2.85) <i>p</i> = 0.0126	HR 1.96 (1.23–3.11) <i>p</i> = 0.0044	Stage (Sex) (Histology) (Site)	[14]
Pancreatic (PAAC)	184 (99) 167 (91)	7816	–10–20% –20–30% + 40–50% + 50–60%	D: <30% U: 40–60%	114 vs. 53	HR 0.44 (0.27–0.74) <i>p</i> = 0.0018	HR 0.49 (0.29–0.82) <i>p</i> = 0.0068	Age Residual tumor Radio therapy (Sex) (Stage)	[15]
Lung	238 (150)	7815	n.s.	n.d. ⁷	n.d.	n.d.	n.d.	n.d.	[16]
Glioblastoma (GBM)	138 (93) 126 (82)	7816	+ 0–10% + 10–20% + 20–30% + 30–40% + 40–50% –60–70% –70–80% –80–90% –90–100%	Methylated/Unmethylated (MUMF) M: >50% UM: <50%	76 vs. 50	HR 0.49 (0.31–0.79) <i>p</i> = 0.0030	HR 0.50 (0.30–0.83) <i>p</i> = 0.0073	Sex Histology Radio therapy	[17]

¹ Number of patients with survival data (death events) (Number of patients with complete clinical data (Death events) used for multiple Cox regression, if not all). ² Number of gene bodies with >10 CpG beta values. ³ Significant frequency deciles with FDR < 5%. “+” means positive impact on unfavorable OS and “–” vice versa. n.s., not significant. ⁴ Definition of the methylation definition factor (MDF) and the methylation-over-unmethylation factor (MUMF) as predictor for overall survival. ⁵ Dichotomization using the optimal cut-point of the MDF/MUMF predictor determined with multiple Cox regression models including all clinical parameters (high MDF/MUMF vs. low MDF/MUMF). ⁶ Clinical parameters finally included into the multiple Cox regression model together with the dichotomized MDF predictor, determined by minimizing the Akaike Information Criterion, AIC (in parentheses, additional clinical parameters considered for the Cox model which were excluded). ⁷ n.d., not determined.

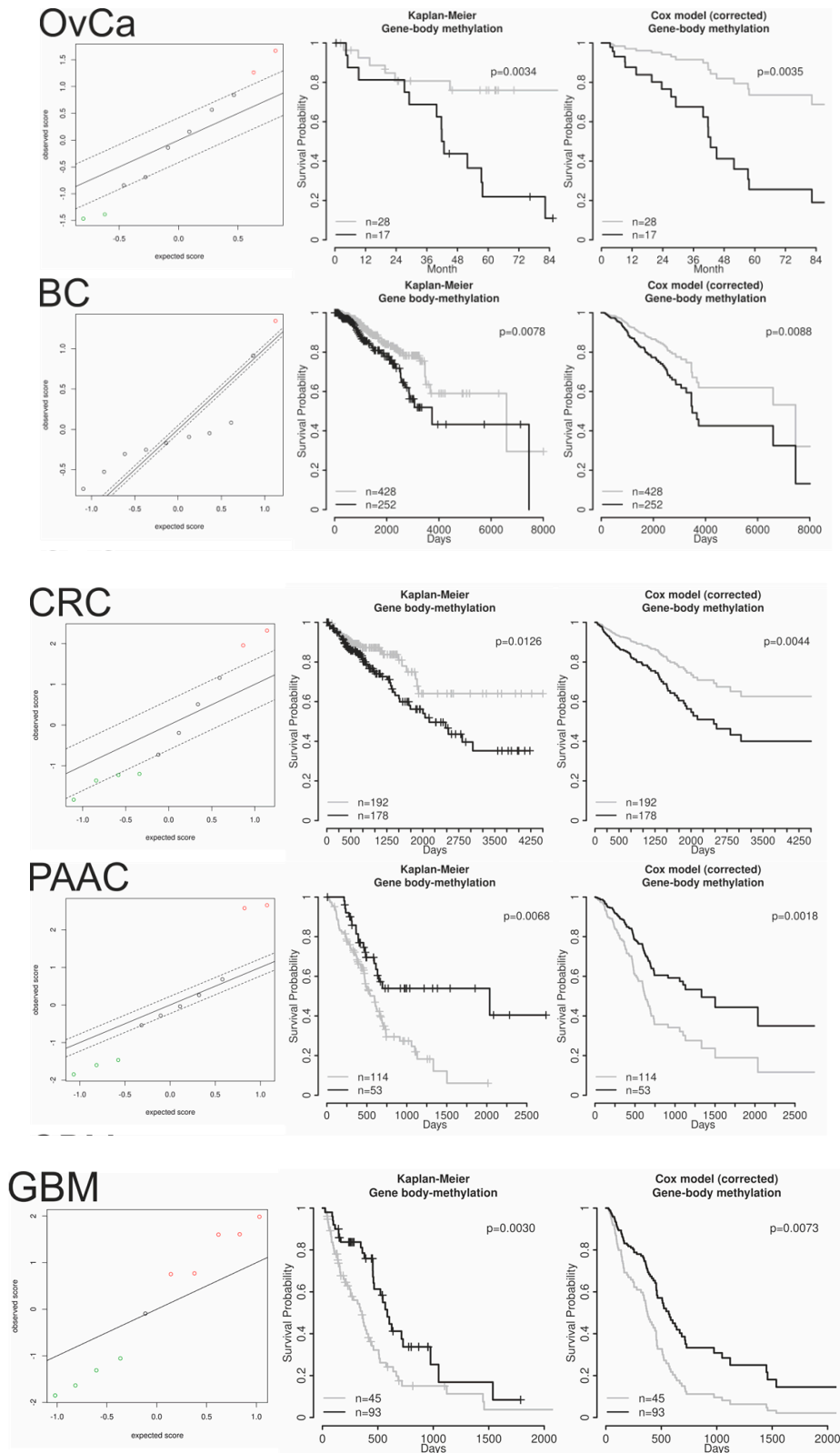


Figure 2. Significance Analysis of Microarray (SAM) plots of Cox regressions (X-axis, expected score; Y-axis, observed score; dashed line, threshold cutoff at 5% false discovery rate, FDR; red dots, associated with unfavorable overall survival (OS); green dots, associated with favorable OS), Kaplan–Meier estimates of the optimally dichotomized MDF/MUMF, and survival curves of the corrected Cox regression models (details are given in Table 1). (No censored patients are indicated for the latter plots, as these are survival estimates from multiple Cox regression models).

A comparison of GB methylation and corresponding CGI methylation (mostly localized in the 5'-untranslated region (5'-UTR) or at the beginning of the first exon) over all samples revealed high ($r > 0.8$) correlations only for 4.45% of GB-CGI combinations, mainly small genes with the CGI inside their corresponding GB. In Figure S3 these results are shown, in (A) the correlations between GB and CGI methylation values over all samples subdivided into ten methylation slots (GBs left plots and CGIs right plots), i.e., 0–10%, 10–20%, ..., 90–100% mean GB or CGI methylation, respectively (colored white to dark grey), in (B) the correlation of corresponding GB and CGI methylation levels averaged over all samples of all GB-CGI combinations, and in (C) the coefficients of variation (CV) of the GB methylation values (in percent) according the GB mean values, indicating more variation in low methylated GBs compared to high methylated ones (probably due to noise derived from the method used for measurement). Interestingly, the highest GB-CGI methylation correlations were observed in low methylated GBs (which showed the highest variation) and medium to high methylated CGIs.

The next question was whether the GB methylation level correlated to the RNA expression of the corresponding genes or whether genes with high GB methylation levels are globally higher expressed compared to genes with low GB methylation levels (both suggested from literature). Figure S4 shows in (A) the histograms of the correlation coefficients between GB methylation values and corresponding gene expression values of all analyzed patients subdivided into ten GB methylation slots, i.e., 0–10%, 10–20%, ..., 90–100% mean GB methylation levels (colored white to dark grey) and in (B) the boxplots of \log_2 expression values of all genes in these GB methylation slots and boxplots of the corresponding gene (i.e., mRNA transcript) lengths of genes in these slots. The latter is shown, as from RNA sequencing data the expression values derived from mapped read counts are dependent on corresponding (mappable) gene lengths, if not corrected for gene lengths during bioinformatical processing. There seems to be no positive correlation between GB methylation and gene expression, neither for individual genes nor globally over all genes with the same GB methylation level. Contrary, genes with very high GB methylation levels are globally less expressed compared to genes with lower or intermediate GB methylation levels (regardless of gene length).

2.2. Comparison of the Methylation Definition Factor (MDF) of Tumor Tissues, Tumor Cell Lines, and Normal Human Tissues

To compare the ratios of the defined/extreme methylation status (averaged beta values of CpGs in GBs < 0.1 , i.e., unmethylated, and averaged beta values of CpGs in GBs > 0.9 , i.e., methylated) to the undefined/intermediate methylation status (averaged beta values between 0.2 and 0.8), the methylation definition factor (MDF), between tumor tissues, tumor cell lines, and normal cells and tissues, RRBS data from the ENCODE consortium were downloaded and processed as above (using corresponding HG19 annotations for CGIs and GBs). In Figure S5 histograms of these MDF values over all samples in each category are shown. Normal cells and tissues showed the most defined MDF values around 2 and up to 4 (with one outlier at 8, perhaps an artefact), followed by cancer cell lines with values around 1 and up to 5.5 and tumor tissues with values all below 1, regardless if measured with DNA isolated from fresh frozen or FFPE tissues. Loss of defined GB methylation, either un- or completely methylated, seems to be a universal event in cancer tissues. Comparing high and low MDF values from tumor tissues concerning overall survival revealed more aggressive tumors with high MDF values, leading to unfavorable overall survival in ovarian cancer patients. In Figure S6 the distribution of all averaged GB methylation beta values with 95%-confidence intervals over all samples are shown for ovarian cancer tissues, cancer cell lines, and normal human primary cells or tissues.

2.3. Validation of the Impact of the MDF on Survival in other Cancer Entities

Unfortunately, the TCGA methylation data for ovarian cancer were performed with Agilent's Human Methylation 27 bead array, comprised of about 27.6K CpGs only, which is by far not enough to validate the results (not even one CpG for each of the ~28K CGIs and ~18K protein coding GBs). Other publicly available DNA methylation data of clinical samples with corresponding survival data

(RRBS, whole genome bisulfite sequencing (WGBS), or Agilent's Human Methylation 450K array) were not available for ovarian cancer; therefore, we thought to investigate if similar correlations of MDF values with survival will be obtained for other solid cancer entities. In Table 1 cancer entities used in this study are summarized. Interestingly, in breast (BC) and colorectal (CRC) cancer comparable results were obtained using GB frequency deciles for predicting overall survival: more defined/extreme states (either highly methylated or unmethylated) showed a negative impact on survival and more undefined/intermediate states (averaged beta values around 0.5) showed a positive impact. A calculated and optimally dichotomized MDF revealed, always, an independent predictor for OS, corrected for relevant clinicopathologic factors, similar as for OvCa. In BC the MDF correlated positively with the basal subtype (Figure S7), but including the subtype information to the multiple Cox regression model even increased the hazard ratio (HR) and the significance of the dichotomized MDF (HR 2.00; $p = 0.003$). In CRC the MDF correlated positively with the CpG Island Methylator Phenotype high (CIMP-H) status (determined from TCGA colorectal cancer samples according to Hinoue et al. [18]; Figures S8 and S9) but the dichotomized MDF remained as independent predictor for OS (HR 2.07; $p = 0.004$) even when the CIMP status was added to the other clinicopathologic factors in the multiple Cox regression model. The CIMP status showed no independent impact on survival. In lung cancer, no significant association of any averaged GB methylation frequency decile with OS was revealed, therefore, no further survival analyses were performed. In pancreatic adenocarcinoma (PAAC) the impact of the averaged GB methylation frequency deciles was the other way round, more defined/extreme states (i.e., unmethylated) showed a positive impact on favorable OS and more undefined/intermediate states showed a positive impact on unfavorable OS. An optimally dichotomized high MDF, therefore, showed alone and corrected for known clinicopathologic factors (independent) significant impact of favorable OS. In glioblastoma multiforme (GBM), all deciles below averaged beta values of 0.5 (unmethylated) showed a negative impact on OS and all deciles above averaged beta values of 0.5 (methylated) revealed positive impact on OS. Therefore a "methylation-over-unmethylation factor" (MUMF) was calculated, dichotomized at the optimal cutoff, and used for OS analyses. High MUMF correlated with favorable OS, alone and corrected for (thus independent from) clinicopathologic factors. In Figure 2 SAM plots of robust Cox regressions, Kaplan–Meier estimates, and multiple Cox regression survival curves, dichotomized using the optimal cutoff, are shown.

2.4. Characterization of Unmethylated, Methylated, and Undefined Methylated Gene Bodies

To assess the methodological impact of the RRBS technology and the Agilent's methylation arrays to the results and to estimate the cancer type specificity of the whole genome GB methylation state, an Isomap (a nonlinear dimensionality reduction method) and a tSNE (t-distributed stochastic neighbor embedding, a machine learning algorithm for visualization) analysis over all tumor entities using averaged beta values was performed. Figure S10 shows the first four dimensions of the Isomap in three plots (dimensions 1 and 2, 2 and 3, and 3 and 4) and in Figures 3 and 4 the first two dimensions of the tSNE are shown. The largest impact on the results revealed the methodology, i.e., RRBS versus methylation array, in dimension 1 of the Isomap. The clearest discrimination between cancer entities are seen in dimensions 3 and 4. In the tSNE representation all cancer entities are nearly completely separated, with OvCa samples isolated on the upper side due to methodological and biological differences. Only a few individual cancer samples are wrongly clustered, especially some lung and CRC samples in the PAAC cluster. Figure 4 shows the tSNE plots colored according the MDF/MUMF values, the optimally dichotomized MDF/MUMF, and each one of the relevant clinicopathologic or subclassification factors. To compare cancer entity specific lists of usually extreme methylated (mean averaged beta values > 0.7), usually undefined/intermediate methylated (mean averaged beta values > 0.3 and < 0.7) or usually extreme unmethylated (mean averaged beta values < 0.3) genes, Venn-like diagrams (UpSet plots) are shown in Figure S11. Interestingly, besides the differences between RRBS data (OvCa) and all methylation array data (all other cancer entities, including GBM),

the overlap over all cancer entities was relatively large, indicating a robust methylation status of GBs, at least in solid cancers.

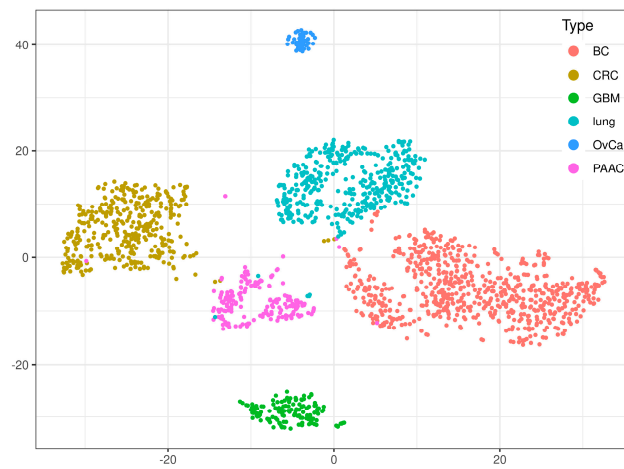


Figure 3. T-distributed stochastic neighbor embedding (tSNE) of the GB-averaged beta values of all cancer entities ($n = 7.403$, measured from all cancer types).

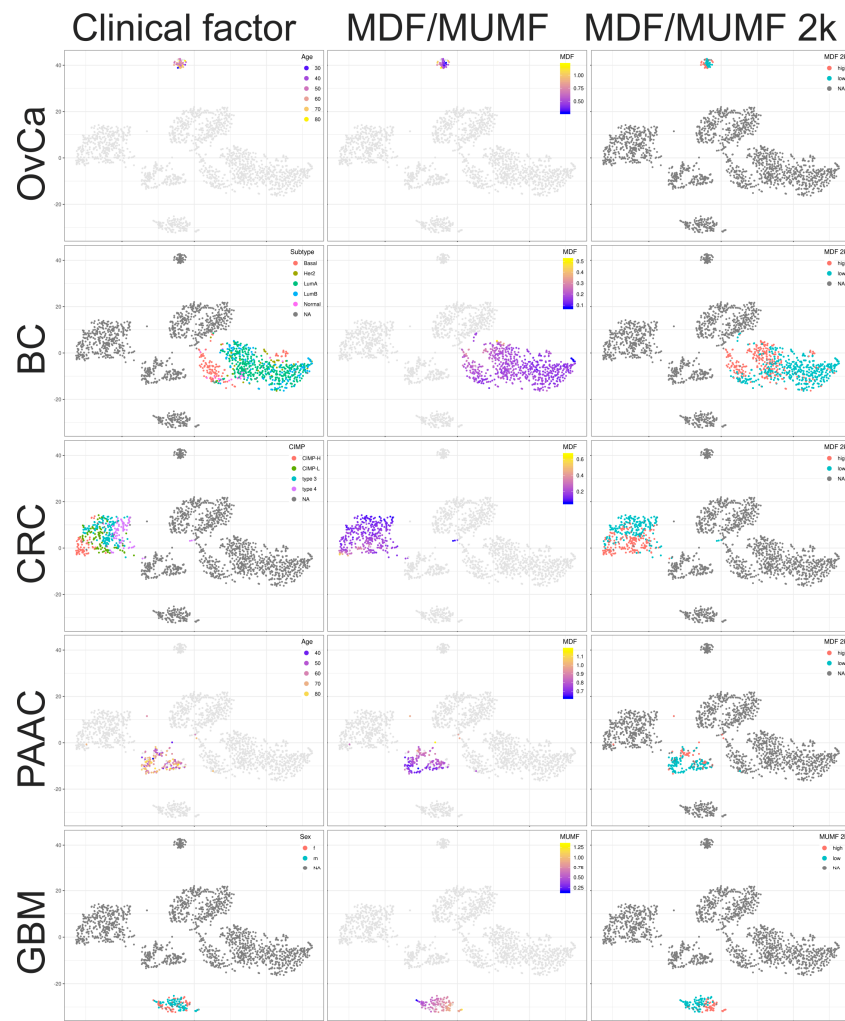


Figure 4. T-distributed stochastic neighbor embedding (tSNE) of the GB-averaged beta values as in Figure 3, colored according one relevant clinicopathologic factor (age for OvCa and PAAC, the subtype for BC, the CIMP status for CRC, and sex for GBM) in the left panel, colored by the MDF/MUMF values in the middle panel, and colored according the optimally dichotomized MDF/MUMF in the right panel.

To biologically annotate genes with either preferentially unmethylated (mean averaged beta values < 0.3), methylated (mean averaged beta values > 0.7), or undefined methylated (mean averaged beta values between 0.3 and 0.7) GBs, lists over all samples per cancer entity where built averaging all averaged beta values per single gene and trichotomized according mean beta cutoffs of 0.3 and 0.7. In Figure 5 gene ontology (GO) and pathway annotations for each category (methylated > 0.7, undefined between 0.3 and 0.7, and unmethylated < 0.3) of every cancer type with GB methylation impacts on OS are shown as word clouds in colors indicating enrichment *p*-values.

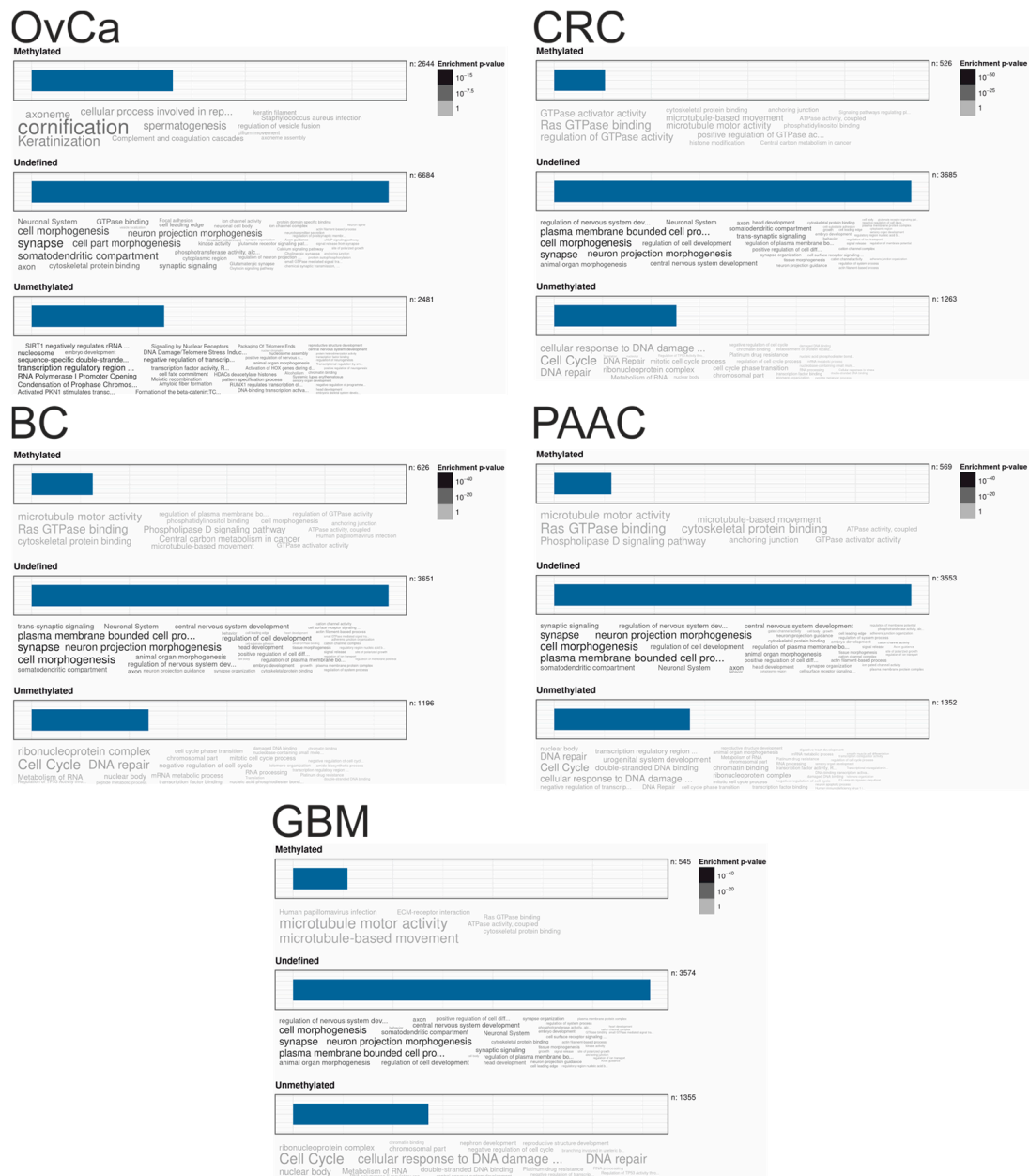


Figure 5. Enrichment plots of gene ontology terms and pathways are shown for lists of genes preferentially (averaged over all samples from each cancer type) methylated (averaged betas > 0.7), unmethylated (averaged betas < 0.3) or undefined methylated (averaged betas between 0.3 and 0.7). Enrichments are indicated in word clouds by size and color (*p*-values). Overlap of gene lists are shown in Figure S11.

To assess the consistency of the GBs in the mean GB methylation slots, i.e., high methylated (>0.7), intermediate methylated (≥0.3 and ≤0.7), and low methylated (<0.3), Figure S12 shows the distribution

of the percentages of all genes consistently annotated to one of these three slots. Interestingly, high percentages of GBs in the same slot are only seen for intermediate methylated GBs in all cancer entities. Therefore, there are only few genes with consistent high or low GB methylation values if cutoffs are applied, despite the fact that GB methylation values correlated substantially over all samples (cf. Figure S2). To analyze and annotate the overlap of consistently high, intermediate, or low methylated GBs, Figures S13 and S14 show the overlap of gene (i.e., GB) lists which are in >90% of samples annotated to the respective GB methylation slot, thus consistently categorized, and the annotations of these gene lists in GO enrichment plots, respectively.

2.5. Correlation of the MDF and MUMF with Gene Expression

Genes whose expressions were significantly correlated to the MDF in OvCa, BC, CRC, and PAAC and to the MUMF in GB were determined from RNA-sequencing data, either from our own data, OvCa [19,20], or from the TCGA database for all other cancer entities. In OvCa (false discovery rate (FDR) < 10%) 1039 genes were positively correlated with the MDF and 730 were negatively correlated; in BC (FDR < 5%) 3177 and 2920 genes, in CRC (<5%) 1405 and 1179 genes, and in PAAC (<5%) 1780 and 2052 genes, respectively. In GB (FDR < 10%) 339 genes were positively correlated with the MUMF and 505 negatively. Overlaps of significantly differentially correlated genes are shown in Figure S15, indicating nearly no relevant overlaps between tumor entities. There were completely no overlaps of significantly positively or negatively correlated genes over all samples, even if GBM was excluded. Only two genes were positively correlated in OvCa, BC, CRC, and negatively correlated in PAAC, namely SLC7A11 and TFAP2A, and only one gene was negatively correlated in OvCa, BC, and CRC and positively correlated in PAAC, namely LIMS2. The rationale for this combination was that the MDF showed an inverse impact on OS in PAAC compared to OvCa, BC, and CRC.

A Signaling Pathway Impact Analysis (SPIA) with lists of significantly with the MDF/MUMF factor associated genes revealed for three cancer entities significant (FDR < 5%) deregulated KEGG pathways (total 202 KEGG pathways considered, Figure S16), for OvCa two activated ones: "Systemic lupus erythematosus" (FDR 0.08%, Figure S17A) and "Staphylococcus aureus infection" (FDR 0.5%), for BC 30 inhibited pathways: "Cytokine-cytokine receptor interaction" (Figure S17B), "Neuroactive ligand-receptor interaction", "Chemokine signaling pathway", "Natural killer cell mediated cytotoxicity", "PI3K-Akt signaling pathway", "Systemic lupus erythematosus", "Pathways in cancer", "Staphylococcus aureus infection", "Th1 and Th2 cell differentiation", "Th17 cell differentiation", "Focal adhesion", "Measles", "NF-kappa B signaling pathway", "Human T-cell leukemia virus 1 infection", "Necroptosis", "Inflammatory bowel disease (IBD)", "Complement and coagulation cascades", "Leukocyte transendothelial migration", "Hepatitis B", "Adipocytokine signaling pathway", "Autoimmune thyroid disease", "Allograft rejection", "Malaria", "Asthma", "Intestinal immune network for IgA production", "Rap1 signaling pathway", "Osteoclast differentiation", "Antigen processing and presentation", "ECM-receptor interaction", FDRs from 10⁻⁸% to 4.9%, and for GB one inhibited "Taste transduction" and one activated "Choline metabolism in cancer" pathway (both FDR 2.5%, Figure S17C) (cf. Table S1). Interestingly, no pathway regulation direction combination overlapped between these cancer entities. For CRC and PAAC no significant KEGG pathways were revealed at all (SPIA results tables are given in Table S1).

For further biological annotations lists of significantly (false discovery rates <5% or <10% for OvCa and GB) positively and negatively with the MDF/MUMF correlated genes for each cancer entity were composed and annotated as above. In Figure 6 these are shown as word clouds in colors indicating enrichment *p*-values.

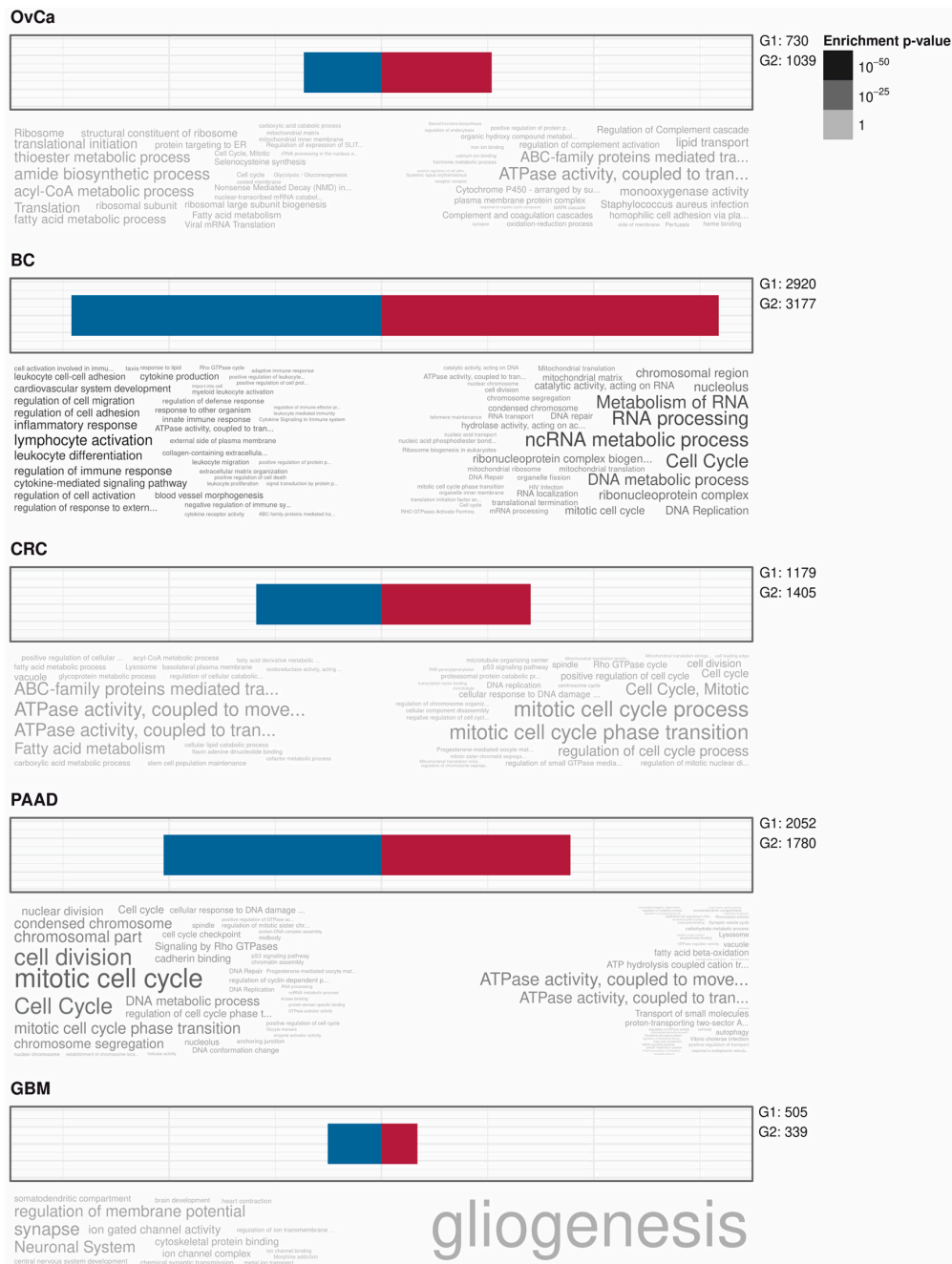


Figure 6. Enrichment plots of gene ontology terms and pathways of negatively (blue; G1) and positively (red; G2) with the MDF/MUMF correlated genes in different cancer entities are shown. For OvCa and GBM a 10% FDR cutoff, for all other cancer entities a 5% cutoff was used. Enrichments are indicated in word clouds by size and color (*p*-values). Overlaps of gene lists are shown in Figure S15.

3. Discussion

DNA methylation is an important event and factor in epigenetic gene expression regulation [21]. Globally, CpG dinucleotides are substantially under-represented in the human genome, except in a few thousand short regions (few hundred to thousand bases long) with enriched CpG-frequencies, known as CpG-islands (CGIs). Non-CGI CpGs are usually methylated in the human genome and CGI CpGs unmethylated in association with active genes. Inactive genes are characterized by methylated CGIs in promoter regions, if a CGI is present at all. Therefore, CGIs are usually involved in gene regulation by a regulator protein (e.g., transcription factors) binding, either at promoter regions or in enhancers, usually following the logic, higher methylation lower regulator protein binding and expression

and vice versa. Globally, solid cancers are characterized by global genome-wide hypo-methylation, i.e., usually methylated CpGs get less methylated in cancer tissues, and hyper-methylation of specific CGIs, especially in front of so-called tumor suppressor genes (TSGs). These TSGs are specifically silenced by hyper-methylation of corresponding CGIs to allow the tumor a more aggressive, malignant, or metastasizing phenotype.

GB methylation, i.e., the methylation status of CpGs throughout the GB, including introns and exons, is a more unclear factor. Usually it is considered that a higher GB methylation is correlated with higher gene expression [5,7], but the concrete mechanisms are not clear. Only the inhibition of spurious transcription initiation of GB methylation of high expressed genes was shown in mouse stem cells [22]. GB methylations at specific regions are also correlated to splicing [8,23].

Using a global approach of correlating GB methylation levels with expression of corresponding genes revealed no proof of a positive correlation. Only genes with high GB methylation levels showed a slightly positive median correlation coefficient, but genes with low and intermediate GB methylation levels showed even a slightly negative median correlation coefficient (Figure S4A).

This work is to our knowledge the first report of the analysis of a global GB methylation status with impact on overall survival in several solid cancer entities. Using averaged methylation beta values of GBs allows near perfect classification of the cancer entities analyzed in this work (Figure 3 and Figure S10), even as GB methylation values were highly correlated over samples and types of tissues, fresh frozen or FFPE (Figure S2), and lists of differentially methylated GBs were rather similar over cancer entities (Figure S11), indicating a similar GB methylation status of individual genes over tissues. There were only few genes where GB methylation correlated to the corresponding CGI methylation (Figure S3) or to gene expression (Figure S4). Only genes with highly methylated GBs were less expressed as a group compared to genes with lowly or intermediately methylated GBs, even considering gene length as possible bias for expression analysis from RNA sequencing data. Nevertheless, a positive impact of a globally more defined/extreme state of GB methylation (i.e., more GBs either highly or lowly methylated compared to GBs with an in-between methylation status), represented by a methylation definition factor (MDF) on unfavorable OS, was initially shown with our own RRBS data in OvCa (Figure 1) and validated with TCGA data in breast (BC) and colorectal (CRC) cancer (Table 1 and Figure 2). Even the exclusive analysis of intronic and exonic regions of the GBs revealed similar results in OvCa compared to the analysis of the complete GB, i.e., a significant impact on OS (cf. Table 1). Interestingly, in pancreatic adenocarcinoma (PAAC), the impact on OS was the other way round, i.e., more defined/extreme methylated GBs correlated with favorable OS (Figure 2). All OS impact results were significant and independent, i.e., even if corrected for cancer type specific clinicopathologic factors and for subtype in BC and for the CpG Island Methylator Phenotype (CIMP) in CRC (even as the MDF was positively correlated to the basal subtype in BC and the CIMP-high status in CRC). For lung cancer no correlation of GB methylation with OS was found (even with a large patient cohort) and for glioblastoma the number of methylated GBs (averaged beta values > 0.5) compared to the number of unmethylated GBs (averaged beta values < 0.5) correlated with favorable OS—methylation-over-unmethylation factor (MUMF)—alone and corrected for clinicopathologic factors. Glioblastoma is not a carcinoma (i.e., derived from epithelial cells) as the other analyzed cancer entities above, but derived from astrocytes, with a rather different (cancer) biology. Therefore, the different impact of GB methylation on OS is not surprising.

Our interpretation of the absence of significant correlations of single GB methylation levels with OS but a highly significant summarized GB methylation measure (i.e., the ratio of high/low methylated to intermediate methylated GBs, described here as MDF and MUMF) is that this measure indicates a malignant epigenetic phenotype dysregulating the whole transcriptional process in the cell but not tumor suppressor genes or oncogenes predominantly. However, this global dysregulation of expression is difficult to analyze with typical RNA-sequencing experiments, as the absolute expression levels of genes in a cell are completely unknown and even more unknowable, only relative differences of expressions between genes and samples can be determined. Probably the epigenetic state described

here also influences global splicing events [8], even splicing events leading to circular RNAs, known to play a role in cancer [24]. To analyze epigenetically driven splicing, not poly(A)-enriched and much deeper sequenced cohorts of tumors would be necessary.

The impact of the MDF/MUMF on gene expression is completely different for all cancer entities with insignificant overlap on gene level and no overlap on dysregulated pathway level at all. Only SLC7A11, TFAP2A, and LIMS2 are commonly regulated by the MDF in carcinomas (not in GBM), but only if the direction of the correlation was inversed for PAAC, whose impact on OS was also inverse. A pathway enrichment analysis also revealed very different results from all cancer entities, with either many significantly dysregulated pathways for BC, only a few significantly dysregulated pathways for OvCa and GMB, or even no significantly dysregulated pathways for CRC and PAAC. Therefore, we want cite again the publication from Lee et al. [8] directly, “Intriguingly, the genes regulated by intragenic methylation in cancer cells are related to cell type-specific functions rather than tumor suppressors” and want add “... or oncogenes”, which seems to be proven again with our results.

The upside down impact of GB methylation in PAAC compared to OvCa, BC, and CRC remains interesting, as does the complete independence of global GB methylation—as analyzed here—and OS in lung cancer, even with a rather large cohort of patients.

4. Materials and Methods

4.1. Aim

Aim of this study was to examine the impact of global methylation states, i.e., global CpG islands (CGI) methylation and global gene body (GB) methylation measures, on the outcome of patients with solid cancers.

4.2. Patient Cohorts

Reduced representation bisulfite sequencing (RRBS) was performed with DNA isolated from 25 fresh frozen and 45 formalin-fixed and paraffin-embedded (FFPE) high grade serous ovarian cancer tumor tissues. The 25 samples were a subset of the 45 cohort. In total all tissue samples were from 45 patients with diagnosed high grade (grade 2 or 3), late stage (FIGO III or IV), and serous epithelial ovarian cancer which were characterized in more detail in several publications [19,20,24–30]. Nineteen patients were already deceased. All patients signed an informed consent for providing tissues to this study according the ethical review board (Ethics Committee of the Medical University of Vienna) Nos. 366/2003, 793/2011, and 1076/2018.

4.3. DNA Isolation and Reduced Representation Bisulfite Sequencing (RRBS)

Total DNA from fresh frozen tumor tissues (frozen and stored in the gas phase of liquid N₂) was isolated with the DNeasy Blood and Tissue Kit (QIAGEN, Venlo, Netherlands) and from FFPE slices after macro- (~80%) or laser capture microdissection (~20%, depending on the size and uniformity of the epithelial tumor areas) of tumor tissues with the AllPrep DNA/RNA FFPE Kit (QIAGEN). DNA amount was measured by NanoDrop™ 8000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and the PicoGreen® dsDNA Quantitation Reagent (Promega, Madison, WI, USA) and the quality was assessed by the Agilent High Sensitivity DNA Kit on the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA USA). RRBS was performed and methylation status of CpGs called from 100 ng total DNA according to the protocols published by the Biomedical Sequencing Facility (BSF) in Vienna [12,31].

4.4. Gene Body and CpG Island Analysis

Gene body (GB) of protein coding genes and CpG island (CGI) coordinates for HG38 and HG19 human genomes were obtained from the UCSC genome browser and the GENCODE V25 annotation. From our RRBS data from high grade serious ovarian cancer, the average beta values

(i.e., the methylation level using the ratio of methylated bases to all called bases at a specific CpG position) of all CpGs mapping to either a CGI or the GB of a protein coding gene was calculated (resulting in one average beta value per each CGI and one average beta value per each GB for each sample). The corresponding sequencing depth ($\log_{10}(\text{calling depth} + 0.1)$) for each CpG was weighted in the calculation of the beta values. Similarly, average beta values for GB were calculated using publicly available TCGA beta values without weighting if at least eleven beta values of CpGs were available for the GB. Beta values and clinicopathologic data from TCGA cohorts were obtained with R-package TCGA2STAT 1.2. For every sample the sum of all GBs and CGIs whose average beta value maps to the deciles (0–10%, 10–20%, ..., 90–100%) were summed up and the sums of all deciles for each sample used for correlation to overall survival. The procedure is summarized in Figure 1. With this approach a measurement for the global methylation status in CGI and GBs was introduced which was further used in the subsequent survival analyses.

4.5. Overall Survival Analysis of Gene Body and CpG Island Decile Numbers

To assess the impact of the global methylation status (measured as sum of GBs or CpG islands in each decile) on overall survival (OS, censored data) the robust (permutation based) Significance Analysis of Microarrays (SAM) method (R-package samr 3.0 [32]) was used with a false discovery rate (FDR) cutoff of 5%. Only the deciles which were significantly associated with OS were considered for the next step, the calculation of a methylation definition factor (MDF) for each sample for carcinomas (OvCa, BC, CRC, and PAAC): by dividing the number of unmethylated (e.g., OvCa: 0–10% decile) or methylated (e.g., OvCa: 90–100% decile), thus extreme or defined methylated, GBs by the number of undefined or intermediate methylated (e.g., OvCa: 50–60% and 60–70% deciles) GBs. The concrete used deciles and formulas for each cancer entity are given in Table 1. For glioblastoma a simple methylation-over-unmethylation factor (MUMF) was calculated, dividing all numbers of methylated GBs (averaged beta values > 50%) by all numbers of unmethylated GBs (averaged beta values < 50%), because all deciles over 50% showed impact on OS in one direction and all deciles below 50% in the opposite direction (*cf.* Table 1). To determine the optimal (most informative) cutoff to dichotomize the MDF or MUMF, a Cox regression model was built including all relevant and available clinicopathologic factors, thus considering in the model (shown in Table 1), and the cutoff with the lowest *p*-value determined using the ‘cut-p’ function from R-package survMisc 0.5.5. Finally, patients were dichotomized according the respective optimal cutoff and single and multiple Cox regression analyses performed, the latter with subsequent stepwise backward selection of factors optimizing the Akaike information criterion (AIC) using function stepAIC from R-package MASS 7.3-51_4 [33]. Factors included together with the optimally dichotomized MDF or MUMF in the final Cox models are indicated by names without parentheses in Table 1. Kaplan–Meier estimates and representations of the final multiple Cox models are shown in Figure 2.

4.6. Annotation of Genes with Different Methylation Status and Genes Which Expressions Correlated with the MDF or MUMF

Isomap and T-distributed Stochastic Neighbor Embedding (tSNE) analyses were performed by R-packages RDRToolbox 1.34.0 (function ‘Isomap’) and Rtsne 0.15 (function ‘Rtsne’ with parameters: ‘pca = FALSE, perplexity = 30, theta = 0.5, dims = 2’) [34], respectively. For biological annotation of genes with different averaged methylation beta values for each cancer entity three lists were composed, one with genes of mean (over all samples of one tumor type) < 0.3 averaged GB methylation beta values (Unmethylated), one with genes of mean > 0.7 averaged GB methylation beta values (Methylated) and the third list with all remaining genes (Undefined). Overlap of these lists were determined and presented in Venn-like UpSet diagrams, prepared by the R-package UpSetR 1.4.0 (function *upset*) [35]. Furthermore, these gene lists were analyzed for enrichments in gene ontology (GO) terms (biological processes, cellular components, and molecular functions) and KEGG and Reactome pathways with function gosummaries from R package GOsummaries 2.22.0 [36].

For the correlation of the MDF or MUMF with gene expression the corresponding RNA-sequencing data were used, for OvCa own data [20] and for all other cancer entities the TCGA data, downloaded again with the TCGA2STAT R-package. Expressed genes were filtered using 0.5 counts per million in halve of samples and remaining raw read counts were normalized and weighted according to quality with R-function ‘voomWithQualityWeights’ from R-package limma 3.40.6 [37] (using the cycloless normalization, i.e., cyclicly applying loess normalization). These genes were correlated to the corresponding MDF or MUMF values and gene lists composed of significantly positively or negatively correlated genes (for OvCa and GBM with an FDR cutoff of 10% and all other cancer entities of 5%). Overlap of these gene lists are shown in Figure S15 using the upset method as above. GO and pathway enrichments of these gene lists were analyzed as above. Differentially regulated KEGG-pathways were determined by a combined gene over-representation and perturbation analysis using Signaling Pathway Impact Analysis (SPIA), implemented in the Bioconductor R-package SPIA v2.38.0 [38], with 10,000 permutations and the normal inversion “norminv” *p*-value combination method. Correction for multiple testing was done by the False Discovery Rate (FDR) method (Benjamini-Hochberg). Significant pathways were illustrated with the Bioconductor R-package pathview v1.26.0 [39].

5. Conclusions

Global gene body methylation measures, in carcinomas the quotient of numbers of defined/extreme methylated GBs (either nearly complete methylated or unmethylated) to numbers of undefined/intermediate methylated GBs—the methylation definition factor (MDF)—correlated significantly with overall survival in ovarian cancer, breast cancer, colorectal cancer, and pancreatic cancer, but not in lung cancer. This impact was always independent from known cancer specific clinicopathologic factors and in the case of breast and colorectal cancer also from the molecular subtype or the CpG island methylator phenotype (CIMP) status, respectively (nevertheless, the MDF was correlated to both factors). Interestingly, in ovarian cancer, breast cancer, and colorectal cancer the impact of more defined/extreme methylated genes was an unfavorable predictive factor and in pancreatic cancer the other way round. A similar “definition” factor of CpG islands showed no such impact on OS, at least in ovarian cancer.

In glioblastoma the quotient of numbers of methylated GBs to numbers of unmethylated GBs—the methylation-over-unmethylation factor (MUMF)—was a significant and independent predictor, if high for unfavorable OS.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-6694/12/8/2257/s1>, Figure S1: Correlation plots of averaged GB methylation values between DNA isolated from fresh frozen tumor tissues and macro/microdissected FFPE tumor tissues of the same patients, Figure S2: (A) Histograms of correlation coefficients of averaged GB methylation values between different samples within (fresh frozen) FF and (formalin fixed and paraffin embedded) FFPE tissues (green color), between different samples across FF and FFPE sample (blue color) and between the matched pairs of samples (from the same patients) from FF and FFPE tissues (red color). The black line indicates the correlation coefficient of the MDF values between matched pairs of FF and FFPE tissues (cf. C). (B) Histograms of the calculated “Methylation Definition Factor” (MDF) from FF and FFPE tissues from the same patients and (C) the correlation between these MDF values of the matched pairs of FF and FFPE tissues, Figure S3: (A) Histograms of correlation coefficients of averaged GB methylation values with corresponding averaged CGI methylation values, split into slots of low to high mean methylation levels, i.e., 0–10%, 10–20%, ..., 90–100% (colored white to dark grey). On the left side, split according GB methylation levels and on the right side according CGI methylation levels. (B) Overall correlation of corresponding averaged GB and CGI methylation levels. (C) Association of the coefficient of variation (CV) with the averaged GB methylation level, Figure S4: (A) Histograms of correlation coefficients between GB methylation values and corresponding gene expression values, split into slots of low to high mean GB methylation levels, i.e., 0–10%, 10–20%, ..., 90–100% (colored white to dark grey). (B) Boxplots of log₂ expression values of genes in slots of 0–10%, 10–20%, ..., 90–100% methylated GBs (different grey colors, left y-axis) and corresponding log₁₀ gene length of genes in these slots (different blue colors, right y-axis), Figure S5: Histograms of methylation definition factors (MDF, calculated as follows: (no average GB betas < 0.1 + no average GB betas > 0.9)/no average GB betas between 0.2 and 0.8) of (A) 41 normal human primary cells and tissues (astrocyte (primary cell), foreskin fibroblast (primary cell), islet of Langerhans (tissue), placenta (tissue), hepatocyte (primary cell), mononuclear cell (primary cell), testis (tissue), adrenal gland (tissue), zone of skin

(tissue), mammary epithelial cell (primary cell), uterus (tissue), non-pigmented ciliary epithelial cell (primary cell), epithelial cell of alveolus of lung (primary cell), iris pigment epithelial cell (primary cell), epithelial cell of esophagus (primary cell), epithelial cell of proximal tubule (primary cell), pericardium (tissue), heart left ventricle (tissue), bronchial epithelial cell (primary cell), kidney epithelial cell (primary cell), lung (tissue), renal cortical epithelial cell (primary cell), epithelial cell of prostate (primary cell), kidney (tissue), cardiac fibroblast (primary cell), skeletal muscle cell (primary cell), stomach (tissue), epidermal melanocyte (primary cell), myoblast (primary cell), pancreas (tissue), skeletal muscle myoblast (primary cell), aortic smooth muscle cell (primary cell), osteoblast (primary cell), cardiac muscle cell (primary cell), brain (tissue), skeletal muscle tissue (tissue), choroid plexus epithelial cell (primary cell), liver (tissue), retinal pigment epithelial cell (primary cell), amniotic epithelial cell (primary cell), breast (tissue)); (B) 43 cancer cell lines (MCF, 10A, LNCaP, K562, IMR-90, MCF-7, HEK293, H1, HTR-8/SVneo, NB4, U-87, MG, BJ, HepG2, Caco-2, HeLa-S3, GM19240, HL-60, BE2C, T47D, CMK, Panc1, A549, GM19239, AG09319, NT2/D1, GM06990, HCT116, SK-N-SH, GM12891, SK-N-MC, AG04450, Ishikawa, AG10803, Jurkat, AG08470, AG09309, GM12878, UCH-1, PFSK-1, OVCAR3, HGPS, GM12892, AG04449, hTERT-HM); (C) fresh frozen ovarian cancer tumor tissues; and (D) formalin fixed paraffin embedded (FFPE) ovarian cancer tumor tissues, Figure S6: Distribution of average gene body methylation values (beta values \pm CI95 over samples) of gene bodies of protein coding genes. (A) 45 ovarian cancer tumor tissues; (B) 43 cancer cell lines (MCF, 10A, LNCaP, K562, IMR-90, MCF-7, HEK293, H1, HTR-8/SVneo, NB4, U-87, MG, BJ, HepG2, Caco-2, HeLa-S3, GM19240, HL-60, BE2C, T47D, CMK, Panc1, A549, GM19239, AG09319, NT2/D1, GM06990, HCT116, SK-N-SH, GM12891, SK-N-MC, AG04450, Ishikawa, AG10803, Jurkat, AG08470, AG09309, GM12878, UCH-1, PFSK-1, OVCAR3, HGPS, GM12892, AG04449, hTERT-HM); (C) 41 normal human primary cells and tissues (astrocyte (primary cell), foreskin fibroblast (primary cell), islet of Langerhans (tissue), placenta (tissue), hepatocyte (primary cell), mononuclear cell (primary cell), testis (tissue), adrenal gland (tissue), zone of skin (tissue), mammary epithelial cell (primary cell), uterus (tissue), non-pigmented ciliary epithelial cell (primary cell), epithelial cell of alveolus of lung (primary cell), iris pigment epithelial cell (primary cell), epithelial cell of esophagus (primary cell), epithelial cell of proximal tubule (primary cell), pericardium (tissue), heart left ventricle (tissue), bronchial epithelial cell (primary cell), kidney epithelial cell (primary cell), lung (tissue), renal cortical epithelial cell (primary cell), epithelial cell of prostate (primary cell), kidney (tissue), cardiac fibroblast (primary cell), skeletal muscle cell (primary cell), stomach (tissue), epidermal melanocyte (primary cell), myoblast (primary cell), pancreas (tissue), skeletal muscle myoblast (primary cell), aortic smooth muscle cell (primary cell), osteoblast (primary cell), cardiac muscle cell (primary cell), brain (tissue), skeletal muscle tissue (tissue), choroid plexus epithelial cell (primary cell), liver (tissue), retinal pigment epithelial cell (primary cell), amniotic epithelial cell (primary cell), breast (tissue)), Figure S7: Correlation of the methylation definition factor (MDF), Y-axis, with the breast cancer subtypes, determined with the PAM50 gene signature. Red line, cutoff for the optimal dichotomization of the MDF as obtained from multiple Cox regression models, Figure S8: The CpG island methylator phenotype (CIMP) status determined from TCGA colorectal cancer samples according to Hinoue et al. [1]. The rRR, CIMP-H; rRL, CIMP-L; rLL, type 3; rLR, type 4, Figure S9: Correlation of the methylation definition factor (MDF), Y-axis, with the CpG Island Methylator Phenotypes determined with the function `bclTree` of R-package RPMM 1.25. Red line, cutoff for the optimal dichotomization of the MDF as obtained from multiple Cox regression models, Figure S10: Isomap representation of all samples from all cancer entities using the averaged gene body beta values (if present and without missing data in all samples, $N = 7403$), Figure S11: Overlap of (A) genes with methylated gene bodies (mean averaged beta values > 0.7), (B) with intermediate methylated gene bodies (mean averaged beta values < 0.7 and > 0.3), and (C) unmethylated (mean averaged beta value < 0.3) over all indicated cancer types. Only gene bodies with available averaged beta values present in all cancer entities were considered; $n = 7.801$, Figure S12: Histograms showing the percentages of genes which are consistently annotated to a specific GB methylation slot, divided into three mean GB methylation slots, low methylated (< 0.3 ; light grey), intermediate methylated (≥ 0.3 and ≤ 0.7 ; grey), and high methylated (> 0.7 ; dark grey), Figure S13: (A) Overlap of gene lists consistently categorized (i.e., in $> 90\%$ of samples) to one of the GB methylation slots (high, > 0.7 ; intermediate; and low, < 0.3) from different cancer entities, derived from data shown in Figure S12. (B) Subset of the first 13 gene overlaps from (A), Figure S14: Enrichment plots of gene ontology terms and pathways of lists of consistently as high, intermediate, or low methylated categorized genes. Enrichments are indicated in word clouds by size and color (p -values). Overlap of gene lists across cancer entities are shown in Figure S13, Figure S15: Overlap of genes positively (red) or negatively (green) correlated to the respective MDF/MUMF in different cancer entities. For OvCa and GBM a 10% FDR cutoff, for all other cancer entities, a 5% FDR cutoff was used, Figure S16: SPIA evidence plot of significantly deregulated KEGG pathways, blue and red dots with KEGG pathway numbers (blue line: cutoff FDR $< 5\%$). For colorectal cancer and pancreatic carcinoma, no significant pathways were revealed (cf. S Table S1), Figure S17: Most significant KEGG pathways associated with the Methylation Definition Factor (MDF) in ovarian and breast cancer or the Methylation-over-Unmethylation Factor (MUMF) for glioblastoma. (A) Inhibited pathway “Systemic Lupus Erythematosus” in ovarian cancer. (B) Inhibited pathway “Cytokine-Cytokine Receptor Interaction” in breast cancer, and (C) two identical significant pathways, one inhibited, “Taste Transduction” and one activated, “Choline Metabolism in Cancer” in glioblastoma. Green-to-red color bar represents \log_2 fold changes correlated to the MDF or MUMF, Tables S1: SPIA results tables.

Author Contributions: D.P. conceived and designed the analysis; E.S. and A.B.-H. prepared DNA and processed samples; S.A. contributed samples and clinical data; S.D. did the pathological review and contributed samples; D.P. performed the analysis and wrote the paper. All authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by funds of the Oesterreichische Nationalbank (Anniversary Fund, project No. 14595) and The Austrian Science Fund (FWF) project no. P28137.

Acknowledgments: RRBS and base calling was performed by Christoph Bock from the Biomedical Sequencing Facility (BSF).

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Fardi, M.; Solali, S.; Farshdousti Hagh, M. Epigenetic mechanisms as a new approach in cancer treatment: An updated review. *Genes Dis.* **2018**, *5*, 304–311. [[CrossRef](#)] [[PubMed](#)]
2. Hanahan, D.; Weinberg, R.A. The hallmarks of cancer. *Cell* **2000**, *100*, 57–70. [[CrossRef](#)]
3. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674. [[CrossRef](#)] [[PubMed](#)]
4. Schubeler, D. Function and information content of DNA methylation. *Nature* **2015**, *517*, 321–326. [[CrossRef](#)] [[PubMed](#)]
5. Yang, X.; Han, H.; De Carvalho, D.D.; Lay, F.D.; Jones, P.A.; Liang, G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **2014**, *26*, 577–590. [[CrossRef](#)] [[PubMed](#)]
6. Hellman, A.; Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **2007**, *315*, 1141–1143. [[CrossRef](#)]
7. Jones, P.A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **2012**, *13*, 484–492. [[CrossRef](#)]
8. Lee, S.M.; Choi, W.Y.; Lee, J.; Kim, Y.J. The regulatory mechanisms of intragenic DNA methylation. *Epigenomics* **2015**, *7*, 527–531. [[CrossRef](#)]
9. Arechederra, M.; Daian, F.; Yim, A.; Bazai, S.K.; Richelme, S.; Dono, R.; Saurin, A.J.; Habermann, B.H.; Maina, F. Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer. *Nat. Commun.* **2018**, *9*, 3164. [[CrossRef](#)]
10. Jorda, M.; Diez-Villanueva, A.; Mallona, I.; Martin, B.; Lois, S.; Barrera, V.; Esteller, M.; Vavouri, T.; Peinado, M.A. The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells. *Genome. Res.* **2017**, *27*, 118–132. [[CrossRef](#)]
11. Klein Hesselink, E.N.; Zafon, C.; Villalmanzo, N.; Iglesias, C.; van Hemel, B.M.; Klein Hesselink, M.S.; Montero-Conde, C.; Buj, R.; Mauricio, D.; Peinado, M.A.; et al. Increased Global DNA Hypomethylation in Distant Metastatic and Dedifferentiated Thyroid Cancer. *J. Clin. Endocrinol. Metab.* **2018**, *103*, 397–406. [[CrossRef](#)] [[PubMed](#)]
12. Sheffield, N.C.; Pierron, G.; Klughammer, J.; Datlinger, P.; Schonegger, A.; Schuster, M.; Hadler, J.; Surdez, D.; Guillemot, D.; Lapouble, E.; et al. DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma. *Nat. Med.* **2017**, *23*, 386–395. [[CrossRef](#)]
13. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **2012**, *490*, 61–70. [[CrossRef](#)] [[PubMed](#)]
14. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **2012**, *487*, 330–337. [[CrossRef](#)] [[PubMed](#)]
15. Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **2017**, *32*, 185–203. [[CrossRef](#)] [[PubMed](#)]
16. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **2014**, *511*, 543–550. [[CrossRef](#)]
17. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **2008**, *455*, 1061–1068. [[CrossRef](#)]
18. Hinoue, T.; Weisenberger, D.J.; Lange, C.P.; Shen, H.; Byun, H.M.; Van Den Berg, D.; Malik, S.; Pan, F.; Noushmehr, H.; van Dijk, C.M.; et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res.* **2012**, *22*, 271–282. [[CrossRef](#)]
19. Auer, K.; Bachmayr-Heyda, A.; Aust, S.; Grunt, T.W.; Pils, D. Comparative transcriptome analysis links distinct peritoneal tumor spread types, miliary and non-miliary, with putative origin, tubes and ovaries, in high grade serous ovarian cancer. *Cancer Lett* **2016**, *388*, 158–166. [[CrossRef](#)]

20. Auer, K.; Bachmayr-Heyda, A.; Aust, S.; Sukhbaatar, N.; Reiner, A.T.; Grimm, C.; Horvat, R.; Zeillinger, R.; Pils, D. Peritoneal tumor spread in serous ovarian cancer-epithelial mesenchymal status and outcome. *Oncotarget* **2015**, *6*, 17261–17275. [[CrossRef](#)]
21. Michalak, E.M.; Burr, M.L.; Bannister, A.J.; Dawson, M.A. The roles of DNA, RNA and histone methylation in ageing and cancer. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 573–589. [[CrossRef](#)] [[PubMed](#)]
22. Neri, F.; Rapelli, S.; Krepelova, A.; Incarnato, D.; Parlato, C.; Basile, G.; Maldotti, M.; Anselmi, F.; Oliviero, S. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **2017**, *543*, 72–77. [[CrossRef](#)] [[PubMed](#)]
23. Lev Maor, G.; Yearim, A.; Ast, G. The alternative role of DNA methylation in splicing regulation. *Trends Genet* **2015**, *31*, 274–280. [[CrossRef](#)]
24. Bachmayr-Heyda, A.; Reiner, A.T.; Auer, K.; Sukhbaatar, N.; Aust, S.; Bachleitner-Hofmann, T.; Mesteri, I.; Grunt, T.W.; Zeillinger, R.; Pils, D. Correlation of circular RNA abundance with proliferation—exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci. Rep.* **2015**, *5*, 8057. [[CrossRef](#)] [[PubMed](#)]
25. Auer, K.; Bachmayr-Heyda, A.; Sukhbaatar, N.; Aust, S.; Schmetterer, K.G.; Meier, S.M.; Gerner, C.; Grimm, C.; Horvat, R.; Pils, D. Role of the immune system in the peritoneal tumor spread of high grade serous ovarian cancer. *Oncotarget* **2016**, *7*, 61336–61354. [[CrossRef](#)] [[PubMed](#)]
26. Aust, S.; Felix, S.; Auer, K.; Bachmayr-Heyda, A.; Kenner, L.; Dekan, S.; Meier, S.M.; Gerner, C.; Grimm, C.; Pils, D. Absence of PD-L1 on tumor cells is associated with reduced MHC I expression and PD-L1 expression increases in recurrent serous ovarian cancer. *Sci. Rep.* **2017**, *7*, 42929. [[CrossRef](#)] [[PubMed](#)]
27. Aust, S.; Knogler, T.; Pils, D.; Obermayr, E.; Reinhaller, A.; Zahn, L.; Radlgruber, I.; Mayerhoefer, M.E.; Grimm, C.; Polterauer, S. Skeletal Muscle Depletion and Markers for Cancer Cachexia Are Strong Prognostic Factors in Epithelial Ovarian Cancer. *PLoS ONE* **2015**, *10*, e0140403. [[CrossRef](#)]
28. Bachmayr-Heyda, A.; Auer, K.; Sukhbaatar, N.; Aust, S.; Deycmar, S.; Reiner, A.T.; Polterauer, S.; Dekan, S.; Pils, D. Small RNAs and the competing endogenous RNA network in high grade serous ovarian cancer tumor spread. *Oncotarget* **2016**, *7*, 39640–39653. [[CrossRef](#)]
29. Bachmayr-Heyda, A.; Aust, S.; Auer, K.; Meier, S.M.; Schmetterer, K.G.; Dekan, S.; Gerner, C.; Pils, D. Integrative Systemic and Local Metabolomics with Impact on Survival in High-Grade Serous Ovarian Cancer. *Clin. Cancer Res.* **2017**, *23*, 2081–2092. [[CrossRef](#)]
30. Sukhbaatar, N.; Bachmayr-Heyda, A.; Auer, K.; Aust, S.; Deycmar, S.; Horvat, R.; Pils, D. Two different, mutually exclusively distributed, TP53 mutations in ovarian and peritoneal tumor tissues of a serous ovarian cancer patient: Indicative for tumor origin? *Cold Spring Harb. Mol. Case Stud.* **2017**, *3*. [[CrossRef](#)]
31. Klughammer, J.; Datlinger, P.; Printz, D.; Sheffield, N.C.; Farlik, M.; Hadler, J.; Fritsch, G.; Bock, C. Differential DNA Methylation Analysis without a Reference Genome. *Cell Rep.* **2015**, *13*, 2621–2633. [[CrossRef](#)] [[PubMed](#)]
32. Tusher, V.G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5116–5121. [[CrossRef](#)]
33. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*; Springer: New York, NY, USA, 2003.
34. van der Maaten, L. Accelerating t-SNE using Tree-Based Algorithms. *J. Mach Learn Res.* **2014**, *15*, 3221–3245.
35. Conway, J.R.; Lex, A.; Gehlenborg, N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2017**, *33*, 2938–2940. [[CrossRef](#)] [[PubMed](#)]
36. Kolde, R.; Vilo, J. GOsummaries: An R Package for Visual Functional Annotation of Experimental Data. *F1000Res* **2015**, *4*, 574. [[CrossRef](#)]
37. Law, C.W.; Chen, Y.; Shi, W.; Smyth, G.K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **2014**, *15*, R29. [[CrossRef](#)]
38. Tarca, A.L.; Draghici, S.; Khatri, P.; Hassan, S.S.; Mittal, P.; Kim, J.S.; Kim, C.J.; Kusanovic, J.P.; Romero, R. A novel signaling pathway impact analysis. *Bioinformatics* **2009**, *25*, 75–82. [[CrossRef](#)]
39. Luo, W.; Brouwer, C. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **2013**, *29*, 1830–1831. [[CrossRef](#)]

