FACULDADE DE MEDICINA
UNIVERSIDADE D
COIMBRA

MESTRADO INTEGRADO EM MEDICINA – TRABALHO FINAL

FRANCISCA RAFAELA MARTINS NUNES

IDENTIFICATION OF NEW EPIGENETIC BIOMARKERS IN UROTHELIAL BLADDER CANCER

ARTIGO CIENTÍFICO ORIGINAL

ÁREA CIENTÍFICA FISIOPATOLOGIA

TRABALHO REALIZADO SOB ORIENTAÇÃO DE:

DOUTORA JOANA DIAS APOLÓNIO

PROFESSOR DOUTOR RICARDO ROMÃO NAZÁRIO LEÃO

Faculdade de Medicina, Universidade de Coimbra

Janeiro, 2022

IDENTIFICATION OF NEW EPIGENETIC BIOMARKERS IN UROTHELIAL BLADDER CANCER

Francisca Rafaela Martins Nunes[1], Joana Dias Apolónio[2,3], Ricardo Romão Nazário Leão[1*]

* Correspondence: romaoleao@gmail.com

[1] Faculty of Medicine, University of Coimbra, Azinhaga de Santa Comba, Celas 3000-548 Coimbra, Portugal

[2] Algarve Biomedical Center Research Institute (ABC-RI), University of Algarve, Campus Gambelas, Bld.2, 8005-139 Faro, Portugal

[3] Algarve Biomedical Center, Campus Gambelas, 8005-139 Faro, Portugal

Abstract

Introduction: Urothelial bladder cancer is a highly molecularly diverse disease presenting different behaviors and therapeutic responses. Because of its heterogeneity, there is an unmet need to identify diagnostic and prognostic biomarkers that could help individualize patient approach. As an epigenetic event, DNA methylation is an early event in oncogenesis and highly prevalent across a tumor type, suggesting its great potential to develop cancer biomarkers. Therefore, in this study we aim to investigate the role of nine epigenetic alterations as diagnostic and prognostic biomarkers in bladder cancer.

Methods: DNA methylation, gene expression and corresponding clinicopathological information for the TCGA Bladder Cancer cohort were retrieved from the TCGA data portal. Methylation values and gene expression were assessed to determine their association with normal urothelium and malignant bladder tissue. Additionally, we studied the association between methylation values and epidemiological and clinicopathological variables to interrogate its potential to differentiate clinical features. For the prognostic model, Kaplan-Meier Survival curves were generated. Lastly, univariate and multivariate analysis were performed to evaluate the simultaneous impact of methylation and clinicopathological variables on the risk of tumor progression and overall survival.

Results: The methylation β-values of the nine CpG sites involved in our study demonstrated notably different methylation signatures when comparing normal urothelium to bladder cancer. Hypermethylated CpGs, namely cg15165122, cg12374721, cg10224098, cg18081940 and cg04475027, were overrepresented in tumor tissue ($p<0.0001$). By contrast, cg12743248, cg27170427, cg17192862 and cg10216717 showed lower methylation values in tumor samples ($p<0.0001$). Surprisingly, higher methylation levels were not routinely associated to lower gene expression values. Cg10216717 was hypermethylated in smokers and in locally advanced disease, while cg15165122 hypermethylation was an evident feature in localized disease ($p<0.0001$). Besides, we verified that $cg12374721^{high}$, $cg12743248^{high}$ and $cg17192862^{low}$ are risk factors for progression free survival. Interestingly, $cg12374721^{high}$ (HR:3.003 (1.283-7.030) also demonstrated to be the most valuable independent risk factor for disease progression and a risk factor for overall survival.

Discussion/Conclusion: We have identified 9 differentially methylated CpG changes that characterize urothelial bladder cancer. Novel DNA methylation markers were identified, of which cg12374721 (*C17orf93*) shows promise as a diagnostic and independent prognostic marker in bladder cancer progression as it might be in other cancer types.

Keywords: Bladder cancer, DNA methylation, Biomarkers, Diagnostic, Prognostic

Introduction

Urothelial bladder cancer (UBC) is the 10th most common malignancy worldwide and the second most common urologic malignancy after prostate cancer, with 573.278 incident cases per year, globally.(1-3) Although its incidence is three times higher in men, women tend to present a more aggressive disease and worse survival rates. According to histological classification, approximately 75% of the urothelial bladder cancer cases are non-muscle invasive (NMIBC), mostly low-grade tumors. Despite its more indolent nature, NMIBC can recur (70%) and/or progress (20%) to muscle-invasive bladder cancer (MIBC) in a still unpredictable way, reason why lifelong cystoscopy and cytology monitoring is mandatory in NMIBC.

Several scoring systems have been developed to help predicting individual recurrence and progression. However, because these are static variables which do not reflect the dynamic behavior of the tumor, they are still insufficient to predict individual prognosis, posing a true challenge when dealing with these patients.(1, 4) As such, urinary cytology and cystoscopy remain the gold-standard diagnostic and surveillance modalities. Nevertheless, because urinary cytology has poor sensitivity, particularly in low-grade tumors, and cystoscopy is a highly costly and invasive procedure, there is an urgent unmet need to identify reliable, less invasive and more specific biomarkers.

25% of UBC are MIBC at the time of their presentation (>T2) and others progress from NMIBC. These patients are usually managed in a multimodal fashion (neo)adjuvant cisplatin-based chemotherapy, radiotherapy and/or radical cystectomy).(1) Since it is such a molecularly heterogenous disease, treatment modalities display a full-spectrum of responses among different patients. To help individualize patient approach, a consensus MIBC molecular classification system has been proposed consisting of six genetically different classes (luminal papillary, luminal nonspecified, luminal unstable, stroma-rich, basal/squamous, and neuroendocrine-like) based on transcriptome and genome analysis. As for clinical characteristics and prognostic features, these very different molecular profiles have suggested variable differential sensitivities to chemotherapy, radiotherapy or even immunotherapy.(5) Although it is now known that genetic and epigenetic mechanisms can explain much of this diversity, new biomarkers are lacking for a more personalized patient management.

As mentioned, genetic alterations are common in bladder cancer. However, epigenetic dysregulation, namely DNA methylation, is a well-known early event in the tumorigenesis of many types of cancers. It refers to the mechanisms by which the cell phenotype can be

reversibly affected without altering the DNA sequence.(6, 7) To date, methylation signatures have been applied as biomarkers in cancers, due to their dynamic behavior across tumor stage and histological grade, stability between individuals and tissue specificity.(8, 9) Several biological epigenetic biomarkers have already been identified specifically in bladder tumor tissue and liquid biopsy samples. From them, non-invasive tests have been developed to be applied in NMIBC, mainly in urine samples as an adjunct to cystoscopy and cytology, consequently avoiding unnecessary interventions.(10-13).

Recently, an investigation conducted by our group revealed seven new genes with potential clinical application in breast cancer, some of them also reported in studies carried out in other cancers.(14-17) Specifically, the DNA methylation pattern of those 7 genes was linked to an abnormal gene expression in tumor tissue samples and also related with clinical outcomes. These encouraging results highlight the potential of epigenetic alterations to be used as cancer biomarkers.

Therefore, in this study, using the The Cancer Genome Atlas (TCGA) publicly available database we aimed to investigate the role of those epigenetic alterations as diagnostic and prognostic biomarkers in bladder tumors.

Materials and Methods

Open access data

The Cancer Genome Atlas (TCGA) data for the TCGA Bladder Cancer (BLCA) cohort was extracted from the TCGA data portal via the UCSC Cancer Browser Xena (https://xenabrowser.net/). These data are freely and publicly available, therefore no ethical issues were involved in the present study.

DNA methylation, gene expression and corresponding clinical information were retrieved from a total of 436 tissue samples [malignant (n=413) – urothelial bladder cancer (UBC) and benign – normal urothelium (NU) (n=23]. The metastatic tissue samples (n=1) were not considered. We also excluded patients that had received neoadjuvant therapy (n=10), in order to eliminate any interference with methylation analysis. Samples with missing data concerning methylation or gene expression values were also not included (n=5, tumor samples and n=6, normal tissue samples). A final sum of 414 patients (UBC n=397; NU n=17) was obtained, whose clinicopathological features are summarized in **Supplementary Table 1**.

DNA methylation and gene expression analysis

The methylation status of 9 CpG sites (cg15165122, cg12743248, cg12374721, cg27170427, cg17192862, cg10224098, cg18081940, cg10216717, cg04475027) was derived from the Illumina Infinium Human Methylation 450k array, via the UCSC Cancer Browser Xena. The methylation score for each CpG site is represented as beta values ranging from 0 to 1, corresponding to unmethylated and completely methylated DNA, respectively. Gene expression data from the 7 genes of interest (*ANKRD53, EFCAB1, C17orf93, RIMBP2, RNF220, TDRD10, TMEM132C*) was derived from Illumina HiSeq 2000 RNA Sequencing, whose gene-level transcription estimates are expressed in log2(x+1) transformed RSEM normalized counts.

Diagnostic and prognostic value analyses

Methylation values were assessed as continuous variables to firstly determine its association with normal and malignant bladder tissue. Additionally, we studied the association between methylation values and smoking habits, UBC pathological T and N staging, as well as lymphovascular invasion, to interrogate its potential to differentiate clinical features. We also assessed gene expression differences between NU and UBC, so we could further correlate CpG island methylation and gene expression.

To assess the association between CpG methylation and gene expression, we dichotomized our samples into a high and low methylation subgroup, by receiver operating characteristic (ROC) analysis. Methylation cut-off values were obtained by calculating the Youden Index for each one of the nine probes (**Supplementary Table 2**). Using these two subgroups, high and low methylated UBC samples were compared in terms of the respective gene expression estimates.

To evaluate the prognostic ability of the nine CpG sites Kaplan-Meier Survival curves were generated considering the cut-off value previously defined for each CpG site (overall survival (OS) and progression free survival (PFS)).

Statistical analysis

To assess methylation and gene expression differences between NU and UBC alone and associated epidemiological and clinicopathological features, a two-tailed non-parametric Student's *t*-test was performed using GraphPad Prism 8.0.1.

For methylation and gene expression comparisons, we also performed a Student's *t*-test. A *p*-value below 0.05 was considered statistically significant.

For the prognostic model, ROC curves were generated to calculate the area under the curve (AUC), with 95% confidence intervals (CI) and including sensitivity and specificity. The optimum cut-off value for prognosis was achieved by maximizing the sum of sensitivity and specificity (Youden index). Kaplan Meier curves were constructed for overall survival and progression free survival. The logrank test was used to assess whether the survival rates were significantly different between the high and low methylation subgroups for each CG site.

For the assessment of variables associated with disease outcomes (OS and PFS), Cox Proportional Hazards (CPH) methods were used to interrogate univariate significance. Smoking habits (Smoker *vs.* Non-Smoker), pT Stage (Localized *vs.* Locally Advanced disease), Nodal involvement (N0 *vs.* N+), Lymphovascular invasion (LVI *vs.* No LVI) and cg12374721 ($cg12374721^{high}$ *vs.* $cg12374721^{low}$) were the chosen variables to incorporate this analysis for OS outcome.

Multivariable Cox Proportional Hazards models were used to evaluate the simultaneous impact of those variables on the risk of tumor progression and overall survival, while adjusting for variables found to be significant at a level < 0.05 on univariable analysis. IMB SPSS version 28 was used to perform the analysis.

Results

Cohort Characteristics

A total of 397 UBC samples were used in our study, corresponding to patients with a mean age at diagnosis of 68 years. A gender predominance was evident, with 73% male patients. Most patients had muscle-invasive disease without nodal involvement and absence of metastasis. The vast majority of tumors harbored in patients with smoking habits (72%). Other variables, including prognostic variables such as OS and PFS, are summarized in **Supplementary Table 1**.

Methylation levels are differently elevated in NU and UBC

The methylation beta-values of the nine CpG sites involved in our study demonstrate notably different methylation signatures when comparing NU to UBC. Specifically, cg15165122 (*ANKDR53*), cg12374721 (*C17orf93*), cg10224098 (*RNF220*), cg18081940 (*TDRD10*) and cg04475027 (*TMEM132C*) had statistically significant higher methylation levels (p<0.0001, **Figure 1**). On the contrary, cg12743248 (*EFCAB1*), cg27170427 (RIMBP2), cg17192862 (*RIMBP2*) and cg10216717 (*TMEM132C*) showed lower methylation values in UBC samples. (p<0.0001, **Figure 1**).

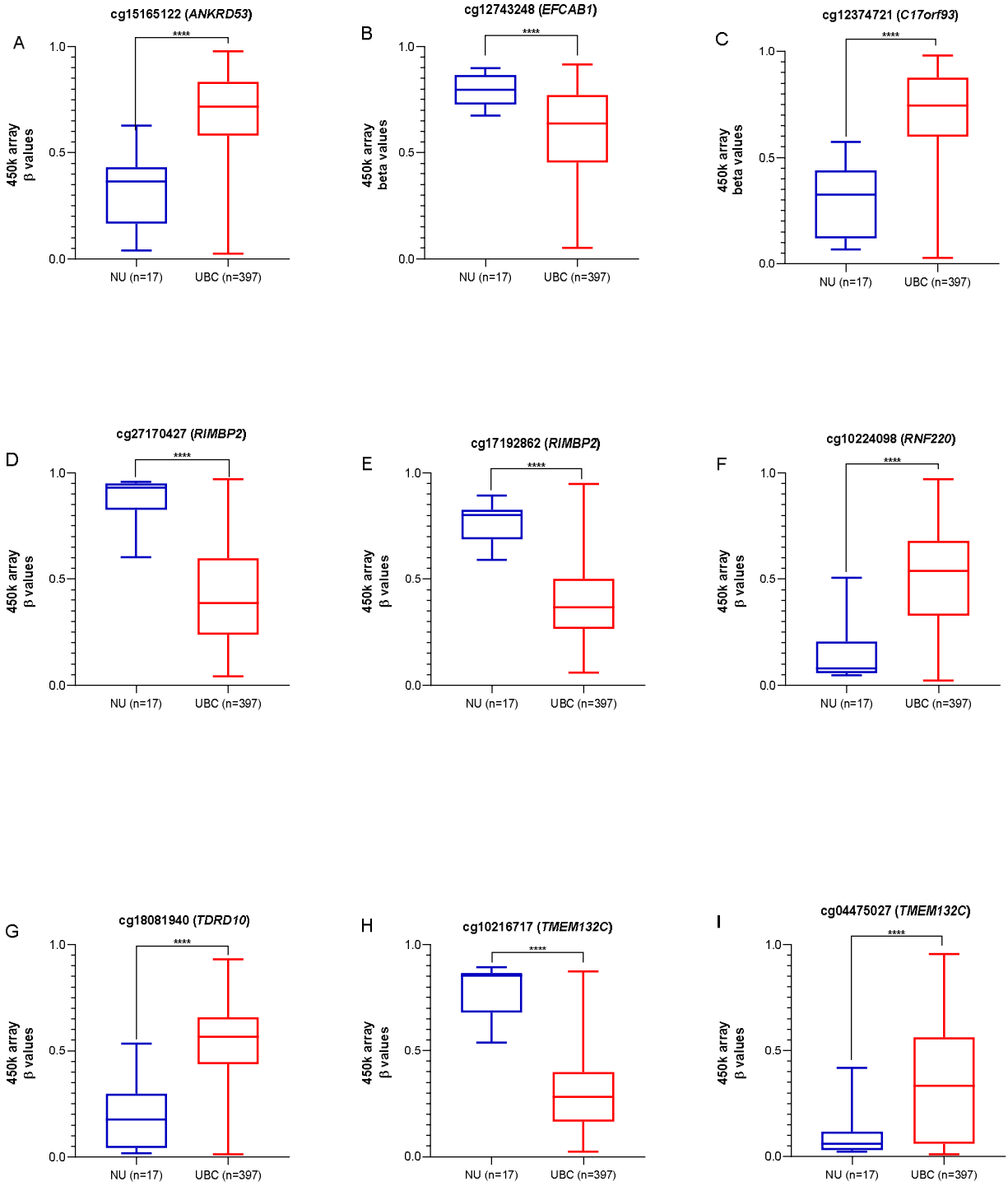


Figure 1. CpG methylation levels in Normal Urothelium (NU) and Urothelial Bladder Cancer (UBC). Methylation levels are expressed in β-values, varying from 0 to 1 (****, p <0.0001).

Methylation levels according to epidemiological and clinical features

As known, smoking is the most important risk factor for the development of UBC. Thus, when looking at the levels of methylation and CpG sites we found that cg10216717 (*TMEM132C*) demonstrated higher methylation values in current and former smokers, when compared to non-smokers (**Supplementary Figure 1**, $p<0.0001$). None of the others CpG sites portraited any difference.

As invasiveness (T stage) and lymph node involvement (N stage) are relevant predictors of UBC tumor progression, we evaluated CpG methylation ability to distinguish these stages. Cg15165122 (*ANKRD53*) site demonstrated higher methylation levels in localized disease ($p<0.001$, **Supplementary Figure 2**) and cg10216717 (*TMEM132C*) presented the highest methylation values in locally advanced disease ($p<0.0001$**, Supplementary Figure 2**).

Regarding lymph node involvement and lymphovascular invasion (LVI), no difference was found in methylation values across the nine CpG sites. (**Supplementary Figures 3 and 4**).

Gene expression levels differ between NU and UBC

Then we analyzed gene expression levels for each one of the genes corresponding to the previous CpG sites evaluated (**Figure 2**).

As observed in **Figure 2**, all the seven genes of interest are differentially expressed between NU and UBC with statistically significant results. Genes *EFCAB1* ($p<0.001$), *ANKRD53, C17orf93, RIMBP2, TDRD10* and *TMEM132C* ($p<0.0001$) were downregulated in UBC samples. By contrast, *RNF220* ($p>0.001$) gene was significantly more expressed in malignant tissue.

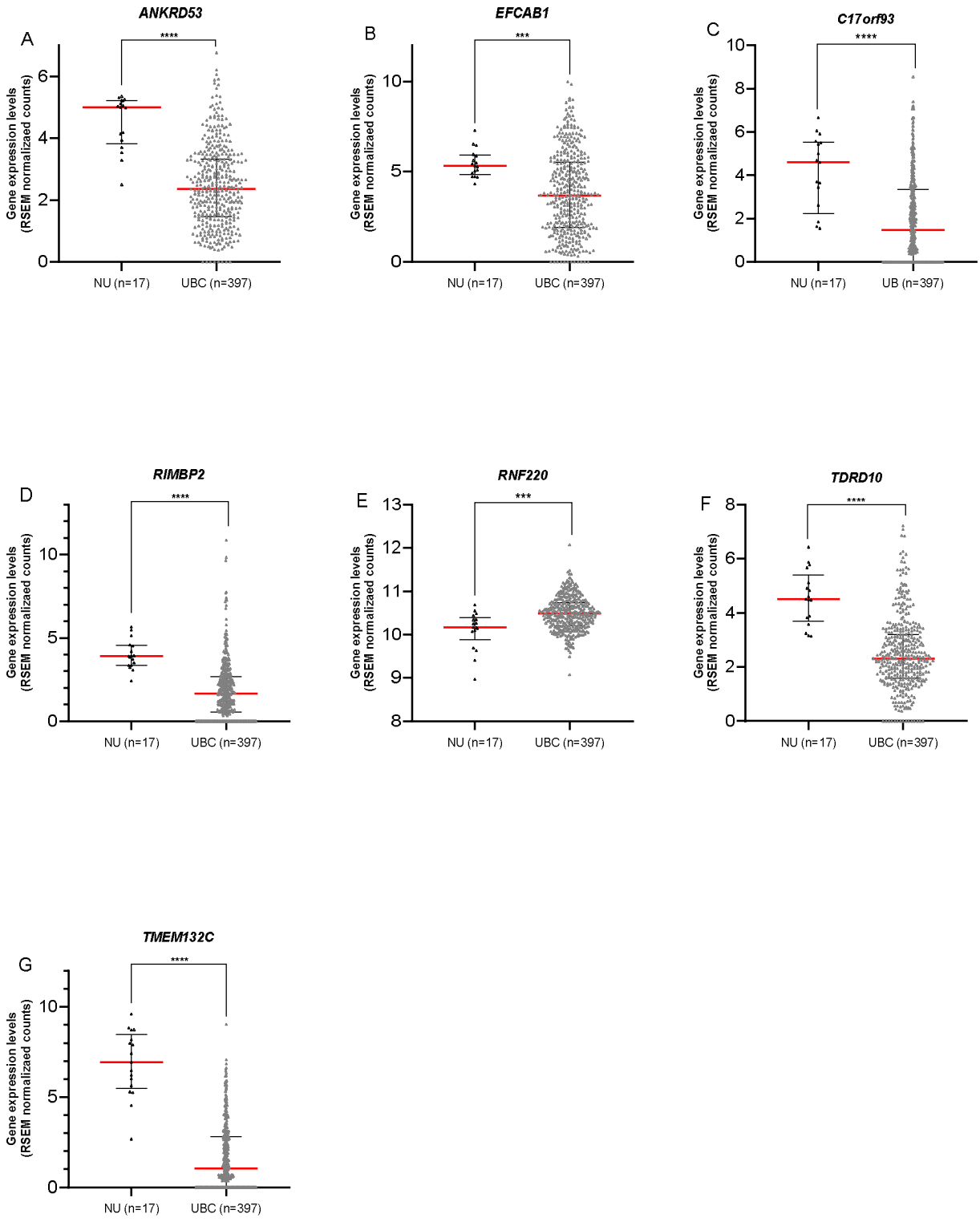


Figure 2. Gene expression levels for each gene of interest in Normal Urothelium (NU) and Urothelial Bladder Cancer (UBC). Gene expression levels are expressed in RSEM normalized counts. (***, $p \leq 0.001$; ****, $p \leq 0.0001$).

Gene expression levels are associated with different methylation patterns

To explore the relationship between DNA methylation and gene expression in UBC, we compared both subgroups of CpG methylation signatures (hyper- vs hypomethylated) with the expression changes in the corresponding genes. Among the 9 differentially methylated CpG sites we identified 7 that were associated with differentially expressed genes.

CpG site cg12743248^{high} subgroup showed to be associated with an increase in *EFCAB1* gene expression ($p<0.0001$, **Figure 3**). The same was observed for both *RIMBP2* corresponding cg27170427^{high} and cg17192862^{high} subgroups with matching levels of significance ($p<0.0001$, **Figure 3**). When looking at gene *TMEM132C*, only cg10216717^{high} subgroup was associated with higher levels of gene expression ($p<0.0001$**, Figure 3**).

The other *TMEM132C* corresponding CpG site (cg04475027) also reached statistical significance ($p<0.05$, **Figure 3**); however, it was the cg04475027^{low} subgroup that seemed to be associated with increased gene expression. *C17orf93* and *RNF220* genes showed no difference in gene expression estimates between the two methylation subgroups.

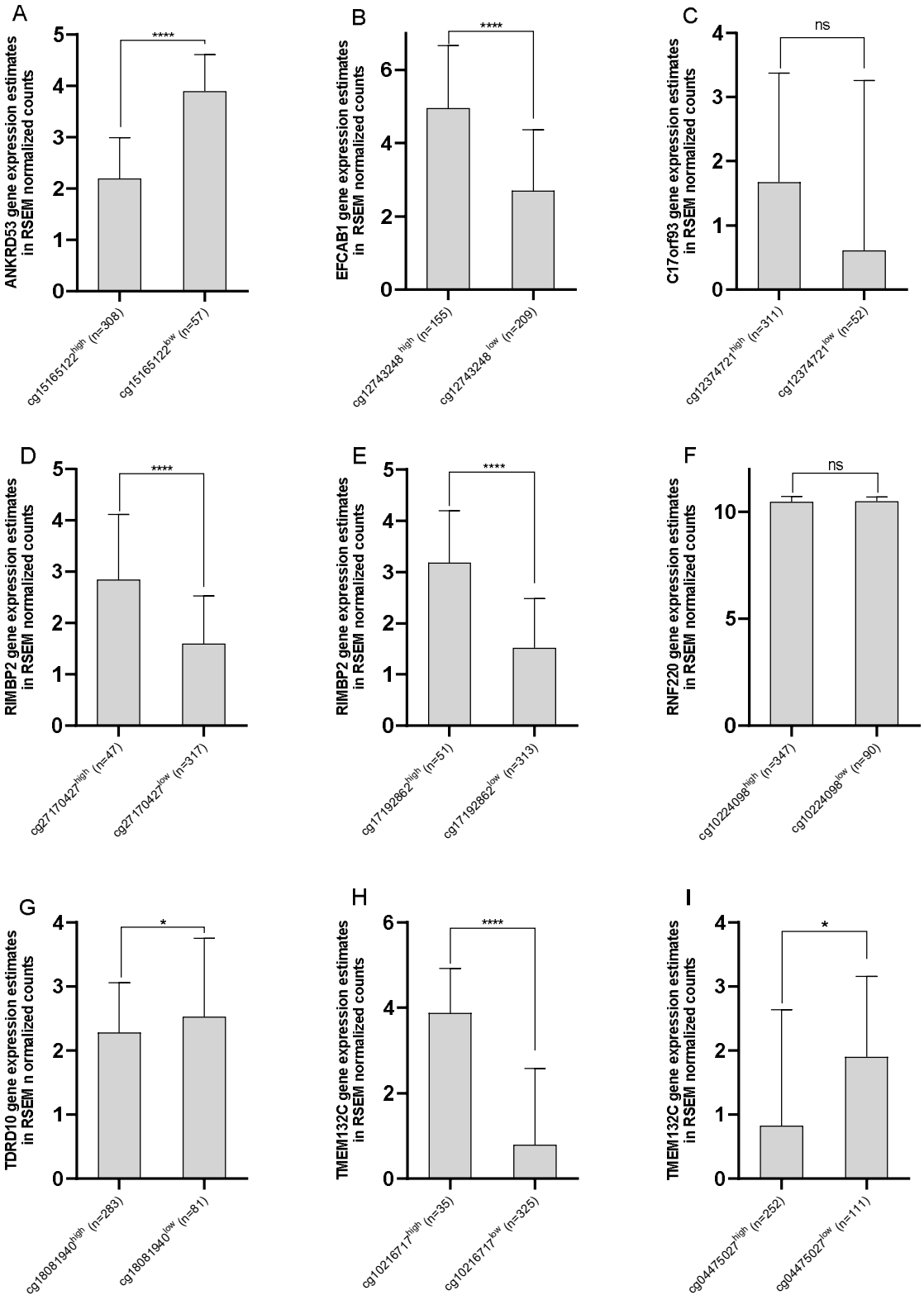


Figure 3. Gene expression in UBC tumors according to CG site hyper and hypomethylated status.

Methylation levels and Clinical Outcomes

Cg12743248 (*EFCAB1*), cg12374721 (*C17orf93*), and cg17192862 (*RIMBP2*) showed differences in PFS. Hypermethylated cg12743248 (HR: 1,419, CI 1,045 to 1,928; Log Rank *p*=0,0220; **Figure 4**) and cg12374721 (HR: 2,989, CI 2,005 to 4,455; Log Rank *p*=0,00012; **Figure 4**) were associated with a significant decrease in PFS. Interestingly, for cg17192862 (*RIMBP2*), whose methylation status is positively correlated with gene expression, the hypomethylated subgroup showed to be related to slightly, yet significant, decreased PFS (HR: 0,6049, CI 0,4028 to 0,9084; Log Rank *p*=0,0409; **Figure 4**). For the remaining CpG sites, no significant differences were observed (**Supplementary Figure 5**).

Hypermethylated CpG site cg12374721 (*C17orf93*) tumors also demonstrated significantly lower overall survival (OS) rates as opposed to the hypomethylated subgroup (HR: 1.727, 95% CI 1.148-2.598; Log Rank *p*=0.0301; **Figure 4D**). Methylation status did not reach statistical significance with respect to OS in other CpG sites **(Supplementary Figure 6).**

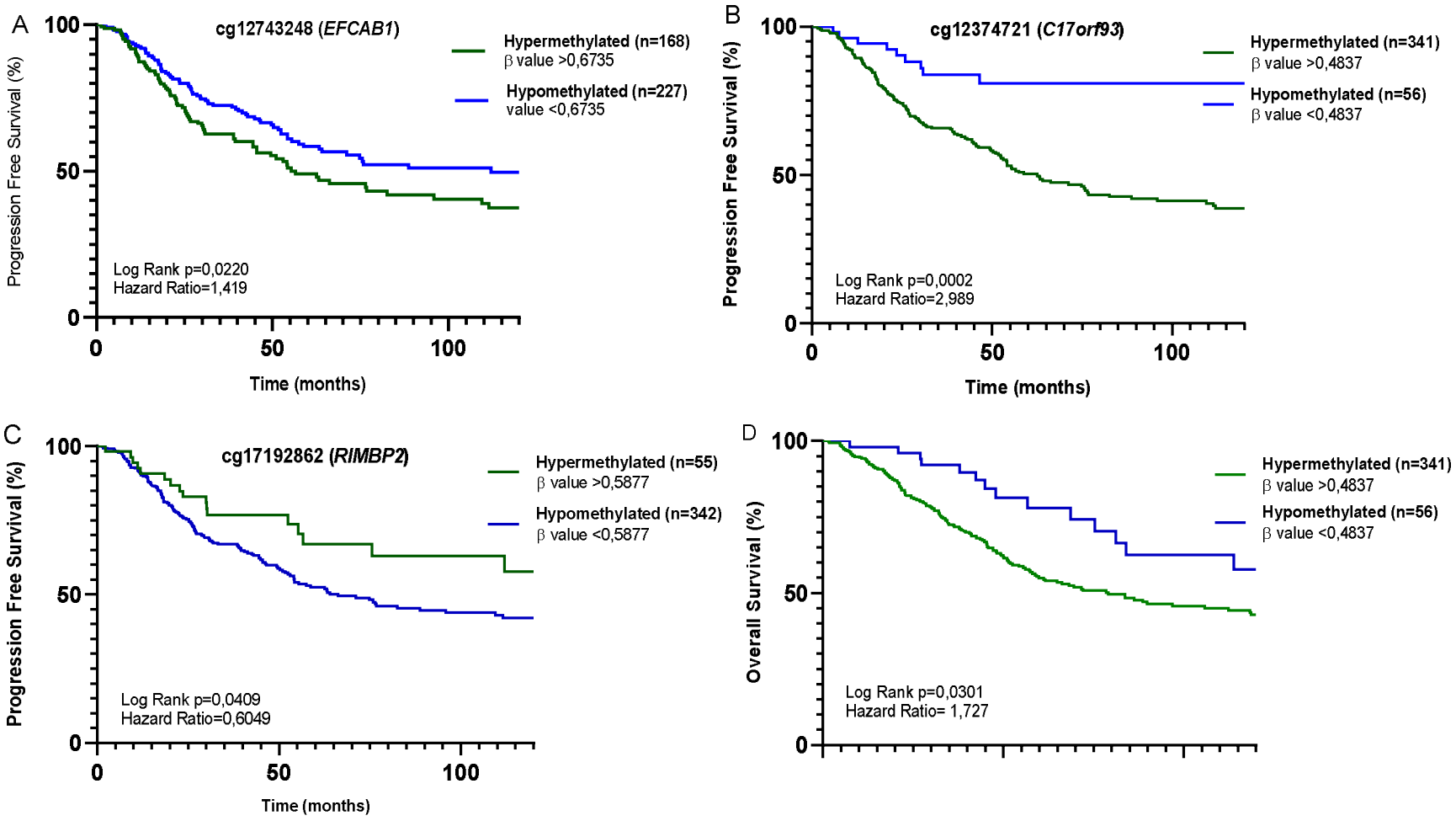


Figure 4. A, B and C. Progression Free Survival according to methylation levels in cg12743248, cg12374721 and cg17192862. Kaplan-Meyer curves in hyper and hypomethylated subgroups (based on Youden Index calculated methylation cut-off values). D. Survival analysis according to cg12374721 methylations levels in (UBC) tumors. (HR: 1.727, 95% CI 1.148-2.598; Log Rank *p*=0.0301).

Univariate and Multivariate analysis for PFS and OS outcomes

In order to assess both univariate and multivariate significance of clinicopathological and CpG methylation covariates in UBC prognosis, univariate and multivariate cox regression analysis were performed for prognostic models. Pathological T and N stage, LVI invasion, cg12374721 (*C17orf93*), cg12743248 (*EFCAB1*) and cg17192862 (*RIMBP2*) methylation status all achieved significance in influencing the PFS of UBC (**Table 1**). Locally advanced disease (2.187 (HR:1.485-3.220), p<0.001), pN+ involvement (HR:2.827 (2.037-3.924), p<0.001) and LVI invasion (HR:2.305 (1.570-3.383), p<0.001) were all associated with approximately twice the chance of progression. Similarly, cg12743248^{high} methylation status granted a 43% risk of progression in comparison to the cg12743248^{low} methylation subgroup (HR:1.428 (1.056-1.932), p=0.021). *C17orf93* associated cg12374721^{high} methylation subgroup was associated with the highest risk of progression (HR:3.181 (1.678-6.029), p<0.001), whereas *RIMBP2* associated cg17192862^{high} subgroup seemed to have a 40% lower risk of progression (HR:0.603 (0.369-0.985), p=0.043). When adjusting these variables for PFS outcome, we realized that not only did pathological T stage and node involvement reach significance, but cg12374721^{high} subgroup could also independently influence PFS (HR:3.003 (1.283-7.030), p=0.011).

In respect to OS outcome, locally advanced disease (HR: 2.101 (1.438-3.070), p<0.001), pN+ involvement (HR: 2.317 (1.686-3.186), p<0.001), LVI invasion (HR: 2.267 (1.553-3.311), p<0.001) and cg12374721^{high} (HR: 1.738 (1.054-2.868), p=0.030) subgroup all reached univariate significance (**Table 2**). The first three clinicopathological covariates showed approximately twice the chance of dying than the opposite variables. Cg12374721^{high} subgroup achieved a lower, yet significant, 73% higher risk of death than the hypomethylated subgroup. However, only pN+ UBC samples could independently predict worse survival (HR: 1.668 (1.060-2.623), p=0.027) in multivariate analysis, as no other covariable reached statistical significance.

Table 1. Univariable and Multivariable Cox proportional hazards models of time for Progression free survival (n=236).

	UNIVARIABLE		MULTIVARIABLE	
Variable	**HR (95%CI)**	***p* value**	**HR (95%CI)**	***p* value**
Smoking Habits (Smokers *vs.* Non-smokers)	1.049 (0.747-1.475)	0.782	-	-
Stage (Locally Advanced *vs.* Localized Disease)	**2.187 (1.485-3.220)**	**<0.001**	**1.977 (1.172-3.336)**	**0.011**
Nodal Involvement (N+ *vs.* N0)	**2.827 (2.037-3.924)**	**<0.001**	**1.960 (1.236-3.109)**	**0.004**
LVI invasion (LVI *vs*. No LVI)	**2.305 (1.570-3.383)**	**<0.001**	1.261 (0.775-2.051)	0.351
cg12374721 methylation status ($cg12374721^{high}$ *vs.* $cg12374721^{low}$)	**3.181 (1.678-6.029)**	**<0.001**	**3.003 (1.283-7.030)**	**0.011**
cg12743248 methylation status ($cg12743248^{high}$ *vs.* $cg12743248^{low}$)	**1.428 (1.056-1.932)**	**0.021**	1.233 (0.833-1.825)	0.296
cg17192862 methylation status ($cg17192862^{high}$ *vs.* $cg17192862^{low}$)	**0.603 (0.369-0.985)**	**0.043**	0.686 (0.368-1.279)	0.235

Table 2. Univariable and Multivariable Cox proportional hazards models of time for Overall Survival (n=236).

	UNIVARIABLE		MULTIVARIABLE	
Variable	**HR (95%CI)**	***p* value**	**HR (95%CI)**	***p* value**
Smoking Habits (Smokers *vs.* Non-smokers)	1.232 (0.869-1.748)	0.241	-	-
pT Stage (Locally Advanced *vs.* Localized Disease)	**2.101 (1.438-3.070)**	**<0.001**	1.419 (0.871-2.313)	0.160
Nodal Involvement (N+ *vs.* N0)	**2.317 (1.686-3.186)**	**<0.001**	**1.668 (1.060-2.623)**	**0.027**
LVI invasion (LVI *vs*. No LVI)	**2.267 (1.553-3.311)**	**<0.001**	1.513 (0.940-2.435)	0.088
cg12374721 methylation status ($cg12374721^{high}$ *vs.* $cg12374721^{low}$)	**1.738 (1.054-2.868)**	**0.030**	1.743 (0.928-3.272)	0.084

Discussion

Alterations in epigenetic control have been associated with several human pathologic conditions including bladder cancer.(13) Transcriptional inactivation by CpG island promoter hypermethylation is a well-established mechanism for gene silencing in cancer and aberrant methylation has long been known to help distinguish clinical and prognostic features.(18) Using the TCGA bladder cancer cohort we acknowledged that all the 9 CpG sites of interest to our study (cg15165122 - *ANKRD53*; cg12743248 and cg17192862 - *EFCAB1*; cg12374721 - *C17orf93*; cg27170427 and cg17192862 - *RIMBP2*; cg10224098 – *RNF220*; cg10216717, cg04475027 - *TMEM132C*) were differentially methylated between NU and UBC. Most CpG sites (cg15165122, cg18081940 and cg04475027) were hypermethylated in UBC samples and associated with downregulation of the corresponding gene (*ANKRD53*, *C17orf93*, *TDRD10* and *TMEM132C*, respectively). Only cg12743248, cg27170427, cg17192862 and cg10216717 hypermethylation in UBC seemed to be related to upregulation of the corresponding gene expression, contradicting this paradigm. Importantly, 3 CpG sites (cg12374721 – *C17orf93;* cg12743248 – *EFCAB1*; cg17192862 – *RIMBP2*) were able to predict tumor progression, with cg12374721 methylation profile also being able to predict survival as an independent risk factor.

MIBC is an aggressive condition characterized by a high risk of relapse and metastasis.(1) At the molecular level, MIBC is also a heterogeneous disease characterized by genomic instability and a high mutation rate.(5, 19, 20) As such, a consensus molecular classification has been proposed, so one can better stratify a patient, according to prognosis and therapeutic options. This classification system converged on six biologically relevant classes: luminal papillary (LumP), luminal nonspecified (LumNS), luminal unstable (LumU), stroma-rich, basal/squamous (Ba/Sq), and neuroendocrine-like, according to gene signatures and tumor microenvironment.(5) This has proven to be a highly interesting stratification system in patient personalized management, opening doors towards a more personalized medicine. Just like molecular gene signatures and microenvironment infiltration, our study demonstrated that epigenetic changes may also be important independent predictors of tumor progression. This can be useful, for example, in assisting treatment modalities.

Our results firstly showed that methylation levels for the studied CpG sites are routinely different between tumor (UBC) and benign tissue (NU). However, and interestingly, methylation levels are not consistently higher in UBC as expected. Cg15165122,

cg12374721, cg10224098, cg18081940 and cg04475027 were significantly higher methylated in tumor tissue when compared to NU, but cg12743248, cg17192862, cg10216717 and cg27170427 displayed lower methylation levels in UBC. These results might be related to each corresponding gene function – tumor suppressors or oncogenes – and their role in bladder cancer oncogenesis.

When looking at high and low-methylation subgroups and their association with gene expression we verified that the highest levels of methylation were associated with the highest levels of expression in UBC for cg12743248 (*EFCAB1*), cg27170427, cg17192862 (*RIMBP2*) and cg10216717 (*TMEM132C*). However, as cited before, all these CpGs seemed to be less methylated and their gene expression downregulated in UBC compared to NU. Lower levels of *EFCAB1* gene expression have already been described in breast cancer and related to worse survival rates.(24) On the contrary, the biological function of *RIMBP2* is still under investigation and there is no scientific literature relating to it yet. *TMEM132C* gene encodes for a transmembrane protein that to date has only been associated with apocrine breast cancer.(16) Altogether, these results suggest an influence of methylation in gene expression in normal urothelial tissue and their eventual role as tumor suppressors in bladder tissue.

On the contrary, cg15165122, cg18081940 and cg04475027 higher methylated were associated with lower levels of the corresponding gene expression in UBC. Accordingly, these same CpG sites were hypermethylated and their genes downregulated in UBC when compared to NU. Hypermethylation of cg15165122 – *ANKDR53* (Ankyrin Repeat Domain 53) CpG site may prevent the binding of transcription activators thus leading to *AKNRD53* gene silencing and aberrant mitosis in tumor tissue. In fact, previous studies have stated that *ANKRD53*-depleted cells show mitosis chromosome unalignment, prolonged mitosis duration and inability to complete cell division.(21) Cg18081940 site is related to *TDRD10* which belongs to the family of *TDRD - Tudor-Domain Containing Proteins*, whose main function is to recognize the N-terminal arginine-rich motifs via their Tudor domains and help in chromatin and transcriptional regulation, genome stability and RNA metabolism.(22) In line with previous findings in breast cancer(14), our results showed a negative association between hypermethylated cg18081940 and *TDRD10* downregulation, denoting its importance in suppressing tumorigenesis.

Clinicopathological features are still the main known risk factors in bladder cancer. We verified that methylation levels were not differently elevated according to lymph node

involvement or LVI status. LVI status is in fact known to be a risk factor for relapse and we were anticipating an association with higher CpG methylation levels.

Localized tumor samples, however, displayed higher levels of methylation in comparison to locally advanced disease, in respect to cg15165122, which might suggest that this epigenetic modification is an early event in tumorigenesis with other alterations prevailing in tumor progression. Intriguingly, when looking at cg10216717 methylation pattern among the different clinicopathologic features, locally advanced UBC and smokers' samples were related to higher methylation levels than the opposite variables. Indeed, these results may indicate that cg10216717 methylation is a late epigenetic event that acts to promote disease progression and could also be susceptible to environmental influences, especially tobacco smoking.

Our results confirm that clinicopathological variables such as advanced disease, LVI and nodal involvement are known risk factors for disease progression and overall survival. In this study we verified that cg12743248^{high}, cg12743248^{high} and cg17192862^{low} are risk factors for PFS. Strikingly, cg12374721^{high} (HR:3.003 (1.283-7.030) is the most valuable independent risk factor for disease progression and also a risk factor for overall survival. As such, cg12743248 encompasses diagnostic (able to differentiate malignant from normal bladder tissue) and prognostic value.

These results state that promoter methylation of *C17orf93,* namely the tumors displaying cg12374721^{high} methylation signature, correlates to worse survival. These tumors also displayed higher methylation values than NU, but those values were not associated with higher gene expression. Similarly, methylation levels of Cg12374721 were not associated either with pT or nodal invasion, this way confirming its independent characteristics. C17orf93 also known as PRCA2 (Prostate Cancer Susceptibility Candidate 2) is a gene highly expressed in prostate, rectum, colon, and testis.(23) Depending on alternative splicing, it may produce a non-coding RNA or may encode a short protein localized in the nucleus. To date, it has only been studied by our group in breast cancer, and no function has been attributed to it yet. Unlike breast cancer, corresponding cg12374721 methylation levels were higher in normal tissue relative to malignant tissue. However, in this TCGA dataset UBC had lower gene expression values. This corroborates the central paradigm of hypermethylation associated tumor suppressor gene repression, although an oncogenic role has been suggested previously. Additional studies, including validation in vitro and in other cancers, might be helpful in clarifying and supporting these results.

As an epigenetic event, DNA methylation is an early event in oncogenesis and highly prevalent across tumor types, suggesting its great potential to develop cancer biomarkers. Future studies might validate cg12374721 for selecting which patients are at higher risk of progression, which patients should be treated with neoadjuvant chemotherapy, and which patients should receive straightforward surgery to avoid early progression and systemic toxicity.

Conclusion

In this study, we reinforced the role of epigenetics, namely DNA methylation, in bladder cancer carcinogenesis. Using a TCGA bladder cancer cohort we could assess 9 CpG methylation differences and associate them with clinicopathological features and clinical outcomes. Accordingly, both CpG hypermethylation and hypomethylation may be crucial events in bladder carcinogenesis, so further studies are needed to validate their biomarker potential in bladder cancer and other cancer types. While all the 9 CpG sites of interest demonstrate diagnostic potential, three CpG sites - cg12374721, cg12743248 and cg17192862 – indicate that they also may be effective prognostic tools in disease progression and survival.

References

1. Babjuk M, Burger M, Compérat EM, Gontero P, Mostafid AH, Palou J, et al. European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and Carcinoma In Situ) - 2019 Update. Eur Urol. 2019;76(5):639-57.

2. Li Y, Sun L, Guo X, Mo N, Zhang J, Li C. Frontiers in Bladder Cancer Genomic Research. Front Oncol. 2021;11:670729.

3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.

4. Seo KW, Kim BH, Park CH, Kim CI, Chang HS. The efficacy of the EORTC scoring system and risk tables for the prediction of recurrence and progression of non-muscle-invasive bladder cancer after intravesical bacillus calmette-guerin instillation. Korean J Urol. 2010;51(3):165-70.

5. Kamoun A, de Reyniès A, Allory Y, Sjödahl G, Robertson AG, Seiler R, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. Eur Urol. 2020;77(4):420-33.

6. Cheng Y, He C, Wang M, Ma X, Mo F, Yang S, et al. Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. Signal Transduct Target Ther. 2019;4:62.

7. Martinez-Useros J, Martin-Galan M, Florez-Cespedes M, Garcia-Foncillas J. Epigenetics of Most Aggressive Solid Tumors: Pathways, Targets and Treatments. Cancers (Basel). 2021;13(13).

8. Loras A, Segovia C, Ruiz-Cerdá JL. Epigenomic and Metabolomic Integration Reveals Dynamic Metabolic Regulation in Bladder Cancer. Cancers (Basel). 2021;13(11).

9. Ruan W, Chen X, Huang M, Wang H, Chen J, Liang Z, et al. A urine-based DNA methylation assay to facilitate early detection and risk stratification of bladder cancer. Clin Epigenetics. 2021;13(1):91.

10. Piatti P, Chew YC, Suwoto M, Yamada T, Jara B, Jia XY, et al. Clinical evaluation of Bladder CARE, a new epigenetic test for bladder cancer detection in urine samples. Clin Epigenetics. 2021;13(1):84.

11. Mancini M, Righetto M, Zumerle S, Montopoli M, Zattoni F. The Bladder EpiCheck Test as a Non-Invasive Tool Based on the Identification of DNA Methylation in Bladder Cancer Cells in the Urine: A Review of Published Evidence. Int J Mol Sci. 2020;21(18).

12. Lozano F, Raventos CX, Carrion A, Trilla E, Morote J. Current status of genetic urinary biomarkers for surveillance of non-muscle invasive bladder cancer: a systematic review. BMC Urol. 2020;20(1):99.

13. Kim YJ, Kim WJ. Can we use methylation markers as diagnostic and prognostic indicators for bladder cancer? Investig Clin Urol. 2016;57 Suppl 1(Suppl 1):S77-88.

14. de Almeida BP, Apolónio JD, Binnie A, Castelo-Branco P. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. BMC Cancer. 2019;19(1):219.

15. Pan Y, An N, Deng X, Zhang Q, Du X. RNF220 promotes the proliferation of leukaemic cells and reduces the degradation of the Cyclin D1 protein through USP22. Blood Cells Mol Dis. 2021;86:102490.

16. Zhang X, Kang X, Jin L, Bai J, Zhang H, Liu W, et al. ABCC9, NKAPL, and TMEM132C are potential diagnostic and prognostic markers in triple-negative breast cancer. Cell Biol Int. 2020;44(10):2002-10.

17. Xu Y, She Y, Li Y, Li H, Jia Z, Jiang G, et al. Multi-omics analysis at epigenomics and transcriptomics levels reveals prognostic subtypes of lung squamous cell carcinoma. Biomed Pharmacother. 2020;125:109859.

18. Kazanets A, Shorstova T, Hilmi K, Marques M, Witcher M. Epigenetic silencing of tumor suppressor genes: Paradigms, puzzles, and potential. Biochim Biophys Acta. 2016;1865(2):275-88.

19. Li HT, Duymich CE, Weisenberger DJ, Liang G. Genetic and Epigenetic Alterations in Bladder Cancer. Int Neurourol J. 2016;20(Suppl 2):S84-94.

20. Leão R, Lee D, Figueiredo A, Hermanns T, Wild P, Komosa M, et al. Combined genetic and epigenetic alterations of the TERT promoter affect clinical and biological behavior of bladder cancer. Int J Cancer. 2019;144(7):1676-84.

21. Kim S, Jang CY. ANKRD53 interacts with DDA3 and regulates chromosome integrity during mitosis. Biochem Biophys Res Commun. 2016;470(3):484-91.

22. Gan B, Chen S, Liu H, Min J, Liu K. Structure and function of eTudor domain containing TDRD proteins. Crit Rev Biochem Mol Biol. 2019;54(2):119-32.

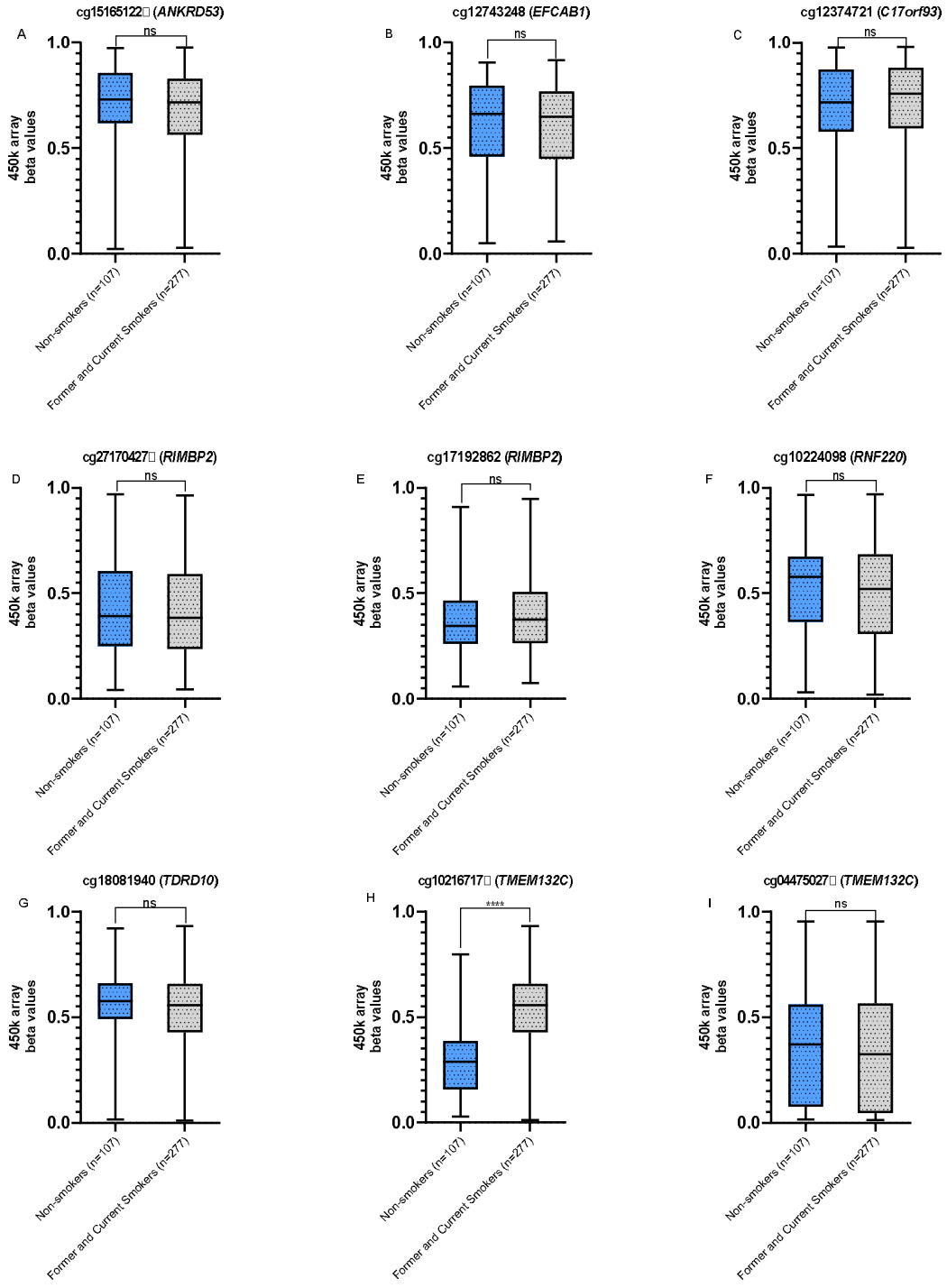
23. Olsson P, Motegi A, Bera TK, Lee B, Pastan I. PRAC2: a new gene expressed in human prostate and prostate cancer. Prostate. 2003;56(2):123-30.

Supplementary tables and figures

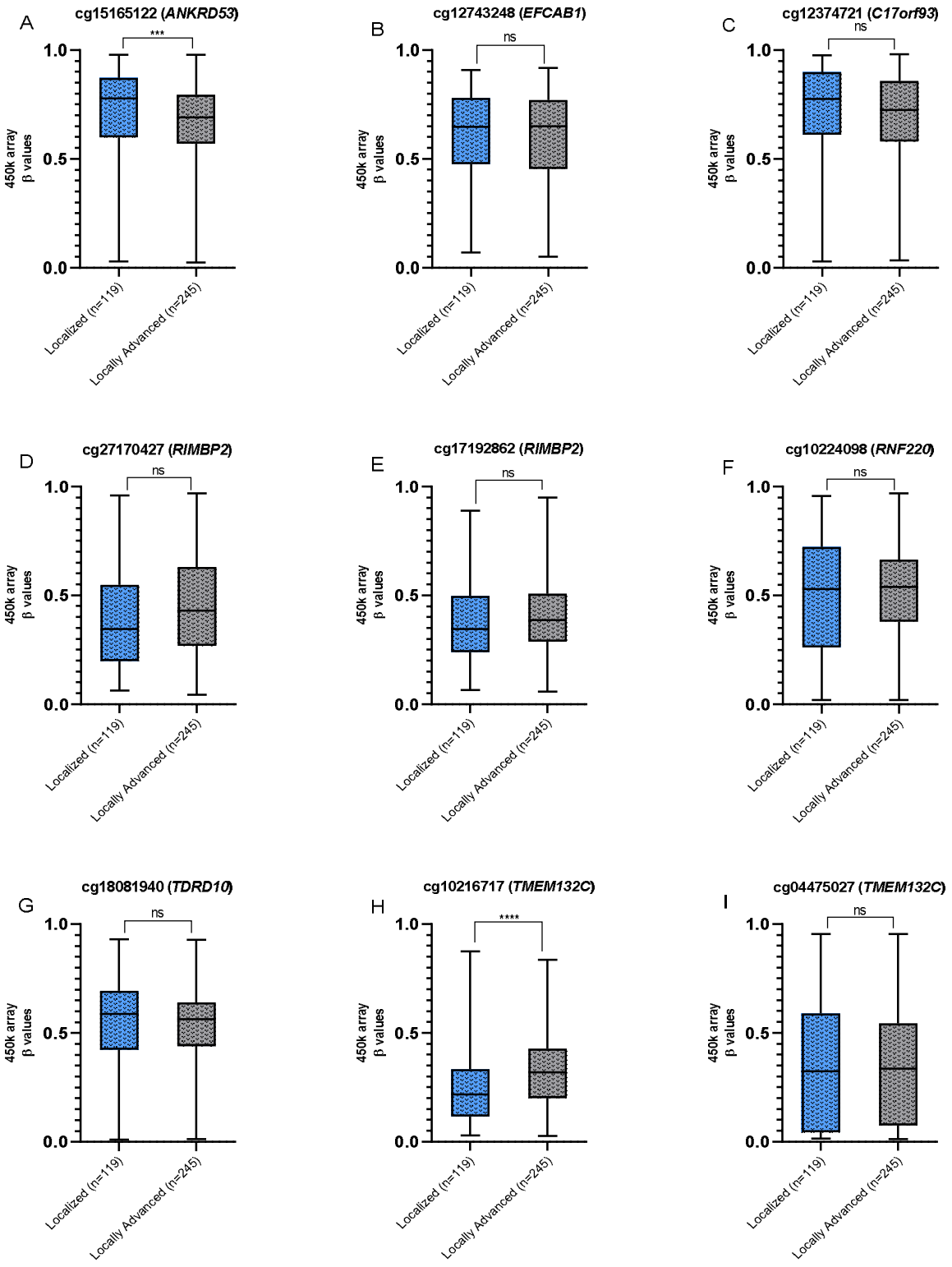
Supplementary Table 1. TCGA Bladder Cancer (BLCA) cohort clinicopathological characteristics

Characteristics	TCGA BLCA Cohort (n=414)
Normal bladder tissue	n=17(4%)
Primary bladder tumor tissue	n=397(96%)
Mean age at diagnosis (IQR)	68(16)
Male	67(15)
Female	70(17)
Sex	
Male	291(73%)
Female	106(27%)
Pathological T stage	
≤pT2	119
>pT2	245
Total	364
Pathological N stage	
pN0	231
pN+	125
Total	356
M stage	
M0	194
M1	11
Total	205
LVI	
Yes	146
No	126
Total	272
Smoking Habits	
Non-Smoker	107(28%)
Former or Current Smoker	277(72%)
Total	384
PFS (months)	
0	228(57%)
1	169(43%)
Mean PFI time (IQR) max, min	59(49)1-413
OS (months)	
0	223(56%)
1	174(44%)
Mean OS time (IQR) max-min	68(55)1-420
Mean months to death (IQR)	47(36)

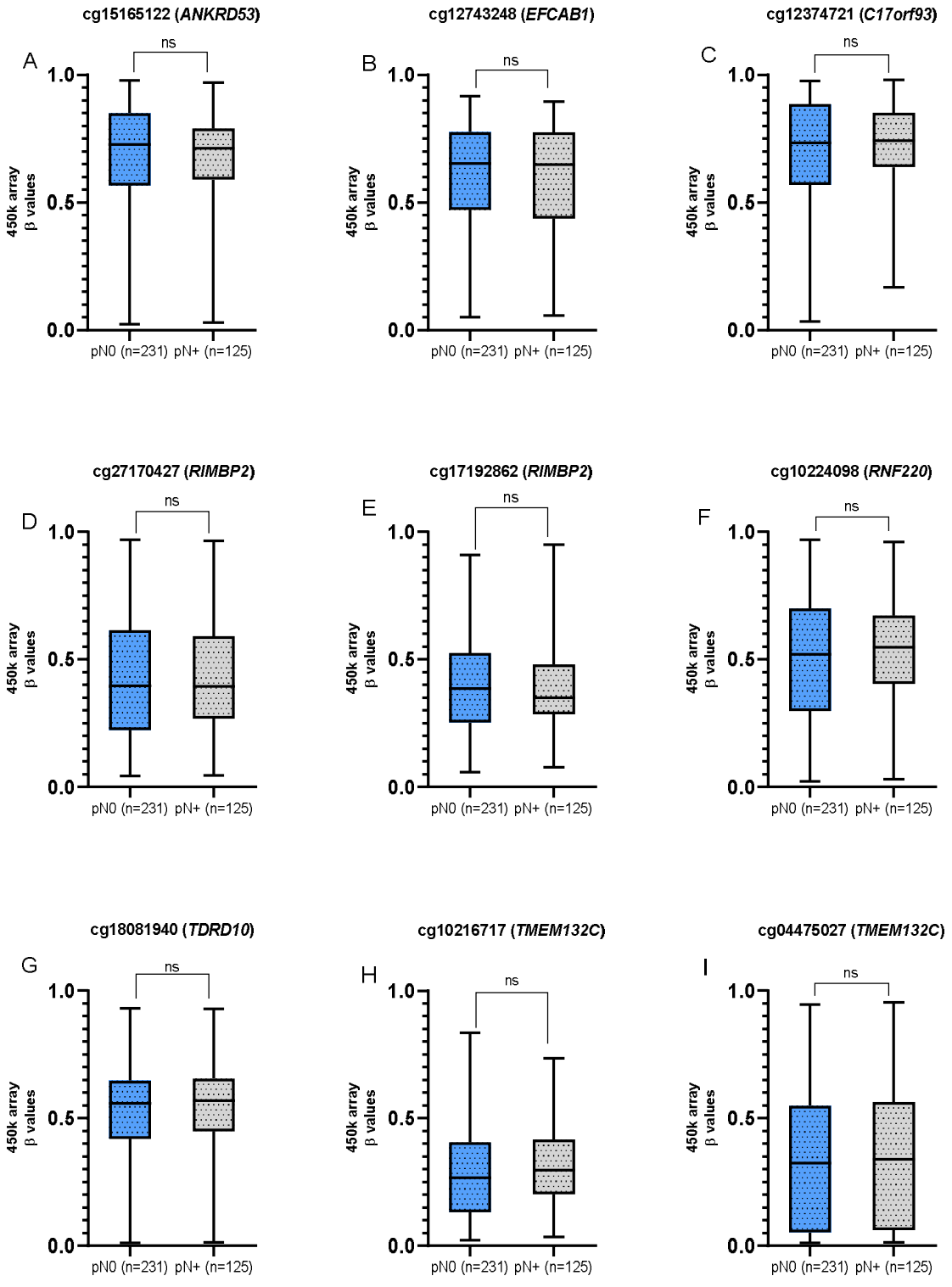
IQR - interquartile range; NA – non-attributable; LVI – lymphovascular invasion; PFS – progression free survival; OS – overall survival.



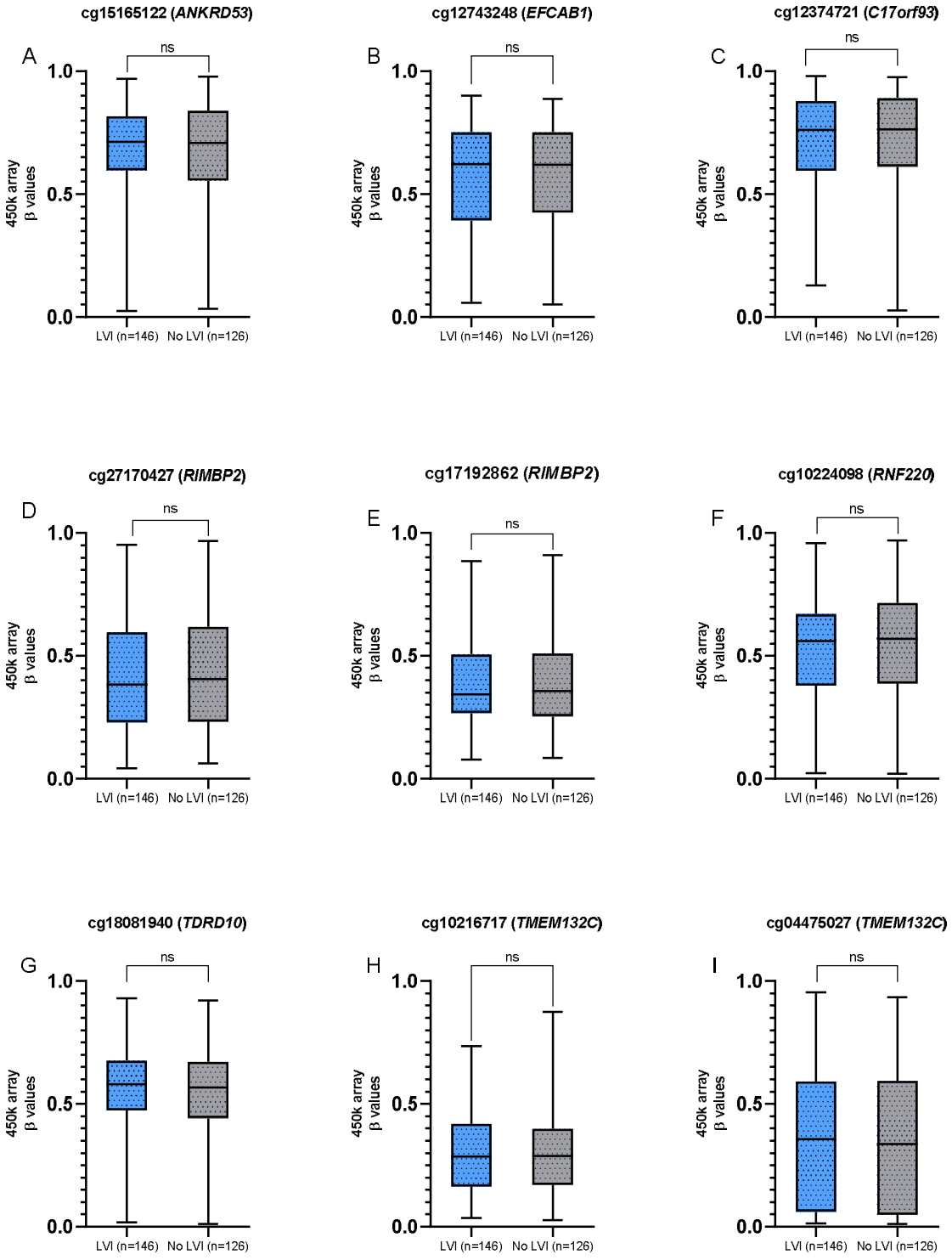
Supplementary Figure 1. Urothelial Bladder Cancer (UBC) methylation levels, according to smoking habits. Non-smokers refers to patients who have smoked less than 100 cigarettes in their lifetime, while Former and Current Smokers corresponds to patients who have had a different kind of smoking exposure during their lives. CG site cg10216717 (TMEM132C gene) was the only one achieving statistical significance. Methylation levels are expressed in β-values, ranging from 0 to 1. (ns, non significant; ****, $p \leq 0.0001$).



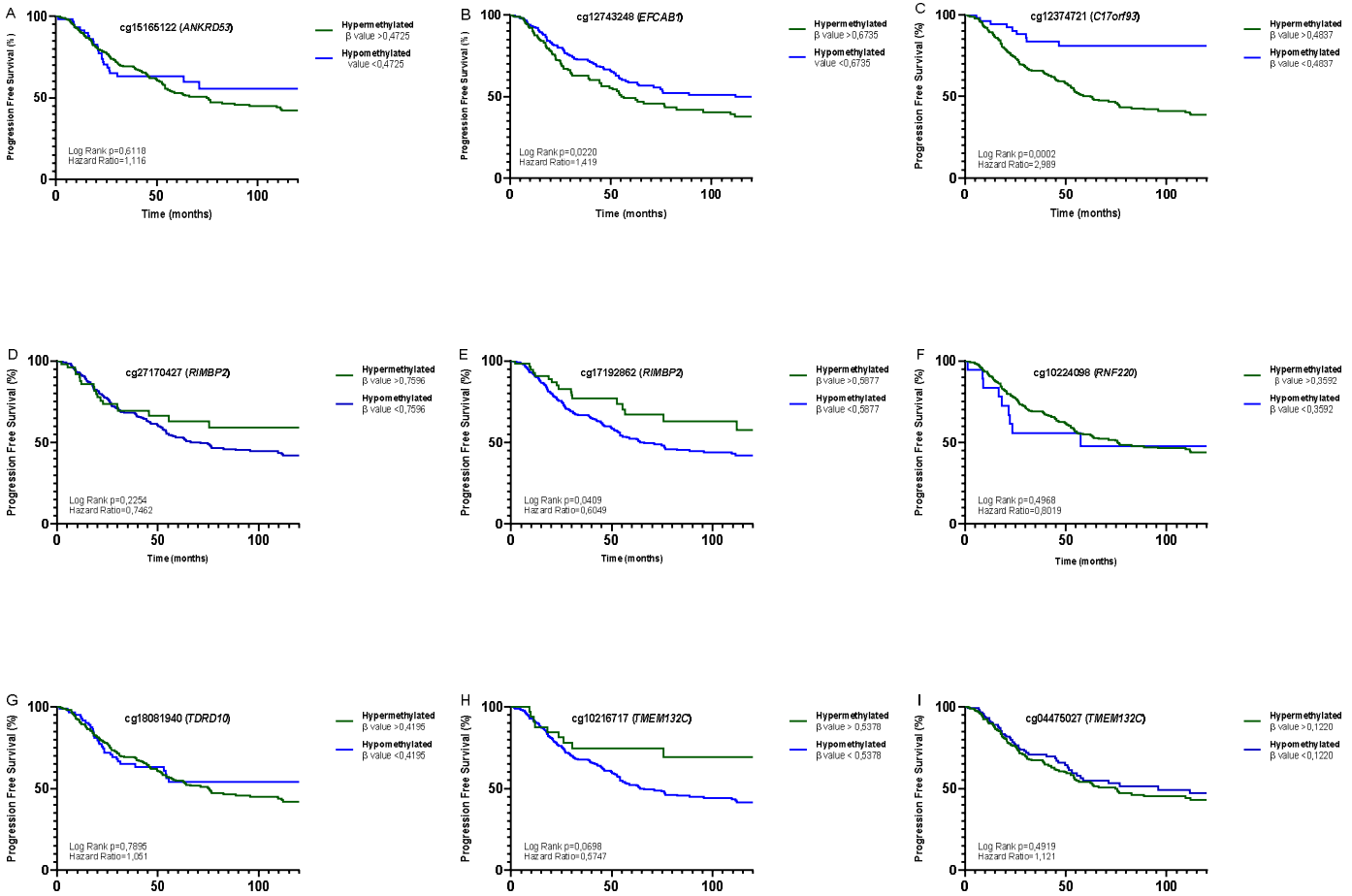
Supplementary Figure 2. Urothelial Bladder Cancer (UBC) methylation levels according to pathological T stage. pT2 disease stands for localized disease, while pT3 and pT4 stages stand for locally advanced disease. Methylation levels are expressed in β values, ranging from 0 to 1. (ns, non significant; ***, $p \leq 0.001$; ****, $p \leq 0.0001$).



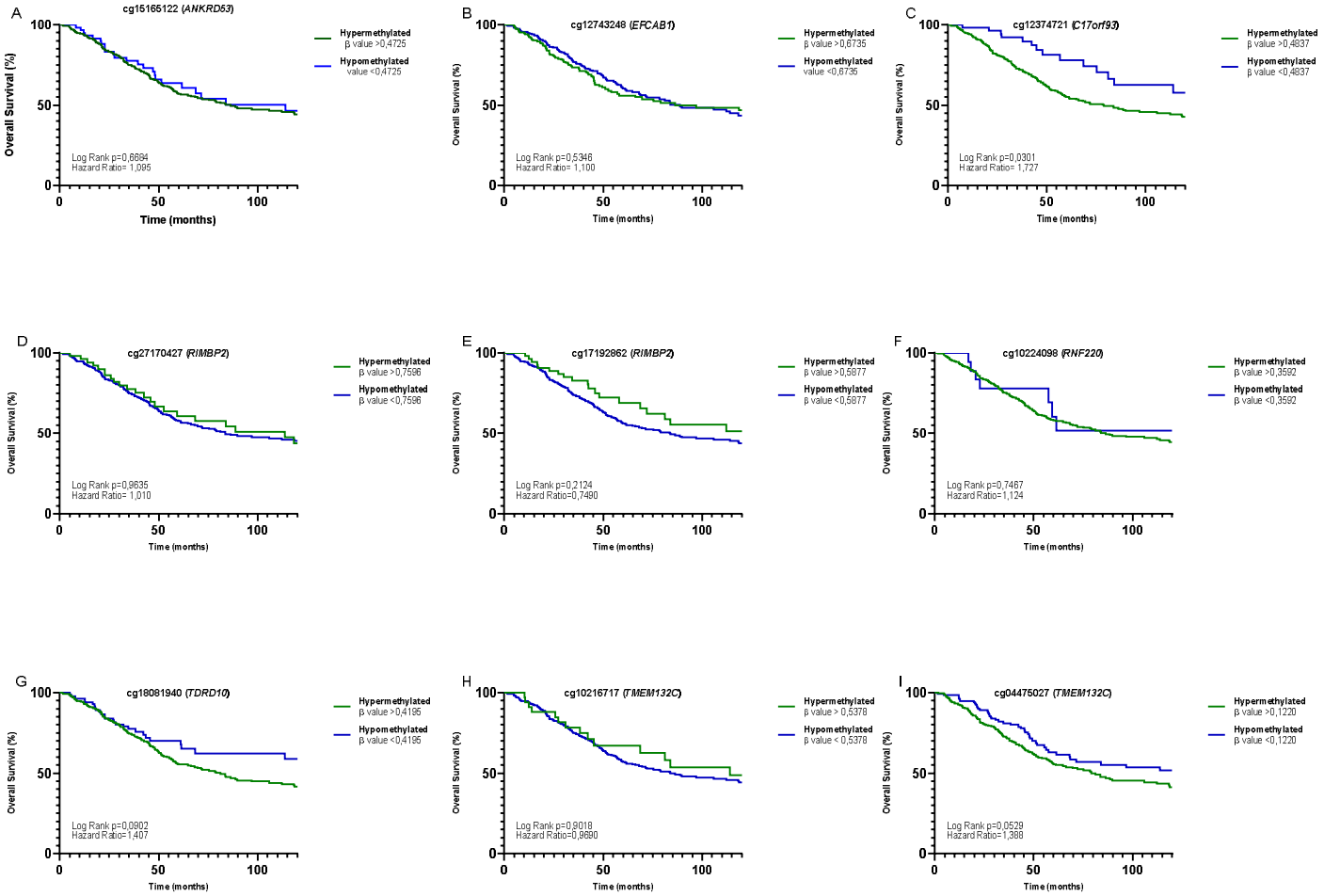
Supplementary Figure 3. Urothelial Bladder Cancer (UBC) methylation levels according to pathological N stage. Methylation levels are expressed in β values, ranging from 0 to 1. pN0=no lymph node involvement, pN+=lymph node involvement (ns, non significant).



Supplementary Figure 4. Urothelial Bladder Cancer (UBC) methylation levels according to lymphovascular invasion. Methylation levels are expressed in β values, ranging from 0 to 1 (*ns,* non significant).



Supplementary Figure 5. Progression Free Survival Analysis of CpG methylation profile in patients with UBC. PFS Kaplan-Meyer curves in hyper and hypomethylated subgroups (based on Youden Index calculated methylation cut-off values).



Supplementary Figure 6. Survival analysis of CpG methylation profile in patients with UBC. Kaplan-Meier analysis for OS in hyper and hypomethylated subgroups (based on Youden index calculated methylation cut-off values).

Supplementary Table 2. AUC and threshold methylation β-values based on Youden Index for prognostic model analysis and gene expression-methylation analysis.

Gene	CpG site	AUC (95%CI)	p value	Sensitivity(%) (95%CI)	Specificity(%) (95%CI)	Youden Index	Threshold β-value
ANKRD53	**cg15165122**	0.9007 (0.8614-0.9401)	<0.0001	84.89 (81.03-88.07)	94.12 (73.02-99.70)	79.01	> 0.4725
EFCAB1	**cg12743248**	0.7965 (0.7243-0.8687)	<0.0001	57.68 (52.77-62.45)	100 (81.57-100.0)	57.68	< 0.6735
C17orf93	**cg12374721**	0.9227 (0.8892-0.9561)	<0.0001	85.89 (82.13-88.98)	94.12 (73.02-99.70)	80.01	> 0.4837
RIMBP2	**cg27170427**	0.9393 (0.9036-0.9749)	<0.0001	87.41 (83.78-90.31)	88.24 (65.66-97.91)	75.65	< 0.7596
	cg17192862	0.9534 (0.9297-0.9771)	<0.0001	86.15 (82.40-89.20)	100 (81.57-100.0)	86.15	< 0.5877
RNF220	**cg10224098**	0.8546 (0.8004-0.9089)	<0.0001	73.8 (69.26-77.89)	94.12 (73.02-99.70)	67.92	> 0.3592
TDRD10	**cg18081940**	0.8696 (0.8217-0.9175)	<0.0001	77.83 (73.49-81.64)	94.12 (73.02-99.70)	71.95	> 0.4195
TMEM132C	**cg10216717**	0.9852 (0.9723-0.9981)	<0.0001	91.18 (87.99-93.59)	100 (81.57-100.0)	91.18	< 0.5378
	cg04475027	0.7267 (0.6555-0.7979)	0.0015	68.77 (64.05-73.13)	82.35 (58.97-93.81)	51.12	> 0.1220