Predicting response to cisplatin-based neoadjuvant chemotherapy for muscle-invasive bladder cancer: transcriptomic features outrank genomic biomarkers

Ariadna Acedo-Terrades[1], Alejo Rodriguez-Vida[1,2], Oscar Buisan[3], Marta Bódalo-Torruella[1], Maria Gabarrós[1], Miquel Clarós[1], Nuria Juanpere[4], Marta Lorenzo[4], Sergio Vázquez Montes de Oca[1], Alejandro Rios-Hoyo[2], Cristina Carrato Moñino[5], Tamara Sanhueza[5], Eduardo Eyras[1,6], Eulàlia Puigdecanet[7], Gottfrid Sjödahl[8], Júlia Perera-Bel[1#], Lara Nonell[9#] & Joaquim Bellmunt[1,10#]

[1]Hospital del Mar Research Institute (HMRI), Barcelona, Spain.

[2]Medical Oncology Department, Hospital del Mar, Barcelona, Spain.

[3]Urology Department, Hospital de Bellvitge, Barcelona, Spain.

[4]Department of Pathology, Hospital del Mar, Barcelona, Spain.

[5]Urology Department, University Hospital Germans Trias i Pujol, Badalona, Spain.

[6]EMBL Australia Partner Laboratory Network at the Australian National University, Canberra Australian Capital Territory, Australia.

[7]IRIS-CC & MoD Lab Research Group. Faculty of Medicine, UVic-UCC, Vic, Spain.

[8]Division of Urological Research, Department of Translational Medicine, Lund University and Department of Urology, Malmö, Sweden.

[9]Bioinformatics Unit, Vall d'Hebron Institute of Oncology, Barcelona, Spain.

[10]Dana Farber Cancer Institute, Harvard Medical School Boston Massachusetts, USA.

#co-corresponding authors

ABSTRACT

Muscle-invasive bladder cancer (MIBC) is associated with poor predictability of response to cisplatin-based neoadjuvant chemotherapy (NAC). Consequently, the benefit of NAC remains unclear for many patients due to the lack of reliable biomarkers predicting treatment response. In order to identify biomarkers and build an integrated and highly accurate model to predict NAC response, we performed a comprehensive transcriptomic and genomic profiling on tumors from 100 MIBC patients. Our results showed that the expression of the top genes associated with response, as well as the expression of growth factor genes and cell cycle regulators are highly correlated with NAC response. Most importantly, we found a novel signature related to the WNT signaling pathway that alone was highly correlated with NAC response and showed high accuracy in predicting NAC response (AUC=0.76). Additionally, mutations in the DNAH family genes (DNAH8, DNAH6 and DNAH10) and deletion in KDM6A were also highly correlated with NAC response. Using our comprehensive molecular analysis as a backbone, we developed two machine learning (ML) models, one incorporating both transcriptomic and genomic features (RF-RW), and the other using only transcriptomic data (RF-R). Both models demonstrated promising performance (AUC=0.82) as predictive models of response to NAC in MIBC. RF-RW and RF-R, after external validation, could potentially change the management of MIBC patients by selecting ideal candidates for NAC.

INTRODUCTION

Bladder cancer (BC) is a complex disease with a huge impact in society being the 10th most common cancer worldwide and the 6th most common cancer in men[1]. The pathologic tumor stage in localized BC could be separated in two main groups according to the depth of invasion into the muscularis propria layer: non-muscle invasive bladder cancer (NMIBC), which includes Tis, Ta and T1, and muscle-invasive bladder cancer (MIBC) tumors, which includes T2, T3 and T4a.

Approximately 20% of new BC patients are diagnosed with MIBC, a clinical setting which is more aggressive than NMIBC and is associated with a poor 5-year survival rate[2-4].

Over the last two decades, the standard-of-care for MIBC has been to administer cisplatin-based neoadjuvant chemotherapy (NAC) followed by radical cystectomy (RC). Despite NAC having demonstrated a benefit in overall survival (OS), less than 50% of patients are actually eligible due to comorbidities of frailty, and many of them proceed to RC[5-7]. Moreover, among patients who are treated with NAC, the percentage of non-responders (NR) is around 40%, with these patients having high recurrence rate and poor 5-year survival rates of around 60%[2,3]. Furthermore, despite the great efforts made in recent years, there is still an absolute lack of reliable biomarkers to predict benefit from NAC[3,8]. Consequently, the inability to predict NAC response is a significant barrier to adequately identify those patients who might benefit from it. In turn, this prevents non-responders from avoiding unnecessary toxicity and allows for the provision of alternative treatment options. Identifying and validating integrated and accurate predictive models of NAC is therefore a highly unmet need in MIBC.

BC is associated with a high molecular diversity, and several transcriptomic subtypes have been described[9-16]. The most accepted molecular subtypes were described in the TCGA BC study, in which Robertson et al. proposed five MIBC transcriptomic subtypes (TCGAclas): basal-squamous, luminal papillary, luminal infiltrated, luminal and neuronal[12]. In a parallel study, the Lund taxonomy (LundTax) has been described as an alternative subtype classification: Urothelial-like (Uro), which includes UroA, UroB and UroC, Genomically Unstable (GU), Basal/Squamous (BaSq), Mesenchymal-like (MES-like) and

Small-cell/Neuroendocrine (ScNE)[13]. Despite differing in the quantity, proportion and characteristics of the subtypes, both subtype classifications show high overlap[14].

Some studies suggest that MIBC molecular subtypes predict NAC response, and can be useful for treatment selection and improving outcomes[17]. However, these findings have been demonstrated to be inconsistent over time. First, Choi et al. reported that both basal and luminal were the subtypes showing the greatest benefit from NAC, whereas p53-like tumors are less likely to respond to NAC[10]. On the other hand, Seiler et al. confirmed that basal tumors had the highest benefit of NAC, but conversely indicated that luminal papillary subtype was associated with a low likelihood of benefit from NAC[18]. Additionally, Kamoun et al. highlighted that the luminal infiltrated subtype appeared to be resistant to cisplatin-based therapy but instead particularly sensitive to immune checkpoint inhibitors (ICI)[14]. Moreover, Lotan et al. indicated that non-luminal tumors received the greatest benefit from NAC, while luminal tumors experienced only a minimal survival benefit[19]. Using a different classification system, the opposite finding was reported by Sjödahl et al. demonstrating that luminal-like subtypes are the ones that more frequently respond to NAC treatment[20]. This lack of consistent results therefore requires uniform validation studies to establish the clinical relevance of molecular subtyping in MIBC as predictive biomarkers of NAC[21].

At the genomic level, MIBC is characterized by genomic instability and a high mutation rate[12,14,22]. The most commonly mutated genes in MIBC are TP53 (48%), KMT2D (28%) and KDM6A (26%). The most common somatic copy number alterations (SCNAs) include the amplification of E2F3, PPARG and MDM2 and deletions in CDKN2A and RB1. Importantly, tumor mutational burden (TMB) and APOBEC-mediated mutagenesis have been associated with OS[12]. However, none of these genomic alterations have been validated as potential predictive biomarkers of NAC in MIBC.

In our present study, we have conducted a comprehensive molecular profiling using RNA sequencing (RNA-Seq) and whole exome sequencing (WES) obtained from a cohort of 100 MIBC patients treated with NAC across different hospitals in Catalonia, Spain. Our primary objective was to characterize the transcriptomic and genomic landscape of MIBC in relation to NAC response. Our secondary objective was to develop a machine learning (ML) model capable of predicting the response to NAC treatment. The ultimate aim of our study was to create a predictive tool which could be potentially translated into daily practice for selecting ideal candidates for NAC and consequently change the management of MIBC.

RESULTS

Patient characteristics and genomic analysis

The cohort of this study, included a total of 100 patients with MIBC retrospectively collected from four Catalan hospitals from 2010 until 2019. All patients received cisplatin-based neoadjuvant chemotherapy (NAC), and were divided into responders (R; n=53) and non-responders (NR; n=47). Response was defined as postsurgical downstaging to non-MIBC (<=ypT1) with no pathological lymph-node involvement (ypN0) observed at cystectomy (Fig. 1A). A detailed summary of clinical and histopathological information is provided in Table1 (and Supplementary Table 1).

The median age of the 100 patients (10 females and 90 males) at diagnosis was 69 years (64;75) (Table 1). No differences were observed in the demographic variables between R and NR (Table 1, Supplementary Table 1). Since these patients were treated before the approval of adjuvant nivolumab in Spain[23], none of them received adjuvant immune checkpoint blockade. There was no difference in the use of additional treatments (prior intravesical therapy, adjuvant chemotherapy) between R and NR (Table 1). The similarity between the administration of complementary treatments to NAC suggests a consistent approach in the treatment management for R and NR, which is crucial for ensuring unbiased results in further analyses.

The median follow-up was 27 months (IQR 13;112 months); during this time, 1 R and 10 NR patients had local pelvic recurrence, 5 R and 23 NR distant metastatic recurrence and 8 R and 30 NR died. Response groups showed significant differences on 5-year OS (Fig. 1B).

Previously reported classic prognostic pathological features such as urethral involvement, lymphovascular invasion and surgical margins showed significant overrepresentation in NR (p=7.16e-04, p=9.87e-09, p=0.004, respectively) (Table 1). Local and metastatic recurrence were also significantly more common among NR, indicating response to NAC is associated with a lower likelihood of recurrence (p=0.006, p=3.07e-05). On the other hand, leukocytes, neutrophils and lymphocytes counts did not show significant differences between R and NR (Table 1), suggesting similar immune cell profiles among MIBC patients. Additionally, the ratio between neutrophils and lymphocytes , previously reported as an indicator of good NAC response[24], was not significantly different between groups (Table 1).

	All patients	Responders	Non-responders	P overall	Sig
	N=100	N=53	N=47		
Age (median, range)	70 [64;75.25]	69 [64;75]	70 [64;75]	0,906	
Sex (n, %)				1	
Male	90 (90.0%)	48 (90.6%)	42 (89.4%)		
Female	10 (10.0%)	5 (9.43%)	5 (10.6%)		
Prior intravesical treatment (n, %)				0,465	
None	83 (83.0%)	45 (84.9%)	37 (78.7%)		
Mitomycin	8 (8.00%)	3 (5.66%)	5 (10.6%)		
Bacillus Calmette-Guerin	7 (7.00%)	4 (7.55%)	3 (6.38%)		
Others	2 (2.00%)	0 (0.00%)	2 (4.26%)		
NA	1 (1.00%)	1 (1.89%)	0 (0.00%)		
Hydronephrosis at baseline (n, %)				0,298	
Yes	25 (25.0%)	16 (30.2%)	9 (19.1%)		
No	75 (75.0%)	37 (69.8%)	38 (80.9%)		
Number of cycles of neoadjuvant chemotherapy (n, %)				0,794	
1-3	83 (83.0%)	43 (81.1%)	40 (85.1%)		
>=4	17 (17.0%)	10 (18.9%)	7 (14.9%)		
Leukocytes (Median, IQR)	7435 [5968;8958]	7800 [5760;8900]	7410 [6275;8965]	0,764	
Neutrophils (Median, IQR)	4375 [3300;5910]	4550 [3200;5900]	4300 [3585;5920]	0,656	
Lymphocytes (Median, IQR)	1700 (1292;2200)	1710 [1209;2180]	1600 [1400;2280]	0,868	
Ratio neutrophils/lymphocytes (Median, IQR)	2.42 [1.82;3.77]	2.44 [1.86;3.14]	2.40 [1.76;4.29]	0,882	
Tumor Stage at TURB (n, %)				0,811	
T1	5 (5.00%)	3 (6.38%)	2 (3.77%)		
T2	90 (90.0%)	41 (87.2%)	49 (92.5%)		
T3	4 (4.00%)	2 (4.26%)	2 (3.77%)		
T4	1 (1.00%)	1 (2.13%)	0 (0.00%)		
ypT at cystectomy (n, %)				0	****
ypT0	36 (36.0%)	0 (0.00%)	36 (67.9%)		
ypTis	11 (11.0%)	0 (0.00%)	11 (20.8%)		
ypTa	2 (2.00%)	0 (0.00%)	2 (3.77%)		
ypT1	4 (4.00%)	0 (0.00%)	4 (7.55%)		
ypT2	20 (20.0%)	20 (42.6%)	0 (0.00%)		
ypT3	23 (23.0%)	23 (48.9%)	0 (0.00%)		
ypT4	4 (4.00%)	4 (8.51%)	0 (0.00%)		
ypN Cystectomy (n, %)				6,48E-10	****
N0	77 (77.0%)	24 (51.1%)	53 (100%)		
N1	7 (7.00%)	7 (14.9%)	0 (0.00%)		
N2	16 (16.0%)	16 (34.0%)	0 (0.00%)		
Urethral involvement (n, %)				7,16E-04	***
Yes	9 (9.00%)	0 (0.00%)	9 (19.1%)		
No	91 (91.0%)	53 (100%)	38 (80.9%)		
Lymphovascular invasion (n, %)				9,87E-09	****
Yes	24 (24.0%)	0 (0.00%)	24 (51.1%)		
No	76 (76.0%)	53 (100%)	23 (48.9%)		
Surgical margins (n, %)				0,004	***
Positive	7 (7.00%)	0 (0.00%)	7 (14.9%)		
Negative	93 (93.0%)	53 (100%)	40 (85.1%)		
Number of resected lymph nodes at cystectomy (n, %)				0,307	
1-9	32 (32.0%)	16 (30.2%)	17 (36.2%)		
10-19	58 (58.0%)	34 (64.2%)	24 (51.1%)		
20-26	9 (9.00%)	3 (5.66%)	6 (12.8%)		
Number of positive lymph nodes at cystectomy (n, %)				9,87E-09	****
1 or more	24 (24.0%)	0 (0.00%)	24 (51.1%)		
0	76 (76.0%)	53 (100%)	23 (48.9%)		
Adjuvant chemotherapy (n, %)				0,403	
No	69 (69.0%)	39 (73.6%)	30 (63.8%)		
Yes	31 (31.0%)	14 (26.4%)	17 (36.2%)		
Local recurrence (n, %)				0,006	**
Yes	11 (11.0%)	1 (1.89%)	10 (21.3%)		
No	89 (89.0%)	52 (98.1%)	37 (78.7%)		
Metastatic recurrence (n, %)				3,07E-05	****
Yes	28 (28.0%)	5 (9.43%)	23 (48.9%)		
No	72 (72.0%)	48 (90.6%)	24 (51.1%)		

Table 1. Overview of the clinical characteristics of the MIBC cohort. *(Significancy: p<0.05 = * , p<0.01 = **, p<0.001 = ***, p<0.0001 = ****, p<0.00001 = *****)*

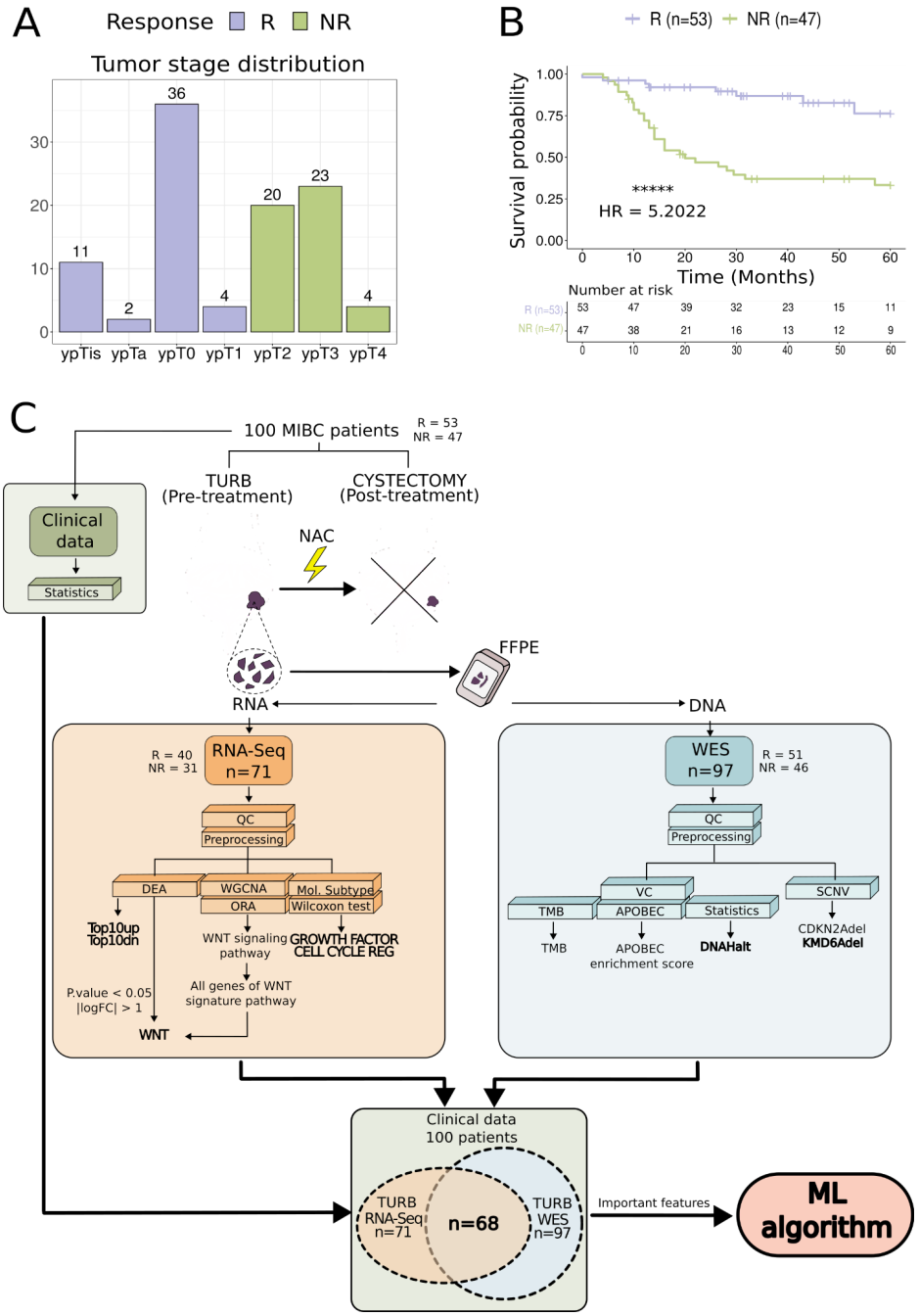

Figure 1. Overview of the MIBC cohort and analysis workflow. A. *Response and tumor stage distribution across the cohort of 100 MIBC patients. Response was defined as dowstaging to non-MIBC status with no pathological lymph-node involvement (<=T1N0) observed at cystectomy.* ***B.*** *Kaplan-Meier survival curve for 5-year overall survival (OS) in MIBC patients stratified by response (p<0.00001; HR=5.20).* ***C.*** *Design of the study. A total number of 100 MIBC patients were retrospectively included in the study: 53 responders (R) and 47 non-responders (NR). All patients received cisplatin-based neoadjuvant chemotherapy. We generated RNA sequencing (RNA-Seq) (n=71; R=40, NR=31) and whole-exome sequencing (WES) (n=97; R=51, NR=46) data from tumor samples at transurethral resection of bladder tumor (TURB). Both analyses followed a first quality control (QC) and were subsequently subjected to a preprocessing step. RNA-Seq data was used to perform different analyses: differential expression analysis (DEA), where two signatures were obtained (Top10up, Top10dn), weighted gene correlation network analysis (WGCNA) followed by an over-representation analysis (ORA) of the module which showed significant correlation with response.*

WNT signaling pathway was over-represented in the ORA results and by intersecting DE results with WNT-related genes we obtained a WNT signature associated with NAC response. Finally, we performed molecular subtype classification using two classifiers (LundTax and TCGAclas). Two of the signatures used to classify MIBC samples in LundTax were significantly associated with NAC response (GROWTH FACTOR and CELL CYCLE REG). WES data was used to perform a variant calling (VC) and somatic copy number variation (SCNV). Following the VC, tumor mutational burden (TMB) and APOBEC scores were obtained but they were not significantly associated with response. However, comparison between R and NR showed that mutations in some DNAH family genes (DNAH8, DNAH6 and DNAH10) were significantly correlated with response. Additionally, SCNV showed that deletion in KMD6A was strongly associated with response. Features obtained from RNA-Seq and WES analyses were used to construct a machine learning (ML) model to predict NAC response using data from 68 patients with both data types.

We generated high-quality RNA sequencing (RNA-Seq) data from 71 patients (R=40; NR=31) and whole-exome sequencing (WES) data from 97 (R=51; NR=46) from tumor samples collected at transurethral resection of bladder tumor (TURB) before starting NAC. High-quality sequencing data was obtained for all samples. To ensure the differences in the analyses were not due to initial tumor quality, we confirmed that tumoral percentage and quantity were not statistically different between R and NR. (Supplementary Table 1). Importantly, no differences in tumoral percentage and quantity were found between R and NR. RNA-Seq data provided insights into gene expression patterns, molecular subtypes and gene signatures associated with NAC response. The WES data was used to investigate the impact of specific mutations, tumor mutational burden (TMB) and mutational signatures on NAC response. We also studied the role of somatic copy number variations (SCNVs). Finally, we integrated the information extracted from RNA-Seq and WES data to create a ML model to predict NAC response in MIBC patients (Fig. 1C).

WNT gene expression signature is associated to lack of response to NAC

The differential expression analysis between R (n=40) and NR (n=31) from our cohort of MIBC pre-treatment samples identified 602 differentially expressed (DE) genes. Among these, a higher number were associated with NR (n=588) compared to a minority that were associated with R (n=14) (Fig. 2A, Supplementary Table 2). To better understand the possible implications for the prediction of response to NAC, and to overcome the lack of adjusted p-values in the previous analysis, we applied the weighted correlation network analysis (WGCNA) method, an unsupervised approach to identify clusters of correlated genes. We identified 14 gene clusters, one of which showed a statistically significant negative association with NAC response (p=0.008, cor=-0.32) (Supplementary Fig. 1A). Genes from this cluster were enriched in several gene sets related to the WNT signaling pathway, cell signaling, angiogenesis and proliferation (Fig. 2B). Interestingly, we found 17

WNT signaling pathway genes among the list of DE genes (SFRP2, FGF10, MIR145, SOX7, ATP6V0C, FOLR1, KREMEN2, TRABD2B, WNT9A, WNT2, AXIN2, PPM1N, SOX2, DLX5, LGR6, LBX2, HESX1) (Fig. 2B).

Analyzing the expression patterns of combinations of genes instead of individual genes can sometimes provide a clearer signal. We created signature scores for the top 10 up (Top10up) and down (Top10dn) regulated genes from the DE analysis. As expected, these signature scores showed significant differences between R and NR (p=0.0012,p=0.0002; respectively) (Fig. 2C). Importantly, the gene signature score derived from the 17 WNT signaling pathway genes also showed a significant difference between R and NR (p=6.2e-07) (Fig. 2C). Beyond the association with NAC response, all three signatures were also capable of stratifying survival outcomes; high expression of the Top10up was associated with better OS (HR=0.38[0.15-0.92]), whereas high expression of Top10dn and WNT signatures were significantly associated with worse OS (HR=3.24[1.42-7.69] and HR=4.53[1.92-11.1], respectively)(Fig. 2C).

Molecular subtype signatures provide further insights into response to NAC

We applied the molecular subtype classification from Robertson et al., 2017. TCGAclas classified the samples as Basal squamous (n=18; R=11, NR=7), Luminal (n=12; R=4, NR=8), Luminal infiltrated (n=13; R=6, NR=7) and Luminal papillary (n=28; R=19, NR=9). Although no significant differences were found between R and NR patients, we observed a trend towards a better response to NAC in luminal papillary and basal squamous subtype (68% and 61% of responders, respectively), compared to luminal (33% responders) and luminal infiltrated (46% responders) (Supplementary Fig. 2A, Supplementary Table 3).

We also applied the molecular subtype classification from LundTax and identified the largest group of samples belonging to the Uro group (n=44, 62%); including the different subgroups of: UroA (n=28), UroB (n=7) and UroC (n=9). 13 samples (18.3%) were classified as BaSq, 12 (16.9%) as GU and 2 as ScNE (2.8%) (Supplementary Fig. 2B). These proportions are in line with the original classification.

Unfortunately, none of Lundtax subtypes showed a significant association with response (Supplementary Table 3), but proportions were in line with what has been described in the literature: patients classified as UroB (3 R vs 4 NR) and BaSq (5 R vs 8 NR) showed a low proportion of responders (approximately 40%), indicating a lower likelihood of responding to NAC. On the contrary, GU (8 R vs 4 NR) and UroA (17 R vs 11 NR) patients showed an

opposite trend, with more than 60% of responders being the 2 subtypes most likely to benefit from NAC (Supplementary Fig. 2B). In addition, whereas luminal papillary in TCGAclas is consistent with the findings in the UroA LundTax subtype, BaSq results contradict our previous findings. Probably, the BaSq TCGAclas subtype includes patients classified as GU in LundTax, which may increase the proportion of responders in this TCGAclas subtype.

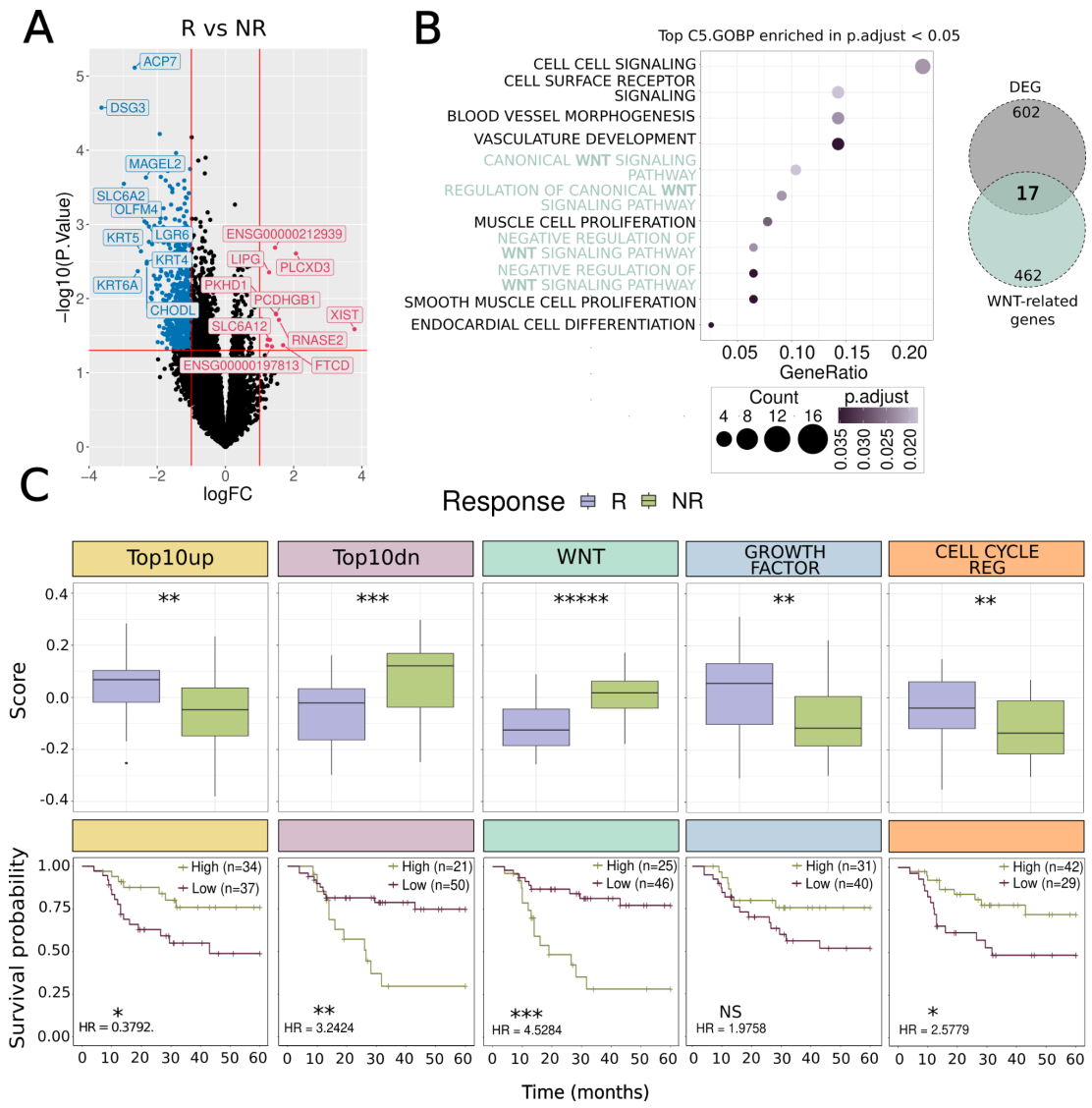

Figure 2. Gene expression signatures are strongly associated with response and 5-year overall survival. A. *Volcano plot showing the top 10 up-regulated (Top10up) and the top 10 down-regulated (Top10dn) genes obtained after differential expression analysis between responders (R) and non-responders (NR).* ***B.*** *Top significantly enriched pathways (p.adj<0.05) in the over-representation analysis (ORA) and 17 genes (WNT signature) rehesulting from the intersection between DE genes (602) and WNT-related genes (462).* ***C.*** *First row: Boxplot of the Top10up (p=0.0012), Top10dn (p=0.0002), WNT (p=6.2e-07), GROWTH FACTOR (p=0.0067) and CELL CYCLE REG (p=0.0056) signatures score and its significant correlation with NAC response. Second row: Kaplan-Meier curves showing the significant association between Top10up (p=0.0267, HR=0.38[0.15-0.92]), Top10dn (p=0.003, HR=3.24[1.42-7.69]), WNT (p=2e-04, HR=4.53[1.92-11.1]), GROWTH FACTOR (p>0.05, HR=1.98[0.81-4.80]) and CELL CYCLE REG (p=0.0217, HR=2.58[1.14-5.96]) signatures score and 5-year overall survival (OS).*

In order to understand the response mechanisms within the molecular subtypes, we derived single sample scores of the individual signatures used by the LundTax classifier (Supplementary Figure 1B). Interestingly, signatures of cell cycle regulation and growth factor (CELL CYCLE REG including FGFR3, CCND1, E2F3, RB1 and CDKN2A genes[25], and GROWTH FACTOR including EGFR, ERBB2 and ERBB3[16] showed the most highly significant differences between R and NR, being positively associated with response and OS. However, the association between GROWTH FACTOR and OS is not significant (Fig. 2C).

The combination of gene expression signatures stratifies response and overall survival

To assess the potential of these signatures as biomarkers for a predictive model, we studied the correlation between the genes from all the identified signatures. Correlation analysis revealed only three genes from the Top10dn signature with high correlation coefficients (KRT6A-KRT5=0.95, KRT6A-DSG3=0.94, KRT5-DSG3=0.89, Fig. 3A). Both KRT6A and KRT5 belong to the keratin family of proteins, which are essential markers for basal epithelial cells. Additionally, while DSG3 is not a keratin family protein per se, it collaborates with other keratin family proteins to maintain the epithelial cells' structure. Besides that, there was an overall weak association among the genes from the identified signatures (|cor| < 0.60, Fig. 3A, Supplementary Table 4), suggesting a need to further explore their potential as combined biomarkers.

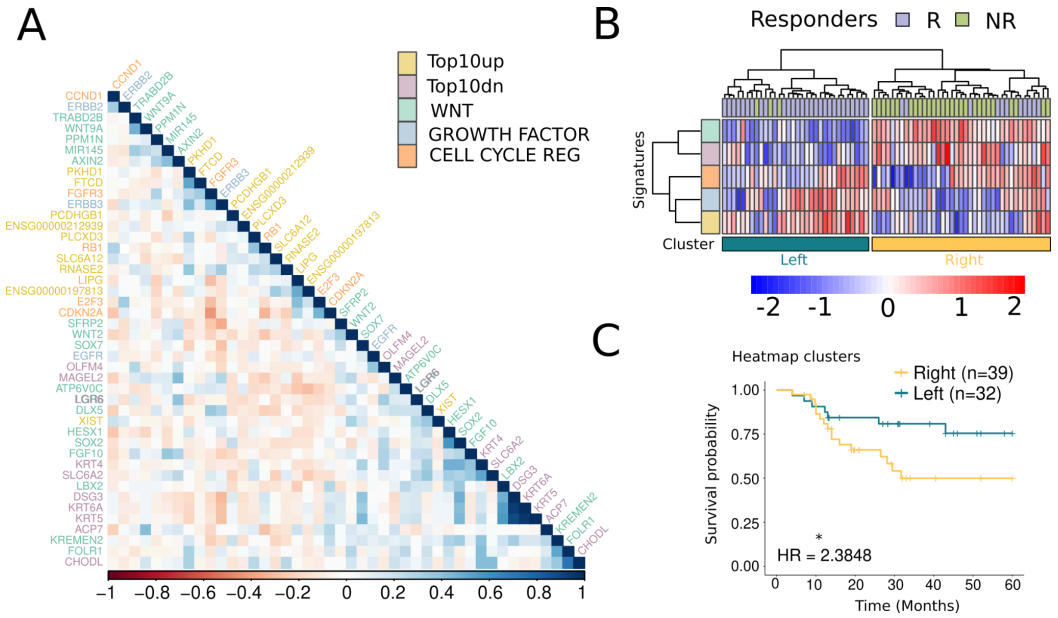

Figure 3. Gene expression signatures can distinguish between R and NR. A. *Correlation plot of the expression of genes constituting the RNA-Seq signatures.* ***B.*** *Heatmap showing the scores of*

RNA-Seq signatures across samples and their impact in response. ***C.*** *Kaplan-Meier showing the significant association between clusters obtained in heatmap and 5-year overall survival (OS) (p=0.049, HR=2.38[0.97-5.82]).*

The combination of the five signatures by unsupervised hierarchical clustering identified two main clusters of patients (Fig. 3B). The left cluster, which mainly contained R, was characterized by low expression of WNT and Top10dn signatures and overexpression of GROWTH FACTOR, CELL CYCLE REG and Top10up signatures. The right cluster contained the majority of NR and was characterized by an overexpression of WNT and Top10dn signatures. Within the later cluster, we could further distinguish a subcluster with high CELL CYCLE REG and GROWTH FACTOR signatures that could discriminate against the few responders with high WNT signature values. In line with the gene-wise correlation results, we observed that the signatures captured distinct relevant aspects associated with NAC response, harboring strong potential features for a predictive model. Similarly to the individual signatures, the combination of the five signatures led to the differentiation of two groups with significantly different 5-year OS (p=0.049, HR=2.3848[0.97-5.82]) (Fig. 3C).

Analysis of mutational landscape

Analysis of WES data identified a total of 27.404 somatic non-synonymous mutations in our cohort. The most frequently mutated genes were KIR2DL3 (54%) and LILRB3 (54%), followed by TP53 (51%), TTN (48%), RAMEF18 (46%), CYP2D6 (44%), MUC3A (41%), MUC5AC (41%), GTPBP6 (38%) and ZBED3 (38%) (Supplementary Fig. 3A).

The most frequent somatic mutations in 121 previously reported BC associated genes (Supplementary Table 5) were TP53, with 49 mutations (51%), followed by KMT2D (24%), ERCC2 (18%), EP300 (16%) and ATM (15%) (Fig. 4A). This frequency is consistent with previously reported in other MIBC cohorts[12,26,27], highlighting their role as BC drivers.

Cancer pathways that have been associated with BC tumorigenesis (e.g. DDR, CM, CRG, TP53) (Supplementary Table 6) were mutated in the majority of patients in our cohort. In contrast, the signatures obtained from RNA-Seq analysis (WNT, GROWTH FACTOR, CELL CYCLE REG, top10up, top10dn) were mutated in a small proportion (10-20%) of patients, being the GROWTH FACTOR and CELL CYCLE REG signatures the most frequently mutated. Overall, the mutational landscape at the pathway level remained similar across the entire cohort and it was independent of response (Fig. 4A).

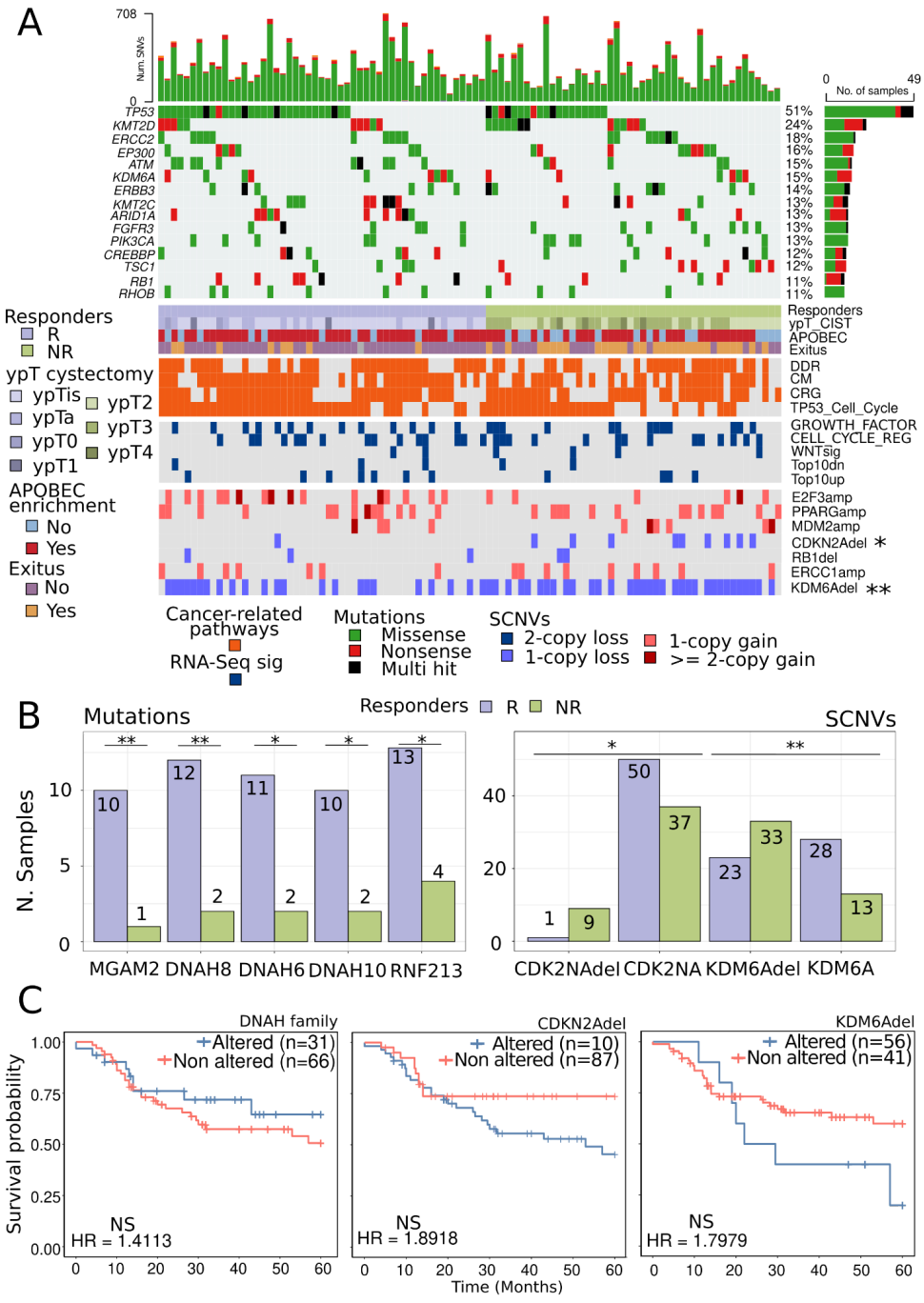

Figure 4: Mutations in DNAH family genes (DNAH8, DNAH6 and DNAH10) and deletions in CDK2NA and KDM6A are significantly correlated with NAC response. ***A.*** *Mutational landscape for MIBC patients with clinical information. Bladder cancer genes mutated in more than 10% of samples. Mutations in important pathways such as* DDR, CM, CRG and TP53CellCycle as well as RNA-Seq signatures were also included. *Somatic copy number variations (SCNVs) previously identified as important features were also visualized in the oncoplot.* ***B.*** *Boxplots of significantly different mutated genes associated with NAC response: MGAM2 (p=0.0086), DNAH8 (p=0.0086), DNAH6 (p=0.0163, DNAH10 (p=0.0301), RNF213 (p=0.0349) and deletions in CDK2NA (p=0.006) and KDM6A (p=0.014) and its significant association with NAC response.* ***C.*** *Kaplan-Meier curve*

showing the correlation of: mutations in any of the DNAH family genes (DNAH8, DNAH6 and DNAH10) (p>0.05) and deletions in CDKN2A (p>0.05) and KDM6A (p>0.05) with 5-year overall-survival.

Somatic mutations in DNAH8, DNAH6 and DNAH10 genes are correlated with response to NAC and OS

We tested whether any association between the mutational profile and response to NAC (R n=51, NR n=46) could be seen. However, none of the frequently mutated genes nor pathways showed significantly different proportions between R and NR. Conversely, we found significant differences in five genes not previously reported to be related to response to NAC in MIBC (Fig. 3B, Supplementary Table 7, Supplementary Fig. 3B,4A): MGAM2 (R=10, NR=1; p=0.0086), DNAH8 (R=12, NR=2; p=0.0086), DNAH6 (R=11, NR=2; p=0.0163, DNAH10(R=10, NR=2; p=0.0301), RNF213 (R=13, NR=4; p=0.0349) (Supplementary Fig. 2B). As expected, mutations in DNAH genes (DNAH8, DNAH6 and DNAH10) tend to co-occur, while mutations in RNF213 and DNAH genes are mutually exclusive. Moreover, mutations in MGAM2 and RNF213 have a high co-occurrence (Supplementary Fig. 4B). Survival analysis of these 5 genes did not yield statistically significant results (Supplementary Fig. 5A), probably due to the low frequency of these events. Grouping patients with mutations in any of the DNAH family genes (DNAH8, DNAH6 and DNAH10) did not reach statistical significance in our cohort either (p=0.369; HR=1.41[0.60-3.00], Fig. 4C). To overcome the limitation due to the low frequency of the events and further investigate the potential prognostic value of these genes, we used different publicly available BC datasets from cBioPortal. With that regard, combined mutations in DNAH8, DNAH6 and DNAH10 genes showed a positive association with better 5-year OS (p=3.89e-04; HR=1.62[:1.29-2.02]) (Supplementary Fig. 5B). Those results suggest that somatic mutations in any of DNAH family genes might predict not only NAC sensitivity, but also better OS.

Biomarkers of response to platinum-based chemotherapy reported in other clinical contexts, such as in the metastatic settings like tumor mutational burden (TMB, total mutations per Mb) and APOBEC mediated-mutation scores (APOBECscore) could not effectively distinguish between R and NR in our cohort. APOBECscore did not show a significant correlation with 5-year OS either, whereas TMB was significantly associated with 5-year OS (p=0.0231, HR=2.1636[1.09-4.28], suggesting that TMB has rather a prognostic than predictive value (Supplementary Fig. 5C).

Deletions in CDKN2A and KDM6A are correlated with response

SCNVs play a crucial role in cancer development, progression and treatment resistance/sensitivity. Amplifications and deletions in genes frequently implicated in BC (E2F3amp, PPARGamp, MDM2amp, ERBB2amp, CDKN2Adel and RB1del)[12,27] occurred across the whole cohort. ERCC1amp and KDM6Adel have previously been identified for their significant role in response to platinum-based chemotherapy[28,29]. Among the SCNVs studied, only CDKN2Adel and KDM6Adel had a statistically significant correlation with NR (p=0.014) (Fig. 4B, Supplementary Table 8). However, survival analysis did not show any significant association between these SCNVs and 5-year OS (Fig. 4C). These findings suggest that these deletions may be more useful for predicting NAC response rather than acting as prognostic markers.

Molecular features can predict NAC response in MIBC

Gene expression signatures from RNA-Seq analysis (WNT, Top10dn, Top10up, GROWTH FACTOR, CELL CYCLE REG) and genomic events from WES (mutations in DNAH8, DNAH6 or DNAH10 and deletions in KDM6A) were used to construct a machine learning (ML) algorithm to predict response to NAC in MIBC patients (Supplementary Table 9). The random forest (RF) model trained with both RNA and WES features (RF-RW) achieved an area under the curve (AUC) of 0.82 (Fig. 5A, Supplementary Table 10). Importantly, random forest (RF) models offer insights into feature importance, which indicates how much every feature contributes to predicting the target variable. The most important features were WNT (0.24) and Top10dn (0.197) signatures, followed by Top10up (0.186), CELL CYCLE REG (0.156) and GROWTH FACTOR (0.145) signatures (Fig. 5B, Supplementary Table 10). The variables that contributed less to the model were KDM6Adel (0.041) and DNAHalt (0.035) (Fig. 5B, Supplementary Table 10).

To test the impact of transcriptomic and genomic features to the model performance, we constructed two separate models. The model considering only WES features (mutations and SCNVs) (RF-W) had an AUC of 0.72 (Fig. 5A), while the model including only RNA-Seq variables (RF-R) had an AUC of 0.82 (Fig. 5A), being the AUC and feature importance of the latter equivalent to the complete model (Fig. 5B, Supplementary Table 10). These findings suggest that transcriptomic data plays a crucial role in predicting response to NAC.

We interrogated the correlation among the model variables to better understand the feature importance and to assess whether we were dealing with redundant features. We observed an overall weak association among the features (Fig. 5C), which would support the inclusion

of all features in the model as they capture distinct phenotypes. However, a model with only WNT signature achieved similar performance as the one combining all transcriptomic signatures (AUC=0.76, Supplementary Fig. 5D, Supplementary Table 10).

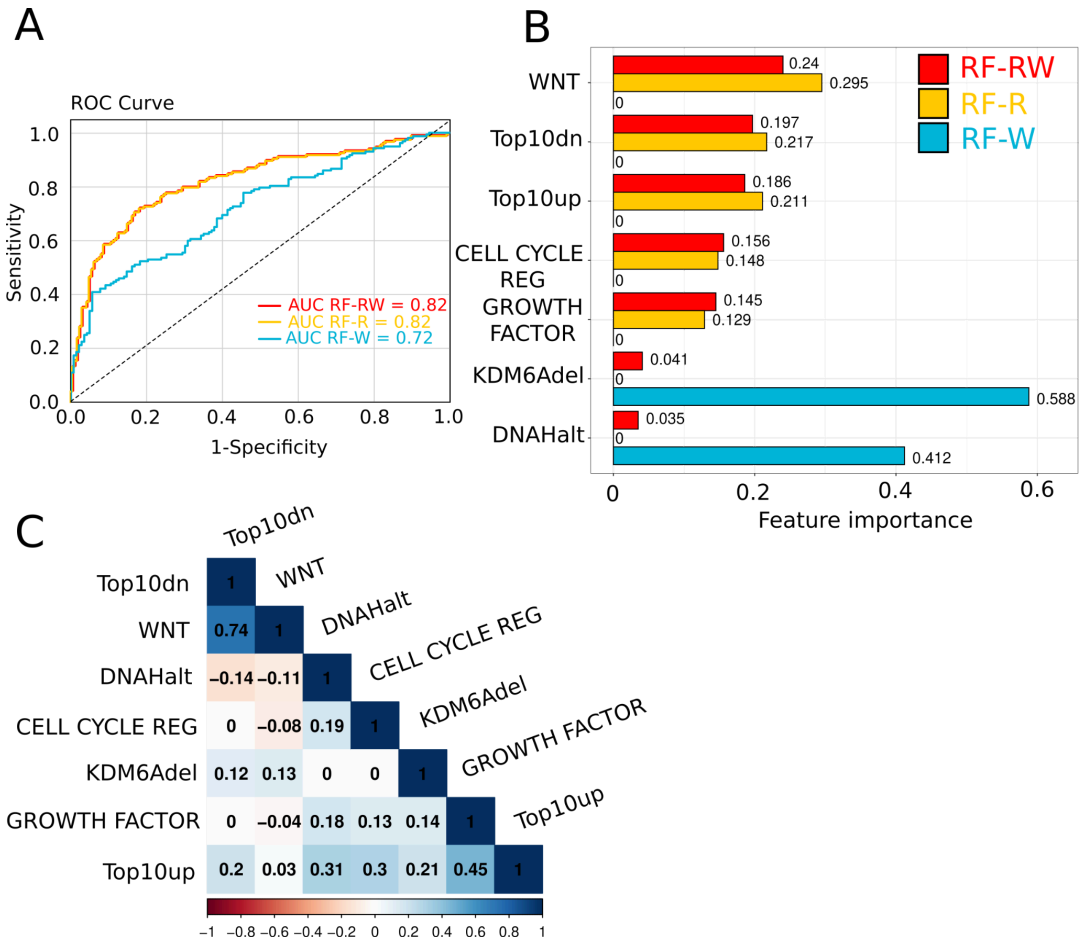

Figure 5: Machine learning models with RNA-Seq and WES data to predict response to NAC. A. *Receiver Operating Curves (ROC) of three Random Forest (RF) models combining RNA-Seq (R) and WES (W) data (RF-RW, RF-R and RF-W). The respective AUCs are: 0.82, 0.82, 0.72.* ***B.*** *RF-RW, RF-R and RF-W features ordered by decreasing feature importance.* ***C.*** *Correlation analysis of features from RNA-Seq and WES data included in the RF models.*

While some studies have shown that molecular subtypes could be potentially used to personalize BC treatment[17,21], we achieved poor AUCs for RF models using the LundTax and TCGAclas subtypes (0.41 and 0.51, respectively; Supplementary Fig. 5D, Supplementary Table 10). Hence, our study suggests that molecular subtypes alone were unable to accurately predict response to NAC in MIBC patients.

DISCUSSION

Identifying robust and accurate predictive biomarkers of NAC is a highly unmet need in MIBC. In our study, we have analyzed the transcriptomic and genomic profiles of a cohort of 100 MIBC patients treated with cisplatin-based NAC and its correlation with response to treatment (<=pT1N0). This is one the largest studies analyzing transcriptomic and genomic markers of benefit to NAC in a real-world cohort of MIBC patients. Importantly, we identified a novel gene signature comprising genes from the WNT signaling pathway strongly associated with non-response to NAC. Some previous studies have shown that the dysfunction of WNT/β-catenin signaling pathway is strongly related to the initiation and progression of different cancer types[30-34]. Furthermore, several studies have suggested a high correlation between altered expression levels of its components with sensitivity or resistance to chemotherapy[35,36]. Here we demonstrated the promising potential of this signature as a biomarker for predicting response to NAC in MIBC, being the most important feature in all the trained models while also showing high performance on its own.

Differential gene expression analysis between R and NR identified 602 DE genes, most of them associated with NR. Many of the top up-regulated genes in R, such as XIST, ENSG00000212939 (LINC01354) and ENSG00000197813 (LINC01420) are long non coding genes (lncRNAs) involved in cell proliferation, increasing the initial sensitivity to cisplatin due to the higher replication rates[37-41]. Additionally, LIPG and RNASE2 genes have also been correlated with proliferation and cancer growth[42,43] and FTCD is a novel candidate to be a tumor suppressor in hepatocellular carcinoma[44].

Despite the molecular subtype distribution in our cohort being in agreement with previous studies[12,14,20], no significant association with response was found, which could be a consequence of our limited cohort size.

On the other hand, several of the top 10 down-regulated genes in R have been highly associated with bad prognosis in BC. For instance, LPGR6, involved in WNT/beta-catenin signaling, has been associated with bad prognosis and chemoresistance[45,46]. KRT6A, KRT5 and DSG3 are markers for basal and squamous-differentiation. Basal-squamous tumors are aggressive and confer a poor survival[12,47]. Some studies have suggested that a basal subgroup may show benefit from NAC[18,19,48]. However, they were not the same subgroups as Basal-squamous in TCGA or LundTax and the finding was based on a survival difference between two cohorts. Notably, several other studies have suggested that high expression of basal markers are correlated with cancer recurrence and chemotherapy resistance[20,49-52].

We identify two signatures of cell cycle regulation (FGFR3, CCND1, E2F3, RB1 and CDKN2A) and growth factor (EGFR, ERBB2 and ERBB3), previously used for molecular subtype classification by Aramendía-Cotillas E. et al., that could discriminate between R and NR[16]. Genes from both signatures play a critical role in regulating cell growth and proliferation[53,54]. Moreover, growth factor genes are associated with luminal subtype and its high expression can sensitize cancer cells improving the response to NAC[50,55]. Our findings suggest that cells which are actively dividing are more susceptible to NAC treatment as chemotherapy can effectively target and destroy cancer cells which proliferate rapidly, inhibiting tumor growth and progression.

The mutational landscape of our cohort was consistent with those found in other MIBC cohorts[12,26,27,56]. While mutations in genes such as TP53, ERCC2, RB1, ATM and FGFR3 are known to correlate with response to cisplatin-based NAC[57-63], in our study, more SNVs in ERCC2, RB1 and ATM were found in R compared to NR, but no statistically significant differentiation was found[59,60,64-67]. These results may be due to the limited number of samples and the influence of other factors to NAC sensitivity[61].

However, we identified significant differences in five genes (MGAM2, DNAH8, DNAH6, DNAH10 and RNF213) not previously linked to NAC response in MIBC. Notably, somatic mutations in the DNAH family genes, involved in cell motility, have been associated with chemotherapy sensitivity in gastric cancer[68]. We showed that mutations in DNAH genes (DNAH8, DNAH6 and DNAH10) were also significantly correlated with 5-year OS in the cBioPortal across several BC datasets. Correlation between RNF213 and MGAM2 genes with NAC response still remains unknown. However, previous studies have suggested that MGAM2 expression is related to immune response with a potential role for predicting response to ICI[69], improving the ongoing research focused on studying the response mechanism to develop algorithms to predict ICI response[70,71].

Regarding somatic copy number variations (SCNVs), our results reveal a significant correlation between deletion in the KDM6A gene and NR, revealing its crucial role in acquiring resistance to chemotherapy. These findings are consistent with some studies which suggest that the loss of KDM6A is correlated with chemotherapy resistance, since it compromises DNA damage repair[29,72].

The main limitation of our study is the limited number of samples in our dataset, which may increase the risk of overfitting and may result in biased models that do not perform equally well in larger, more diverse populations. For instance, some signatures like Top10up and Top10dn may be specific to our MIBC cohort, as they were identified through DE analysis

and play a significant role in distinguishing between R and NR. In contrast, the WNT signature is more likely to be applied in different datasets since it was identified using an unsupervised approach, making it less likely to be cohort-specific, and it has a strong support in the literature. Therefore, despite the high AUC scores, these results should be interpreted cautiously and be validated in future research when new cohorts are available. Yet, we employed robust internal validation that showed our results to be sound. Additionally, the retrospective nature of the cohort and the use of FFPE samples can introduce some biases in the data.

In our study, we developed three ML models (RF-RW, RF-R and RF-WNT) using RNA-Seq and WES data. All of them showed promising performance with high AUC scores (RF-RW 0.82, RF-R 0.82 and RF-WNT 0.76). Feature importance revealed the Top10dn and WNT signatures are the most important ones, suggesting the critical role of transcriptomic data in distinguishing between R and NR. Most of the prior studies in MIBC have been focused on finding potential biomarkers, molecular subtypes or signatures which can be associated with response to cisplatin-based NAC in MIBC[12,14,18,20,59,60,64,73-75]. However, only few models are publicly available to predict cisplatin-based NAC response using transcriptomics. We identified novel transcriptomic signatures with better performance than previously reported ones[76-78], being the WNT signature the most promising one having the highest feature importance and adaptability for several datasets.

Additionally, the combination of any of these models with other diagnostic tools and clinical biomarkers could in the future help to enhance their predictive capabilities in order to provide a more comprehensive approach to patient care and personalized therapy. This will be paramount for transitioning these models from research settings into practical tools that can support clinical decision-making.

In conclusion, our analysis shows that RF-R, with its simpler and more cost-effective data requirements, could be the one selected as a promising candidate for being translated into clinical practice. Future work should aim to validate these models on larger, more diverse datasets to ensure robustness, accuracy, and applicability in varied clinical scenarios. With this purpose, we have made the ML models available via our GitHub repository to promote future external validation.

MATERIALS AND METHODS

Patients and samples

A total of 130 patients with MIBC were retrospectively collected from four Catalan hospitals in Spain. Samples and data from patients included in this study were provided by the Hospital del Mar Biobank (MARBiobanc) and IGTP-HUGTP Biobank, both integrated in the Spanish National Biobanks and Biomodels Network of Instituto de Salud Carlos III (PT20/00023 and PT20/00050) and Tumor Bank Network of Catalonia. They were processed following standard operating procedures with the appropriate approval of the Ethical and Scientific Committees.

All patients received neoadjuvant chemotherapy (NAC) with either cisplatin-gemcitabine or dose dense MVAC (Methotrexate, Vinblastine, Adriamycin and Cisplatin). Patients achieving a dowstaging to non-MIBC status with no pathological lymph-node involvement (<=pT1N0) observed at cystectomy were defined as responders. For each patient, formalin-fixed paraffin-embedded (FFPE) pre-treatment samples were obtained after transurethral removal of bladder tumor (TURB). Out of the total 130 patients, 100 were eligible for molecular profiling. Among these, 100 pre-treatment samples were available to perform RNA-seq analysis (n=71) and WES analysis (n=97). A flow chart with details about samples and patients included is shown in Figure 1.

Statistical analyses

Descriptive statistics of several clinical variables within a cohort of 100 patients, comparing between responders and non-responders, was performed using CompareGroups R package v.4.8.0. Kaplan–Meier survival curves were generated with survival v.3.4.0 and survminer v.0.4.9 R packages and p.values were obtained using a log-rank test.

RNA and DNA sequencing

RNA extraction and library preparation was performed by the MARGenomics core facility at Hospital del Mar Institute for Medical Research. Samples were extracted using the AllPrep DNA/RNA FFPE kit from Qiagen following the manufacturer's instructions. RNA samples were degraded as expected due to the formalin and paraffin protocols to preserve tissue's histology. For RNA libraries, a combination of two Illumina protocols was utilized: TruSeq RNA Exome for the amplification steps (steps 1-6 of the published protocol), followed by the RNA Prep with Enrichment protocol for pool exome enrichment (step 5 to 10, the last of the

published protocol). For highly degraded FFPE samples, the best results were obtained by combining these two protocols rather than exclusively using the RNA Prep with Enrichment protocol*. Sequencing of RNA libraries was conducted at the Centre de Regulació Genòmica (CRG) core facility using NextSeq 2000, with 2x50 read length and an average of 25 million reads per sample. qPCR was performed in order to properly quantify the libraries.

RNA bioinformatics analyses

Raw fastq files were quality controlled using fastQC and fastqscreen. Subsequently, alignment was performed using STAR v.2.7.8 using GRCh38 genome as a reference and version 41 of hg38 GTF from gencode as annotation. Additionally, Picard v.2.25.1 was used to check the quality of the alignment. Quantification was performed using featureCounts from Subread package v.2.0.3 and version 41 of hg38 GTF from gencode as annotation. Minimum library size was 1.5 million counts. Lowly expressed genes (rowSums(counts.m>10) >= 42) were removed for further analyses. Principal Component Analysis (PCA) and Hierarchical Clustering (HC) were performed in order to check any outliers or groups due to either clinical or technical variables.

Differential expression analysis

Limma package v.3.54.2 was used to perform a differential expression analysis between responders and non-responders using TMM normalized counts obtained by edgeR v.3.40.2. Age, sex and batch were used to adjust the model as a fixed effect, while hospital variable was added to the model as a random effect. Voom function was used to model the mean-variance relationship. False discovery rate was used to correct for multiple testing. Due to the lack of significant adjusted p-values, genes were considered differentially expressed with p-value < 0.05 and |logFC| >1.

Molecular subtype classification

LundTax molecular subtype classification was performed using LundTax2023Classifier R package v.1.1.1 and TPM normalized counts. TCGAclas molecular subtype classification was also performed using BLCAsubtyping R package v.2.1.1 using logTPMs. Differences in molecular subtype proportions between responders and non-responders were calculated using Fisher's exact test. Singscore R package v.1.18.0 was employed to perform a single-sample gene signature scoring of the different LundTax and TCGAclas classifiers' signatures and results were compared between R and NR using Wilcoxon test. In order to facilitate understanding the functionality of the two most significantly different signatures between R and NR, we changed the names of ERBB by GROWTH FACTOR and Circuit by

CELL CYCLE REG. Survival analysis for LundTax signatures which are different between R and NR was also performed using survival v.3.4.0 and survminer v.0.4.9 R packages. Optimal cut-off point to separate high and low expression for the signatures scores was calculated using surv_cutpoint from survminer v.0.4.9 R package.

Weighted gene correlation network analysis (WGCNA)

Weighted gene correlation network analysis (WGCNA) using the R WGCNA package v1.72-5 was used to find clusters (modules) of highly correlated genes using the table of counts in TMM and setting the split number to 3 with a minimum cluster size of 25 genes. Pearson correlation was performed to calculate correlation between the first principal component of cluster genes and clinical variables. Over-representation analysis (ORA) was conducted on genes within the cluster that showed significant correlation with the response variable (green cluster, n=77), using the *enricher* function from clusterProfiler R package v.4.6.2 using an adjusted p-value <0.05 and Hallmark, C2 and C5 GO BP MSigDB collections v7.5.1. Genes related to WNT signaling pathway from C5 collection (GOBP_CANONICAL_WNT_SIGNALING_PATHWAY,
GOBP_CELL_CELL_SIGNALING_BY_WNT,
GOBP_REGULATION_OF_CANONICAL_WNT_SIGNALING_PATHWAY,
GOBP_REGULATION_OF_WNT_SIGNALING_PATHWAY,GOBP_NEGATIVE_REGULATION_OF_CANONICAL_WNT_SIGNALING_PATHWAY,
GOBP_NEGATIVE_REGULATION_OF_WNT_SIGNALING_PATHWAY) were put together and intersected with DE genes resulting into 17 genes.

Top10up, Top10dn, GROWTH FACTOR, CELL CYCLE REG and WNT gene expression signatures

Singscore R package v.1.18.0 was employed to perform a single-sample gene signature scoring of the Top10up, Top10dn, GROWTH FACTOR, CELL CYCLE REG and WNT gene expression signatures (Supplementary Table 9). Signature scores were compared between R and NR using the Wilcoxon rank sum test. Optimal cut-off point to separate high and low expression for the signature scores was calculated using surv_cutpoint from survminer v.0.4.9 R package. Survival analysis was also performed as previously described.

A heatmap was performed with the expression of the five signature genes (Top10up, Top10dn, GROWTH FACTOR, CELL CYCLE REG and WNT) using the score obtained by Singscore R package v.1.18.0 to detect any cluster based on the variable response. Additionally, correlation among genes from all identified signatures in the RNA-Seq data was

performed to avoid the multicollinearity, enhancing the robustness and stability of the future ML algorithm.

Whole exome sequencing (WES) bioinformatics analyses

Fastq files were processed with the nf-core/sarek pipeline v.3.4.0 with default parameters and reference genome for GRCh38. Raw fastq files were quality controlled using fastQC v.0.12.1 and fastp v.0.23.4. Subsequently, an alignment step was performed using BWA 0.7.17 and its quality was checked using Picard v.2.25.1. Variant calling (VC) was performed using mutect2 from GATK4 v.4.4.0.0 in tumor-only mode. Additionally, the quality of the identified variants was assessed using the FilterMutectCalls function. VCF files were annotated using VEP v.10.2.0 and transformed to MAF objects using vcf2maf v.1.6.21. Only variants with quality filter “PASS” were included in downstream analyses. Additionally, we removed variants annotated in gnomAD v.2.1.1 or exhibiting a gnomAD allele frequency (AF) > 0.01. We also filtered out variants with low sequencing depth (DP < 20) and those with AF > 0.95 or AF < 0.05.

Somatic mutation landscape

MAFtools R package v.2.18.1 was used to analyze somatic mutations from MAF files. The top 15 bladder cancer mutated genes (Supplementary Table 5) were plotted using oncoplot function. Additionally, TMB was calculated using the tmb function. TMB from our cohort was compared to the TCGAclas-BLCA cohort using the TCGAclasCompare function. The APOBEC score was calculated using the trinucleotideMatrix function. Moreover, the function mafCompare was used to detect significantly different mutated genes between R and NR. Differences in TMB and APOBEC score across the groups were assessed using the Wilcoxon rank sum test. Survival analysis was also performed using survival v.3.4.0 and survminer v.0.4.9 R packages separating patients by mutated and non-mutated genes. Moreover, in order to validate the importance of DNAH alterations, a survival analysis was performed using cBioPortal for Cancer Genomics (https://www.cbioportal.org/). We selected those patients in any of the following datasets: MSK Eur Urol 2014, MSK J Clin Oncol 2013, MSK Nat Genet 2016, MSK/TCGA 2020, TCGA Cell 2017, BGI Nat Genet 2013, DFCI/MSK Cancer Discov 2014, TCGA PanCancer Atlas, BCAN/HCRN Nat Commun 2022, and Cornell/Trento Nat Genet 2016. Patients with mutations in DNAH8, DNAH6 and DNAH10 were compared against patients without mutations in those genes.

SCNV

Somatic copy number variation (SCNVs) analysis was conducted using CNVkit v.0.09.10 in tumor-only mode. For each sample, .cns files obtained by CNVKit were merged in a GenomicRanges object splitted by sample using GenomicRanges R package v.1.50.2. Segmented log2 copy number ratios were transformed into the following states: 0, 2-copy loss; 1, 1-copy loss; 2: normal; 3, 1-copy gain; 4, >= 2-copy gain. Genes related to bladder cancer were selected and plotted using cnvOncoPrint function from CNVRanger R package v.1.14.0 to get the landscape of amplifications and deletions. Important genes previously reported in the literature were also chosen to perform a statistical analysis to detect differences between R and NR in indels and amplifications using CompareGroups R package v.4.8.0. Survival analysis was also performed using survival v.3.4.0 and survminer v.0.4.9 R packages separating patients by the presence of deletions or amplifications.

Predictive models of response to NAC

The machine learning framework developed for this project consists of two different steps. The first one encompasses all preprocessing steps, including scaling and splitting data. During the training phase, optimal hyperparameters for the ML algorithm are determined through 15-fold cross-validation, repeating the process multiple times to ensure the robustness and avoid overfitting for a specific seed. Subsequently, the model is evaluated using the traditional train/test split method, where 70% of our data is used for training and the remaining 30% is reserved for testing. Moreover, bootstrapping was also used to reduce overfitting and improve the accuracy.

We selected a set of variables significantly associated with R, ensuring low correlation among themselves and avoiding the overlap in case of gene signatures. The final model (RF-RW) includes the following variables: WNT, Top10up, Top10dn, GROWTH FACTOR and CELL CYCLE REG signatures (signscore), alterations in DNAH8, DNAH6 or DNAH10 (0=non-altered, 1 altered), deletions in KDM6A (2=no deletion, 1=deletion).

Finally, a random forest (RF) model was used for the final model. RF was chosen because it is very robust with overfitting. Additionally, to mitigate the consequences of our limited sample size, we tested the model using 1000 different seeds to achieve more reliable and consistent scores. Moreover, the Bootstrap .632+ method was applied as an internal validation strategy, providing information about how the model might perform on datasets not included in the training phase. We have also run the model using different features in order

to know whether the AUC is stable or not. We tested models using only RNA-Seq features (RF-R),only WES features (RF-W) and using only the WNT score.

Data availability

We have made available three independent L models (RF-RW , RF-R and RF-WNT), accessible via our GitHub repository (ongoing). WES and RNA-Seq data are available in EGA (ongoing).

Acknowledgements

The work was supported by the following grants and agencies: Project PI19/00004, funded by Instituto de Salud Carlos III (ISCIII) and co-funded by the European Union; a grant from FIS-ISCIII (FI20/00095), 2021SGR00042 by Generalitat de Catalunya;

This work was also supported by the “Xarxa de Bancs de Tumors” sponsored by Pla Director d’Oncologia de Catalunya (XBTC). We want to particularly acknowledge the patients and the IGTP-HUGTP Biobank integrated in the Spanish National Biobanks and BiomodelsNetwork of Instituto de Salud Carlos III (PT20/00050) and Tumor Bank Network of Catalonia for its collaboration.

Competing interests

Potential conflicts of interest: J. Bellmunt has served in consulting or advisory roles for Astellas Pharma, AstraZeneca/MedImmune, Bristol Myers Squibb, Genentech, Novartis, Pfizer, Pierre Fabre, and the healthcare business of Merck KGaA, Darmstadt, Germany; has received travel and accommodation expenses from Ipsen, Merck & Co., Kenilworth, NJ, and Pfizer; reports patents, royalties, other intellectual property from UpToDate; reports stock and other ownership interests in Rainier Therapeutics; has received honoraria from UpToDate; and has received institutional research funding from Millennium, Pfizer, Sanofi, and the healthcare business of Merck KGaA, Darmstadt, Germany. A. Rodriguez-Vida has served in consulting or advisory roles for Astellas Pharma, Bristol Myers Squibb, Novartis, Pfizer, Johnson&Johnson, Merck, Bayer, MSD, Ipsen; has received travel and accommodation expenses from Ipsen, Merck, Johnson&Johnson and Bayer.

References

1. Lenis, A. T., Lec, P. M., Chamie, K., & Mshs, M. (2020). Bladder Cancer: A Review. *JAMA*, *324*(19), 1980. https://doi.org/10.1001/jama.2020.17598.
2. Grossman, H. B., Natale, R. B., Tangen, C. M., Speights, V. O., Vogelzang, N. J., Trump, D. L., White, R. W. deVere, Sarosdy, M. F., Wood, D. P., Raghavan, D., & Crawford, E. D. (2003). Neoadjuvant Chemotherapy plus Cystectomy Compared with Cystectomy Alone for Locally Advanced Bladder Cancer. *New England Journal of Medicine*, *349*(9), 859–866. https://doi.org/10.1056/NEJMoa022148.
3. Patel, V. G., Oh, W. K., & Galsky, M. D. (2020). Treatment of muscle-invasive and advanced bladder cancer in 2020. *CA: A Cancer Journal for Clinicians*, *70*(5), 404–423. https://doi.org/10.3322/caac.21631.
4. Iyer, G., Tully, C. M., Zabor, E. C., Bochner, B. H., Dalbagni, G., Herr, H. W., Donat, S. M., Russo, P., Ostrovnaya, I., Regazzi, A. M., Milowsky, M. I., Rosenberg, J. E., & Bajorin, D. F. (2020). Neoadjuvant Gemcitabine-Cisplatin Plus Radical Cystectomy-Pelvic Lymph Node Dissection for Muscle-invasive Bladder Cancer: A 12-year Experience. *Clinical Genitourinary Cancer*, *18*(5), 387–394. https://doi.org/10.1016/j.clgc.2020.02.014.
5. Cowan, N. G., Chen, Y., Downs, T. M., Bochner, B. H., Apolo, A. B., Porter, M. P., La Rochelle, J. C., Amling, C. L., & Koppie, T. M. (2014). Neoadjuvant Chemotherapy Use in Bladder Cancer: A Survey of Current Practice and Opinions. *Advances in Urology*, *2014*, 1–6. https://doi.org/10.1155/2014/746298.
6. Huo, J., Ray-Zack, M. D., Shan, Y., Chamie, K., Boorjian, S. A., Kerr, P., Jana, B., Freedland, S. J., Kamat, A. M., Mehta, H. B., & Williams, S. B. (2019). Discerning Patterns and Quality of Neoadjuvant Chemotherapy Use Among Patients with Muscle-invasive Bladder Cancer. *European Urology Oncology*, *2*(5), 497–504. https://doi.org/10.1016/j.euo.2018.07.009.
7. Flegar, L., Kraywinkel, K., Zacharis, A., Aksoy, C., Koch, R., Eisenmenger, N., Groeben, C., & Huber, J. (2022). Treatment trends for muscle-invasive bladder cancer in Germany from 2006 to 2019. *World Journal of Urology*, *40*(7), 1715–1721. https://doi.org/10.1007/s00345-022-04017-z.
8. Duplisea, J. J., Mokkapati, S., Plote, D., Schluns, K. S., McConkey, D. J., Yla-Herttuala, S., Parker, N. R., & Dinney, C. P. (2019). The development of interferon-based gene therapy for BCG unresponsive bladder cancer: from bench to bedside. *World Journal of Urology*, *37*(10), 2041–2049. https://doi.org/10.1007/s00345-018-2553-7.
9. Damrauer, J. S., Hoadley, K. A., Chism, D. D., Fan, C., Tiganelli, C. J., Wobker, S. E., Yeh, J. J., Milowsky, M. I., Iyer, G., Parker, J. S., & Kim, W. Y. (2014). Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proceedings of the National Academy of Sciences*, *111*(8), 3110–3115. https://doi.org/10.1073/pnas.1318376111.
10. Choi, W., Porten, S., Kim, S., Willis, D., Plimack, E. R., Hoffman-Censits, J., Roth, B., Cheng, T., Tran, M., Lee, I.-L., Melquist, J., Bondaruk, J., Majewski, T., Zhang, S., Pretzsch, S., Baggerly, K., Siefker-Radtke, A., Czerniak, B., Dinney, C. P. N., & McConkey, D. J. (2014). Identification of Distinct Basal and Luminal Subtypes of Muscle-Invasive Bladder Cancer with Different Sensitivities to Frontline

Chemotherapy. *Cancer Cell*, *25*(2), 152–165. https://doi.org/10.1016/j.ccr.2014.01.009.

11. Rebouissou, S., Bernard-Pierrot, I., De Reyniès, A., Lepage, M.-L., Krucker, C., Chapeaublanc, E., Hérault, A., Kamoun, A., Caillault, A., Letouzé, E., Elarouci, N., Neuzillet, Y., Denoux, Y., Molinié, V., Vordos, D., Laplanche, A., Maillé, P., Soyeux, P., Ofualuka, K., … Radvanyi, F. (2014). EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. *Science Translational Medicine*, *6*(244). https://doi.org/10.1126/scitranslmed.3008970.
12. Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., Hinoue, T., Laird, P. W., Hoadley, K. A., Akbani, R., Castro, M. A. A., Gibb, E. A., Kanchi, R. S., Gordenin, D. A., Shukla, S. A., Sanchez-Vega, F., Hansel, D. E., Czerniak, B. A., Reuter, V. E., … Lerner, S. P. (2017). Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell*, *171*(3), 540-556.e25. https://doi.org/10.1016/j.cell.2017.09.007.
13. Sjödahl, G., Eriksson, P., Liedberg, F., & Höglund, M. (2017). Molecular classification of urothelial carcinoma: global mRNA classification versus tumour-cell phenotype classification. *The Journal of Pathology*, *242*(1), 113–125. https://doi.org/10.1002/path.4886.
14. Kamoun, A., De Reyniès, A., Allory, Y., Sjödahl, G., Robertson, A. G., Seiler, R., Hoadley, K. A., Groeneveld, C. S., Al-Ahmadie, H., Choi, W., Castro, M. A. A., Fontugne, J., Eriksson, P., Mo, Q., Kardos, J., Zlotta, A., Hartmann, A., Dinney, C. P., Bellmunt, J., … Zlotta, A. (2020). A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. *European Urology*, *77*(4), 420–433. https://doi.org/10.1016/j.eururo.2019.09.006.
15. Mo, Q., Li, R., Adeegbe, D. O., Peng, G., & Chan, K. S. (2020). Integrative multi-omics analysis of muscle-invasive bladder cancer identifies prognostic biomarkers for frontline chemotherapy and immunotherapy. *Communications Biology*, *3*(1), 784. https://doi.org/10.1038/s42003-020-01491-2.
16. Aramendía Cotillas, E., Bernardo, C., Veerla, S., Liedberg, F., Sjodahl, G., & Eriksson, P. *A versatile and upgraded version of the LundTax classification algorithm applied to independent cohorts*. Preprint at https://www.biorxiv.org/content/10.1101/2023.12.15.571519v1 (2023).
17. Griffin, J., Down, J., Quayle, L. A., Heath, P. R., Gibb, E. A., Davicioni, E., Liu, Y., Zhao, X., Swain, J., Wang, D., Hussain, S., Crabb, S., Catto, J. W., & the GUSTO Trial Management Group. (2024). Verification of molecular subtyping of bladder cancer in the GUSTO clinical trial. *The Journal of Pathology: Clinical Research*, *10*(2), e12363. https://doi.org/10.1002/2056-4538.12363.
18. Seiler, R., Ashab, H. A. D., Erho, N., Van Rhijn, B. W. G., Winters, B., Douglas, J., Van Kessel, K. E., Fransen Van De Putte, E. E., Sommerlad, M., Wang, N. Q., Choeurng, V., Gibb, E. A., Palmer-Aronsten, B., Lam, L. L., Buerki, C., Davicioni, E., Sjödahl, G., Kardos, J., Hoadley, K. A., … Black, P. C. (2017). Impact of Molecular Subtypes in Muscle-invasive Bladder Cancer on Predicting Response and Survival after Neoadjuvant Chemotherapy. *European Urology*, *72*(4), 544–554. https://doi.org/10.1016/j.eururo.2017.03.030.
19. Lotan, Y., De Jong, J. J., Liu, V. Y. T., Bismar, T. A., Boorjian, S. A., Huang, H.-C., Davicioni, E., Mian, O. Y., Wright, J. L., Necchi, A., Dall'Era, M. A., Kaimakliotis, H. Z., Black, P. C., Gibb, E. A., & Boormans, J. L. (2022). Patients with Muscle-Invasive

Bladder Cancer with Nonluminal Subtype Derive Greatest Benefit from Platinum Based Neoadjuvant Chemotherapy. *Journal of Urology*, *207*(3), 541–550. https://doi.org/10.1097/JU.0000000000002261.

20. Sjödahl, G., Abrahamsson, J., Holmsten, K., Bernardo, C., Chebil, G., Eriksson, P., Johansson, I., Kollberg, P., Lindh, C., Lövgren, K., Marzouka, N.-D., Olsson, H., Höglund, M., Ullén, A., & Liedberg, F. (2022). Different Responses to Neoadjuvant Chemotherapy in Urothelial Carcinoma Molecular Subtypes. *European Urology*, *81*(5), 523–532. https://doi.org/10.1016/j.eururo.2021.10.035.
21. Satyal, U., Sikder, R. K., McConkey, D., Plimack, E. R., & Abbosh, P. H. (2019). Clinical implications of molecular subtyping in bladder cancer. *Current Opinion in Urology*, *29*(4), 350–356. https://doi.org/10.1097/MOU.0000000000000641.
22. Kang, H. W., Kim, W.-J., & Yun, S. J. (2020). The therapeutic and prognostic implications of molecular biomarkers in urothelial carcinoma. *Translational Cancer Research*, *9*(10). https://doi.org/10.21037/tcr-20-1243.
23. Bajorin, D. F., Witjes, J. A., Gschwend, J. E., Schenker, M., Valderrama, B. P., Tomita, Y., Bamias, A., Lebret, T., Shariat, S. F., Park, S. H., Ye, D., Agerbaek, M., Enting, D., McDermott, R., Gajate, P., Peer, A., Milowsky, M. I., Nosov, A., Neif Antonio, J., … Galsky, M. D. (2021). Adjuvant Nivolumab versus Placebo in Muscle-Invasive Urothelial Carcinoma. *The New England Journal of Medicine*, *384*(22), 2102–2114. https://doi.org/10.1056/NEJMoa2034442.
24. Sudoł, D., Widz, D., Mitura, P., Płaza, P., Godzisz, M., Kuliniec, I., Yadlos, A., Cabanek, M., Bar, M., & Bar, K. (2022). Neutrophil-to-lymphocyte ratio as a predictor of overall survival and cancer advancement in patients undergoing radical cystectomy for bladder cancer. *Central European Journal of Urology*, *75*(1), 41–46. https://doi.org/10.5173/ceju.2022.0273.
25. Lindgren, D., Sjödahl, G., Lauss, M., Staaf, J., Chebil, G., Lövgren, K., Gudjonsson, S., Liedberg, F., Patschan, O., Månsson, W., Fernö, M., & Höglund, M. (2012). Integrated Genomic and Gene Expression Profiling Identifies Two Major Genomic Circuits in Urothelial Carcinoma. *PLoS ONE*, *7*(6), e38863. https://doi.org/10.1371/journal.pone.0038863.
26. Cooley, L. F., Glaser, A. P., & Meeks, J. J. (2022). Mutation signatures to Pan-Cancer Atlas: Investigation of the genomic landscape of muscle-invasive bladder cancer. *Urologic Oncology: Seminars and Original Investigations*, *40*(7), 279–286. https://doi.org/10.1016/j.urolonc.2020.01.019.
27. Pérez-Montiel, M. D., Cerrato-Izaguirre, D., Sánchez-Pérez, Y., Diaz-Chavez, J., Cortés-González, C. C., Rubio, J. A., Jiménez-Ríos, M. A., Herrera, L. A., Scavuzzo, A., Meneses-García, A., Hernández-Martínez, R., Vaca-Paniagua, F., Ramírez, A., Orozco, A., Cantú-de-León, D., & Prada, D. (2023). Mutational Landscape of Bladder Cancer in Mexican Patients: KMT2D Mutations and chr11q15.5 Amplifications Are Associated with Muscle Invasion. *International Journal of Molecular Sciences*, *24*(2), 1092. https://doi.org/10.3390/ijms24021092.
28. Chiu, T.-J., Chen, C.-H., Chien, C.-Y., Li, S.-H., Tsai, H.-T., & Chen, Y.-J. (2011). High ERCC1 expression predicts cisplatin-based chemotherapy resistance and poor outcome in unresectable squamous cell carcinoma of head and neck in a betel-chewing area. *Journal of Translational Medicine*, *9*(1), 31. https://doi.org/10.1186/1479-5876-9-31.

29. Matar, M., Prince, G., Hamati, I., Baalbaky, M., Fares, J., Aoude, M., Matar, C., & Kourie, H. R. (2023). Implication of *KDM6A* in Bladder Cancer. *Pharmacogenomics*, *24*(9), 509–522. https://doi.org/10.2217/pgs-2023-0027.
30. Zhan, T., Rindtorff, N., & Boutros, M. (2017). Wnt signaling in cancer. *Oncogene*, *36*(11), 1461–1473. https://doi.org/10.1038/onc.2016.304.
31. Cao, M.-Q., You, A.-B., Zhu, X.-D., Zhang, W., Zhang, Y.-Y., Zhang, S.-Z., Zhang, K., Cai, H., Shi, W.-K., Li, X.-L., Li, K.-S., Gao, D.-M., Ma, D.-N., Ye, B.-G., Wang, C.-H., Qin, C.-D., Sun, H.-C., Zhang, T., & Tang, Z.-Y. (2018). miR-182-5p promotes hepatocellular carcinoma progression by repressing FOXO3a. *Journal of Hematology & Oncology*, *11*(1), 12. https://doi.org/10.1186/s13045-018-0555-y.
32. Spaan, I., Raymakers, R. A., Van De Stolpe, A., & Peperzak, V. (2018). Wnt signaling in multiple myeloma: a central player in disease with therapeutic potential. *Journal of Hematology & Oncology*, *11*(1), 67. https://doi.org/10.1186/s13045-018-0615-3.
33. Zhang, M., Weng, W., Zhang, Q., Wu, Y., Ni, S., Tan, C., Xu, M., Sun, H., Liu, C., Wei, P., & Du, X. (2018). The lncRNA NEAT1 activates Wnt/β-catenin signaling and promotes colorectal cancer progression via interacting with DDX5. *Journal of Hematology & Oncology*, *11*(1), 113. https://doi.org/10.1186/s13045-018-0656-7.
34. Zhang, W., Ruan, X., Li, Y., Zhi, J., Hu, L., Hou, X., Shi, X., Wang, X., Wang, J., Ma, W., Gu, P., Zheng, X., & Gao, M. (2022). KDM1A promotes thyroid cancer progression and maintains stemness through the Wnt/β-catenin signaling pathway. *Theranostics*, *12*(4), 1500–1517. https://doi.org/10.7150/thno.66142.
35. Castagnoli, L., Cancila, V., Cordoba-Romero, S. L., Faraci, S., Talarico, G., Belmonte, B., Iorio, M. V., Milani, M., Volpari, T., Chiodoni, C., Hidalgo-Miranda, A., Tagliabue, E., Tripodo, C., Sangaletti, S., Di Nicola, M., & Pupa, S. M. (2019). WNT signaling modulates PD-L1 expression in the stem cell compartment of triple-negative breast cancer. *Oncogene*, *38*(21), 4047–4060. https://doi.org/10.1038/s41388-019-0700-2.
36. Moreno-Londoño, A. P., Castañeda-Patlán, M. C., Sarabia-Sánchez, M. A., Macías-Silva, M., & Robles-Flores, M. (2023). Canonical Wnt Pathway Is Involved in Chemoresistance and Cell Cycle Arrest Induction in Colon Cancer Cell Line Spheroids. *International Journal of Molecular Sciences*, *24*(6), 5252. https://doi.org/10.3390/ijms24065252.
37. Li, J., He, M., Xu, W., & Huang, S. (2019). LINC01354 interacting with hnRNP-D contributes to the proliferation and metastasis in colorectal cancer through activating Wnt/β-catenin signaling pathway. *Journal of Experimental & Clinical Cancer Research*, *38*(1), 161. https://doi.org/10.1186/s13046-019-1150-y.
38. Yang, G., Yang, C., She, Y., Shen, Z., & Gao, P. (2019). LINC01354 enhances the proliferation and invasion of lung cancer cells by regulating miR-340-5p/ATF1 signaling pathway. *Artificial Cells, Nanomedicine, and Biotechnology*, *47*(1), 3737–3744. https://doi.org/10.1080/21691401.2019.1667816.
39. Cui, H., Ruan, M., Xu, H., Qi, J., Ruan, L., Gao, X., Sun, X., Zhang, S., Zuo, R., & Yin, Y. (2021). LINC01420 Serves as a Novel Prognostic Biomarker and Promotes Cell Proliferation, Migration, and Invasion by Suppressing miR-149-5p in Gastric Cancer. *Critical Reviews in Eukaryotic Gene Expression*, *31*(4), 49–58. https://doi.org/10.1615/CritRevEukaryotGeneExpr.2021038743.
40. Ghafouri-Fard, S., Shirvani-Farsani, Z., Hussen, B. M., Taheri, M., & Jalili Khoshnoud, R. (2022). Emerging role of non-coding RNAs in the regulation of KRAS. *Cancer Cell International*, *22*(1), 68. https://doi.org/10.1186/s12935-022-02486-1.

41. Li, J., Ming, Z., Yang, L., Wang, T., Liu, G., & Ma, Q. (2022). Long noncoding RNA XIST: Mechanisms for X chromosome inactivation, roles in sex-biased diseases, and therapeutic opportunities. *Genes & Diseases*, *9*(6), 1478–1492. https://doi.org/10.1016/j.gendis.2022.04.007.
42. Slebe, F., Rojo, F., Vinaixa, M., García-Rocha, M., Testoni, G., Guiu, M., Planet, E., Samino, S., Arenas, E. J., Beltran, A., Rovira, A., Lluch, A., Salvatella, X., Yanes, O., Albanell, J., Guinovart, J. J., & Gomis, R. R. (2016). FoxA and LIPG endothelial lipase control the uptake of extracellular lipids for breast cancer growth. *Nature Communications*, *7*(1), 11199. https://doi.org/10.1038/ncomms11199.
43. Wu, T., Chen, Y., Yang, L., Wang, X., Chen, K., & Xu, D. (2022). Ribonuclease A Family Member 2 Promotes the Malignant Progression of Glioma Through the PI3K/Akt Signaling Pathway. *Frontiers in Oncology*, *12*, 921083. https://doi.org/10.3389/fonc.2022.921083.
44. Chen, J., Chen, Z., Huang, Z., Yu, H., Li, Y., & Huang, W. (2019). Formiminotransferase Cyclodeaminase Suppresses Hepatocellular Carcinoma by Modulating Cell Apoptosis, DNA Damage, and Phosphatidylinositol 3-Kinases (PI3K)/Akt Signaling Pathway. *Medical Science Monitor*, *25*, 4474–4484. https://doi.org/10.12659/MSM.916202.
45. Ruan, X., Liu, A., Zhong, M., Wei, J., Zhang, W., Rong, Y., Liu, W., Li, M., Qing, X., Chen, G., Li, R., Liao, Y., Liu, Q., Zhang, X., Ren, D., & Wang, Y. (2019). Silencing LGR6 Attenuates Stemness and Chemoresistance via Inhibiting Wnt/β-Catenin Signaling in Ovarian Cancer. *Molecular Therapy - Oncolytics*, *14*, 94–106. https://doi.org/10.1016/j.omto.2019.04.002.
46. Kong, Y., Ou, X., Li, X., Zeng, Y., Gao, G., Lyu, N., & Liu, P. (2020). LGR6 Promotes Tumor Proliferation and Metastasis through Wnt/β-Catenin Signaling in Triple-Negative Breast Cancer. *Molecular Therapy - Oncolytics*, *18*, 351–359. https://doi.org/10.1016/j.omto.2020.06.020.
47. Eckstein, M., Wirtz, R., Gross-Weege, M., Breyer, J., Otto, W., Stoehr, R., Sikic, D., Keck, B., Eidt, S., Burger, M., Bolenz, C., Nitschke, K., Porubsky, S., Hartmann, A., & Erben, P. (2018). mRNA-Expression of KRT5 and KRT20 Defines Distinct Prognostic Subgroups of Muscle-Invasive Urothelial Bladder Cancer Correlating with Histological Variants. *International Journal of Molecular Sciences*, *19*(11), 3396. https://doi.org/10.3390/ijms19113396.
48. Helal, D. S., Darwish, S. A., Awad, R. A., Ali, D. A., & El-Guindy, D. M. (2023). Immunohistochemical based molecular subtypes of muscle-invasive bladder cancer: association with HER2 and EGFR alterations, neoadjuvant chemotherapy response and survival. *Diagnostic Pathology*, *18*, 11. https://doi.org/10.1186/s13000-023-01295-y.
49. Ricciardelli, C., Lokman, N. A., Pyragius, C. E., Ween, M. P., Macpherson, A. M., Ruszkiewicz, A., Hoffmann, P., & Oehler, M. K. (2017). Keratin 5 overexpression is associated with serous ovarian cancer recurrence and chemotherapy resistance. *Oncotarget*, *8*(11), 17819–17832. https://doi.org/10.18632/oncotarget.14867.
50. Jütte, H., Reike, M., Wirtz, R. M., Kriegmair, M., Erben, P., Tully, K., Weyerer, V., Eckstein, M., Hartmann, A., Eidt, S., Wezel, F., Bolenz, C., Tannapfel, A., Noldus, J., & Roghmann, F. (2021). KRT20, KRT5, ESR1 and ERBB2 Expression Can Predict Pathologic Outcome in Patients Undergoing Neoadjuvant Chemotherapy and Radical Cystectomy for Muscle-Invasive Bladder Cancer. *Journal of Personalized Medicine*, *11*(6), 473. https://doi.org/10.3390/jpm11060473.

51. Shinomiya, Y., Kouchi, Y., Harada-Kagitani, S., Ishige, T., Takano, S., Ohtsuka, M., Ikeda, J., & Kishimoto, T. (2024). ECM1 and KRT6A are involved in tumor progression and chemoresistance in the effect of dexamethasone on pancreatic cancer. *Cancer Science*, cas.16175. https://doi.org/10.1111/cas.16175.
52. Xu, Q., Yu, Z., Mei, Q., Shi, K., Shen, J., Gao, G., Liu, S., & Li, M. (2024). Keratin 6A (KRT6A) promotes radioresistance, invasion, and metastasis in lung cancer via p53 signaling pathway. *Aging*. https://doi.org/10.18632/aging.205742.
53. Oshima, J., & Campisi, J. (1991). Fundamentals of Cell Proliferation: Control of the Cell Cycle. *Journal of Dairy Science*, *74*(8), 2778–2787. https://doi.org/10.3168/jds.S0022-0302(91)78458-0.
54. Tilsed, C. M., Fisher, S. A., Nowak, A. K., Lake, R. A., & Lesterhuis, W. J. (2022). Cancer chemotherapy: insights into cellular and tumor microenvironmental mechanisms of action. *Frontiers in Oncology*, *12*, 960317. https://doi.org/10.3389/fonc.2022.960317.
55. Baez-Navarro, X., van Bockstal, M. R., Jager, A., & van Deurzen, C. H. M. (2024). HER2-low breast cancer and response to neoadjuvant chemotherapy: a population-based cohort study. *Pathology*, *56*(3), 334–342. https://doi.org/10.1016/j.pathol.2023.10.022.
56. The Cancer Genome Atlas Research Network. (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*, *507*(7492), 315–322. https://doi.org/10.1038/nature12965.
57. Qureshi, K. N., Griffiths, T. R., Robinson, M. C., Marsh, C., Roberts, J. T., Hall, R. R., Lunec, J., & Neal, D. E. (1999). TP53 accumulation predicts improved survival in patients resistant to systemic cisplatin-based chemotherapy for muscle-invasive bladder cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *5*(11), 3500–3507.
58. Ma, X., Rousseau, V., Sun, H., Lantuejoul, S., Filipits, M., Pirker, R., Popper, H., Mendiboure, J., Vataire, A.-L., Le Chevalier, T., Soria, J. C., Brambilla, E., Dunant, A., & Hainaut, P. (2014). Significance of *TP53* mutations as predictive markers of adjuvant cisplatin-based chemotherapy in completely resected non-small-cell lung cancer. *Molecular Oncology*, *8*(3), 555–564. https://doi.org/10.1016/j.molonc.2013.12.015.
59. Van Allen, E. M., Mouw, K. W., Kim, P., Iyer, G., Wagle, N., Al-Ahmadie, H., Zhu, C., Ostrovnaya, I., Kryukov, G. V., O'Connor, K. W., Sfakianos, J., Garcia-Grossman, I., Kim, J., Guancial, E. A., Bambury, R., Bahl, S., Gupta, N., Farlow, D., Qu, A., … Rosenberg, J. E. (2014). Somatic *ERCC2* Mutations Correlate with Cisplatin Sensitivity in Muscle-Invasive Urothelial Carcinoma. *Cancer Discovery*, *4*(10), 1140–1153. https://doi.org/10.1158/2159-8290.CD-14-0623.
60. Plimack, E. R., Dunbrack, R. L., Brennan, T. A., Andrake, M. D., Zhou, Y., Serebriiskii, I. G., Slifker, M., Alpaugh, K., Dulaimi, E., Palma, N., Hoffman-Censits, J., Bilusic, M., Wong, Y.-N., Kutikov, A., Viterbo, R., Greenberg, R. E., Chen, D. Y. T., Lallas, C. D., Trabulsi, E. J., … Ross, E. A. (2015). Defects in DNA Repair Genes Predict Response to Neoadjuvant Cisplatin-based Chemotherapy in Muscle-invasive Bladder Cancer. *European Urology*, *68*(6), 959–967. https://doi.org/10.1016/j.eururo.2015.07.009.
61. Liu, D., Plimack, E. R., Hoffman-Censits, J., Garraway, L. A., Bellmunt, J., Van Allen, E., & Rosenberg, J. E. (2016). Clinical Validation of Chemotherapy Response

Biomarker *ERCC2* in Muscle-Invasive Urothelial Bladder Carcinoma. *JAMA Oncology*, *2*(8), 1094. https://doi.org/10.1001/jamaoncol.2016.1056.

62. Taber, A., Christensen, E., Lamy, P., Nordentoft, I., Prip, F., Lindskrog, S. V., Birkenkamp-Demtröder, K., Okholm, T. L. H., Knudsen, M., Pedersen, J. S., Steiniche, T., Agerbæk, M., Jensen, J. B., & Dyrskjøt, L. (2020). Molecular correlates of cisplatin-based chemotherapy response in muscle invasive bladder cancer by integrated multi-omics analysis. *Nature Communications*, *11*(1), 4858. https://doi.org/10.1038/s41467-020-18640-0.
63. Timmerman, D. M., Eleveld, T. F., Gillis, A. J. M., Friedrichs, C. C., Hillenius, S., Remmers, T. L., Sriram, S., & Looijenga, L. H. J. (2021). The Role of TP53 in Cisplatin Resistance in Mediastinal and Testicular Germ Cell Tumors. *International Journal of Molecular Sciences*, *22*(21), 11774. https://doi.org/10.3390/ijms222111774.
64. Groenendijk, F. H., De Jong, J., Fransen Van De Putte, E. E., Michaut, M., Schlicker, A., Peters, D., Velds, A., Nieuwland, M., Van Den Heuvel, M. M., Kerkhoven, R. M., Wessels, L. F., Broeks, A., Van Rhijn, B. W. G., Bernards, R., & Van Der Heijden, M. S. (2016). ERBB2 Mutations Characterize a Subgroup of Muscle-invasive Bladder Cancers with Excellent Response to Neoadjuvant Chemotherapy. *European Urology*, *69*(3), 384–388. https://doi.org/10.1016/j.eururo.2015.01.014.
65. Gil-Jimenez, A., Van Dorp, J., Contreras-Sanz, A., Van Der Vos, K., Vis, D. J., Braaf, L., Broeks, A., Kerkhoven, R., Van Kessel, K. E. M., Ribal, M. J., Alcaraz, A., Wessels, L. F. A., Seiler, R., Wright, J. L., Mengual, L., Boormans, J., Van Rhijn, B. W. G., Black, P. C., & Van Der Heijden, M. S. (2023). Assessment of Predictive Genomic Biomarkers for Response to Cisplatin-based Neoadjuvant Chemotherapy in Bladder Cancer. *European Urology*, *83*(4), 313–317. https://doi.org/10.1016/j.eururo.2022.07.023.
66. Taber, A., Christensen, E., Lamy, P., Agerbæk, M., Jensen, J. B., & Dyrskjøt, L. (2021). Reply to: Reconciling differences in impact of molecular subtyping on response to cisplatin-based chemotherapy. Nature Communications, 12(1), 4834. https://doi.org/10.1038/s41467-021-24839-6
67. Christensen, E., Birkenkamp-Demtröder, K., Sethi, H., Shchegrova, S., Salari, R., Nordentoft, I., Wu, H.-T., Knudsen, M., Lamy, P., Lindskrog, S. V., Taber, A., Balcioglu, M., Vang, S., Assaf, Z., Sharma, S., Tin, A. S., Srinivasan, R., Hafez, D., Reinert, T., … Dyrskjøt, L. (2019). Early Detection of Metastatic Relapse and Monitoring of Therapeutic Efficacy by Ultra-Deep Sequencing of Plasma Cell-Free DNA in Patients With Urothelial Bladder Carcinoma. Journal of Clinical Oncology, 37(18), 1547–1557. https://doi.org/10.1200/JCO.18.02052
68. Zhu, C., Yang, Q., Xu, J., Zhao, W., Zhang, Z., Xu, D., Zhang, Y., Zhao, E., & Zhao, G. (2019). Somatic mutation of DNAH genes implicated higher chemotherapy response rate in gastric adenocarcinoma patients. *Journal of Translational Medicine*, *17*(1), 109. https://doi.org/10.1186/s12967-019-1867-6.
69. Xu, S., Feng, Y., & Zhao, S. (2019). Proteins with Evolutionarily Hypervariable Domains are Associated with Immune Response and Better Survival of Basal-like Breast Cancer Patients. *Computational and Structural Biotechnology Journal*, *17*, 430–440. https://doi.org/10.1016/j.csbj.2019.03.008.
70. Okobi, T. J., Uhomoibhi, T. O., Akahara, D. E., Odoma, V. A., Sanusi, I. A., Okobi, O. E., Umana, I., Okobi, E., Okonkwo, C. C., & Harry, N. M. (2023). Immune Checkpoint Inhibitors as a Treatment Option for Bladder Cancer: Current Evidence. *Cureus*. https://doi.org/10.7759/cureus.40031.

71. Boll, L. M., Vázquez Montes De Oca, S., Camarena, M. E., Castelo, R., Bellmunt, J., Perera-Bel, J., & Albà, M. M.. *Predicting immunotherapy response in advanced bladder cancer: a meta-analysis of six independent cohorts*. Preprint at https://www.biorxiv.org/content/10.1101/2024.04.18.589711v1.full (2024).
72. Stief, S. M., Hanneforth, A.-L., Weser, S., Mattes, R., Carlet, M., Liu, W.-H., Bartoschek, M. D., Domínguez Moreno, H., Oettle, M., Kempf, J., Vick, B., Ksienzyk, B., Tizazu, B., Rothenberg-Thurley, M., Quentmeier, H., Hiddemann, W., Vosberg, S., Greif, P. A., Metzeler, K. H., … Spiekermann, K. (2020). Loss of KDM6A confers drug resistance in acute myeloid leukemia. *Leukemia*, *34*(1), 50–62. https://doi.org/10.1038/s41375-019-0497-6.
73. Teo, M. Y., Bambury, R. M., Zabor, E. C., Jordan, E., Al-Ahmadie, H., Boyd, M. E., Bouvier, N., Mullane, S. A., Cha, E. K., Roper, N., Ostrovnaya, I., Hyman, D. M., Bochner, B. H., Arcila, M. E., Solit, D. B., Berger, M. F., Bajorin, D. F., Bellmunt, J., Iyer, G., & Rosenberg, J. E. (2017). DNA Damage Response and Repair Gene Alterations Are Associated with Improved Survival in Patients with Platinum-Treated Advanced Urothelial Carcinoma. *Clinical Cancer Research*, *23*(14), 3610–3618. https://doi.org/10.1158/1078-0432.CCR-16-2520.
74. Pfister, C., Gravis, G., Fléchon, A., Soulié, M., Guy, L., Laguerre, B., Mottet, N., Joly, F., Allory, Y., Harter, V., & Culine, S. (2021). Randomized Phase III Trial of Dose-dense Methotrexate, Vinblastine, Doxorubicin, and Cisplatin, or Gemcitabine and Cisplatin as Perioperative Chemotherapy for Patients with Muscle-invasive Bladder Cancer. Analysis of the GETUG/AFU V05 VESPER Trial Secondary Endpoints: Chemotherapy Toxicity and Pathological Responses. *European Urology*, *79*(2), 214–221. https://doi.org/10.1016/j.eururo.2020.08.024.
75. Gil-Jimenez, A., Van Dorp, J., Contreras-Sanz, A., Van Der Vos, K., Vis, D. J., Braaf, L., Broeks, A., Kerkhoven, R., Van Kessel, K. E. M., Ribal, M. J., Alcaraz, A., Wessels, L. F. A., Seiler, R., Wright, J. L., Mengual, L., Boormans, J., Van Rhijn, B. W. G., Black, P. C., & Van Der Heijden, M. S. (2023). Assessment of Predictive Genomic Biomarkers for Response to Cisplatin-based Neoadjuvant Chemotherapy in Bladder Cancer. *European Urology*, *83*(4), 313–317. https://doi.org/10.1016/j.eururo.2022.07.023.
76. Murphy, N., Shih, A. J., Shah, P., Yaskiv, O., Khalili, H., Liew, A., Lee, A. T., & Zhu, X.-H. (2022). Predictive molecular biomarkers for determining neoadjuvant chemosensitivity in muscle invasive bladder cancer. *Oncotarget*, *13*, 1188–1200. https://doi.org/10.18632/oncotarget.28302.
77. Font, A., Domenech, M., Ramirez, J. L., Marqués, M., Benítez, R., Ruiz De Porras, V., Gago, J. L., Carrato, C., Sant, F., Lopez, H., Castellano, D., Malats, N., Calle, M. L., & Real, F. X. (2023). Predictive signature of response to neoadjuvant chemotherapy in muscle-invasive bladder cancer integrating mRNA expression, taxonomic subtypes, and clinicopathological features. *Frontiers in Oncology*, *13*, 1155244. https://doi.org/10.3389/fonc.2023.1155244.
78. Lerner, S. P., McConkey, D. J., Tangen, C. M., Meeks, J. J., Flaig, T. W., Hua, X., Daneshmand, S., Alva, A. S., Lucia, M. S., Theodorescu, D., Goldkorn, A., Milowsky, M. I., Choi, W., Bangs, R., Gustafson, D. L., Plets, M., & Thompson, I. M. (2024). Association of Molecular Subtypes with Pathologic Response, PFS, and OS in a Phase II Study of COXEN with Neoadjuvant Chemotherapy for Muscle-invasive Bladder Cancer. *Clinical Cancer Research*, *30*(2), 444–449. https://doi.org/10.1158/1078-0432.CCR-23-0602.