Universiteit
Leiden
The Netherlands

On preoperative systemic treatment of muscle-invasive bladder cancer

Dorp, J. van

Citation
Dorp, J. van. (2025, January 10). *On preoperative systemic treatment of muscle-invasive bladder cancer*. Retrieved from https://hdl.handle.net/1887/4175499

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4175499

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 3

Assessment of predictive genomic biomarkers for response to cisplatin-based neoadjuvant chemotherapy in bladder cancer

Alberto Gil-Jimenez*, Jeroen van Dorp*, Alberto Contreras-Sanz, Kristan van der Vos, Daniel J. Vis, Linde Braaf, Annegien Broeks, Ron Kerkhoven, Kim E.M. van Kessel, María José Ribal, Antonio Alcaraz, Lodewyk F.A. Wessels, Roland Seiler, Jonathan L. Wright, Lourdes Mengual, Joost Boormans, Bas W.G. van Rhijn, Peter C. Black, Michiel S. van der Heijden

** These authors contributed equally*

European Urology, 2023

ABSTRACT

Cisplatin-based neoadjuvant chemotherapy (NAC) followed by radical cystectomy is recommended for patients with muscle-invasive bladder cancer (MIBC). Somatic deleterious mutations in *ERCC2,* gain-of-function mutations in *ERBB2*, and alterations in *ATM*, *RB1*, and *FANCC* have been shown to correlate with pathological response to NAC in MIBC.

The objective of this study was to validate these genomic biomarkers in pre-treatment transurethral resection (TUR) material from an independent retrospective cohort of 165 MIBC patients who had subsequently undergone NAC and radical surgery. Patients with ypT0/Tis/Ta/T1N0 disease after surgery were defined as responders.

Somatic deleterious mutations in *ERCC2* were found in 9/68 (13%) evaluable responders and in 2/95 (2%) evaluable non-responders (p=0.009, FDR=0.03). No correlation was observed between response and alterations in *ERBB2* or in *ATM*, *RB1* or *FANCC* alone or in combination. In an exploratory analysis, no additional genomic alterations discriminated between responders and non-responders to NAC. No further associations were identified between the aforementioned biomarkers and complete pathological complete response (ypT0N0) after surgery.

In conclusion, we observed a positive association between deleterious mutations in *ERCC2* and pathological response to NAC, but not overall survival or recurrence-free survival. Other previously reported genomic biomarkers were not validated.

PATIENT SUMMARY

It is currently unknown which patients will respond to chemotherapy before definitive surgery for bladder cancer. Previous studies described several gene mutations in bladder cancer that correlated with chemotherapy response. This study confirmed that patients with bladder cancer with a mutation in the *ERCC2* gene often respond to chemotherapy.

INTRODUCTION

Neoadjuvant cisplatin-based chemotherapy (NAC) followed by radical cystectomy is recommended for patients with muscle-invasive bladder cancer (MIBC)[1]. Pathological response after treatment with NAC is strongly associated with recurrence-free survival (RFS) and overall survival (OS)[2]. Currently, clinicians are unable to identify which patients will benefit from NAC. Genomic biomarkers have been described to correlate with response to NAC, including somatic deleterious mutations in *ERCC2,* gain-of-function mutations in *ERBB2*, and alterations in *ATM*, *RB1* and *FANCC*[3-7]. However, none of these biomarkers have been validated in larger independent cohorts and are consequently not used in clinical practice[1,8].

Here, we set out to to validate these genomic biomarkers in an independent multicenter retrospective cohort. Pre-treatment tissue derived from five centers was sequenced at the Netherlands Cancer Institute (NKI cohort, n=117) or Vancouver Prostate Centre (Vancouver cohort, n=48, Supplementary Figure 1). All patients were diagnosed with MIBC (cT2-4aN0M0 and/or cT1-4aN1-3M0) by transurethral resection (TUR) and were treated with at least two cycles of cisplatin-based NAC, followed by radical cystectomy. The primary endpoint of this study was pathological response, defined as ypT0/Tis/Ta/T1N0 after surgery[2,9]. Seventy of 165 patients (42%) were categorized as responders. Pathological complete response after surgery, defined as ypT0N0, was used as a secondary endpoint which was observed in 51 of 165 patients (31%).

RESULTS

Baseline age, gender, chemotherapy regimen, and number of cycles of chemotherapy did not differ between response groups, however cT-stage at baseline was higher in the non-responders (Table 1). Furthermore, baseline cT-stage and chemotherapy regimen differed between cohorts (Supplementary Table 1). Tumor DNA extracted from TUR samples obtained prior to NAC was sequenced using a targeted capture-based panel for the NKI cohort and whole exome sequencing for the Vancouver cohort. Somatic variants in *ERCC2*, *ERBB2*, *ATM*, *RB1* and *FANCC* were inferred from population databases (Supplementary Methods). Mutations were predicted to be functional (deleterious or gain-of-function) using the annotation databases OncoKB, ClinVar, SIFT, FATHMM, and PolyPhen-2 (Supplementary Methods). A high concordance between the observed and the TCGA mutation rates was observed (Supplementary Table 2).

After filtering, deleterious mutations in *ERCC2* were found in nine of 68 (13%) evaluable responders and in two of 95 (2%) evaluable non-responders (p=0.009, Figure 1A). We found relevant gain-of-function mutations in *ERBB2* in nine of 69 (13%) evaluable responders and five of 95 (5%) evaluable non-responders (p=0.09, Figure 1A). Twenty-seven of 70 (39%) responders had ≥1 alteration in *ATM*, *RB1* or *FANCC* compared to 25 of 95 (26%) non-responders (p=0.13, Figure

1A). Nine of eleven patients (82%) with a deleterious mutation in *ERCC2* had a pathological response after treatment with NAC, as opposed to 62 of 154 patients (40%) without any relevant mutations in *ERCC2* (Supplementary Table 3). After correction for multiple hypothesis testing (three hypotheses), mutations in *ERCC2* were significantly enriched in responders (FDR=0.03, Figure 1B). The association remained when adjusted for cT-stage in a multivariable logistic regression model (p_{ERCC2}=0.008, p_{cT2}=p_{cT3}=p_{cT4}>0.9), or when patients that received less than 3 cycles of NAC were excluded (Supplementary Figure 2). Baseline clinical difference between *ERCC2* mutated and wild-type patients were not identified (Supplementary Table 3).

Table 1 | Baseline characteristics and response of 165 patients with muscle-invasive bladder cancer treated with neoadjuvant chemotherapy and radical cystectomy.

	NKI		Vancouver		
	Responders (ypT0/Tis/Ta/T1N0)	Non-responders (≥ypT2N0)	Responders (ypT0/Tis/Ta/T1N0)	Non-responders (≥ypT2N0)	Significance Responders vs Non-responders (full cohort)[a]
Number of patients	53	64	17	31	-
Median age in years (IQR)	71.0 (61.0, 75.0)	71.0 (61.0, 77.3)	61.2 (56.0, 66.0)	65.5 (58.3, 73.0)	0.2
Male sex (% of pts)	40 (76%)	38 (59%)	15 (88%)	24 (77%)	0.08
cT-stage (% of pts)					0.04
cT1	1 (2%)	0 (0%)	0 (0%)	0 (0%)	
cT2	27 (51%)	19 (30%)	2 (12%)	4 (13%)	
cT3	21 (40%)	26 (40%)	7 (41%)	20 (65%)	
cT4	4 (7%)	19 (30%)	8 (47%)	7 (22%)	
cN-stage (% of pts)					0.6
cN0	31 (59%)	40 (63%)	10 (59%)	12 (39%)	
cN+	22 (41%)	24 (37%)	7 (41%)	19 (61%)	
Chemotherapy regimen (% of pts)					0.16
Cis/Gem	40 (75%)	41 (64%)	14 (82%)	28 (90%)	
MVAC	11 (21%)	23 (36%)	3 (18%)	3 (10%)	
CMV	2 (4%)	0 (0%)	0 (0%)	0 (0%)	
Chemotherapy cycles received (% of pts)					0.8
2	2 (4%)	2 (3%)	0 (0%)	2 (6%)	
3	10 (19%)	13 (20%)	6 (35%)	7 (23%)	
4	39 (74%)	46 (72%)	10 (59%)	17 (55%)	
>4	2 (4%)	3 (5%)	1 (6%)	5 (16%)	
Pathological response (% of pts)					-
ypT0N0 (complete response)	40 (75%)	0 (0%)	13 (76%)	0 (0%)	
ypTis/Ta/T1N0	13 (25%)	0 (0%)	4 (24%)	0 (0%)	
≥ypT2N0 (non-response)	0 (0%)	64 (100%)	0 (0%)	31 (100%)	

[a] Fisher's exact test for binary predictors, t-test for numerical predictors. All the statistical tests were two-sided. No adjustments were made for multiple hypothesis testing. Significant associations are highlighted on bold. *Cis/Gem = cisplatin + gemcitabine; MVAC = methotrexate + vinblastine + docorubicin + cisplatin; CMV = cisplatin + methotrexate + vinblastine; IQR = interquartile range.*

In contrast, alterations in *ERCC2*, *ERBB2*, or in any one of *ATM*, *RB1* or *FANCC* did not associate with a pathological complete response (ypT0N0) after correcting for multiple hypothesis testing (FDR_{ERCC2}=0.09, FDR_{ERBB2}=0.07, $FDR_{ATM/RB1/FANNC}$=0.07, Supplementary Figure 3).

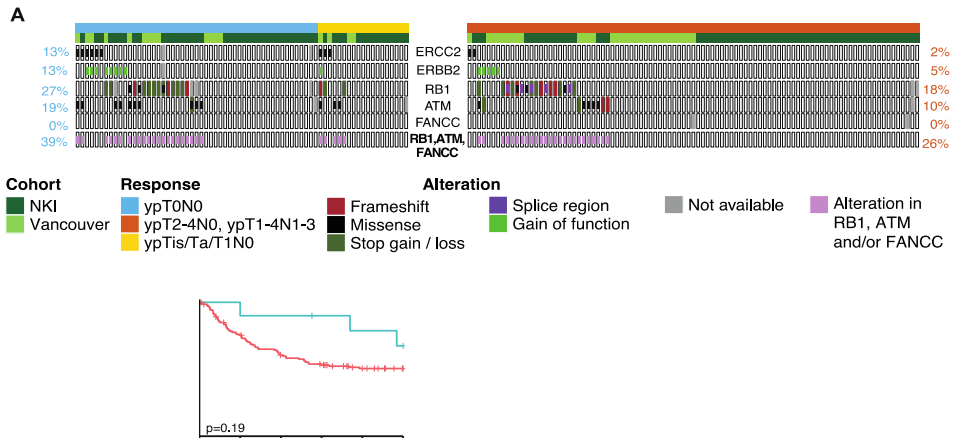


Figure 1 | Somatic mutations in the genes *ERCC2*, *ERBB2, ATM, RB1 and FANCC* in patients with muscle-invasive bladder cancer treated with neoadjuvant chemotherapy. A, Overview of relevant mutations for each patient (**Supplementary Methods - Variant calling and inference of somatic and functional variants**). Left panel shows patients with ypT0N0 (light blue, n=51) or ypTis/Ta/T1N0 (yellow, n=19) after neoadjuvant chemotherapy (responders), and right panel shows non-responders (orange, n=95). Percentages represent the number of patients with a relevant mutation relative to the total number of eligible patients for that specific gene for responders (left) and non-responders (right). Patients with an alteration in any one of *ATM*, *RB1* or *FANCC* are indicated in the last row. **B,** Five-year overall survival (OS) for patients with (blue) and without (red) mutations in *ERCC2.* The p-value indicates statistical significance by a log-rank test. **C,** 5-year recurrence-free survival (RFS) for patients with (blue) and without (red) mutations in *ERCC2*. The p-value inside the survival plot indicates statistical significance by a log-rank test. *NKI = patients from the NKI-cohort; wt = wild type; mut = mutant; Not available = gene-coverage below 20 reads.*

The median duration of follow-up for patients using reverse censoring was 7.2 years. The 5-year OS rates for patients with and without mutations in *ERCC2* were 75% (95% confidence interval (CI): 50%-100%) and 52% (95% CI: 45%-62%), respectively (p=0.2, Figure 1C). The 5-year RFS rates were 65% (95% CI: 39%-100%) and 49% (95% CI: 42%-59%), respectively (p=0.2, Figure 1D). Thus, while the Kaplan Meier curves appear to separate according to *ERCC2* mutation status, we could not demonstrate a statistical difference for either OS or RFS, possibly due to the low frequency of *ERCC2* mutations.

Following earlier analyses by Plimack and colleagues[6], we assessed copy number alterations (CNA) for *ATM*, *RB1,* and *FANCC* by shallow whole genome sequencing for patients from the NKI cohort (n=117, Supplementary Methods). CNA on the Vancouver cohort could not be confidently assessed

due to a lack of germline data. We found seven CNA in *ATM*, *RB1* and/or *FANCC* in all evaluable patients. Together with the previously described mutations, 22 of 53 (42%) responders had ≥1 alteration in *ATM*, *RB1* or *FANCC* versus 20 of 64 (31%) non-responders (p=0.052, Supplementary Figure 4).

In a further exploratory analysis, mutations frequently occurring in MIBC were assessed for their correlation with response to NAC (Supplementary Figure 5). This analysis included *FGFR3*, which was previously associated with negative outcome after chemotherapy (Supplementary Figure 6)[10]. No association with response was identified after correction for multiple hypothesis testing (Supplementary Figure 5).

DISCUSSION

There are several limitations to this study. The genomic data were derived using different sequencing technologies at different centers, leading to potential biases in the mutation frequency. Furthermore, we lacked germline data and somatic variants were filtered with the help of population databases to remove benign germline variants. As germline DNA is often unavailable, this approach is common practice and was also used in the original studies of *ERBB2* and *ATM*/*RB1*/*FANCC*[4,6]. Multiple definitions of response have been used in previous studies, thus introducing heterogeneity between studies. Complete pathological response (ypT0N0) and pathological downstaging (ypT0/Tis/Ta/T1N0) are commonly used. Long-term clinical outcome is favorable in both groups, although patients with ypT0/TisN0 may have a modest survival benefit over patients with ypT0/Tis/Ta/T1N0[2,9].

In summary, we attempted to validate mutations in *ERCC2*, *ERBB2*, *ATM, RB1* and *FANCC* as predictive markers of pathological response in a cohort of 165 patients treated with NAC. We confirmed a positive association of deleterious mutations in *ERCC2* with pathological response (ypT0/Tis/Ta/T1N0), but not with complete response (ypT0N0), OS or RFS. Prospective evaluation of *ERCC2* mutations as a biomarker for response to NAC is needed to confirm our results.

ACKNOWLEDGMENTS

We acknowledge the Genomics Core Facility (Roel Kluin), the Core Facility Molecular Pathology and Biobanking, and the Research High-Performance Computing facility, all at the Netherlands Cancer Institute; the sequencing facility at the Vancouver Prostate Center, and all the clinical teams involved in the 5-center patient cohort.

REFERENCES

1. Witjes, J.A., *et al.* European Association of Urology Guidelines on Muscle-invasive and Metastatic Bladder Cancer: Summary of the 2020 Guidelines. *Eur Urol* 79, 82-104 (2021).
2. Rosenblatt, R., *et al.* Pathologic downstaging is a surrogate marker for efficacy and increased survival following neoadjuvant chemotherapy and radical cystectomy for muscle-invasive urothelial bladder cancer. *Eur Urol* 61, 1229-1238 (2012).
3. Van Allen, E.M., *et al.* Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov* 4, 1140-1153 (2014).
4. Groenendijk, F.H., *et al.* ERBB2 Mutations Characterize a Subgroup of Muscle-invasive Bladder Cancers with Excellent Response to Neoadjuvant Chemotherapy. *Eur Urol* 69, 384-388 (2016).
5. Li, Q., *et al.* ERCC2 Helicase Domain Mutations Confer Nucleotide Excision Repair Deficiency and Drive Cisplatin Sensitivity in Muscle-Invasive Bladder Cancer. *Clin Cancer Res* 25, 977-988 (2019).
6. Plimack, E.R., *et al.* Defects in DNA Repair Genes Predict Response to Neoadjuvant Cisplatin-based Chemotherapy in Muscle-invasive Bladder Cancer. *Eur Urol* 68, 959-967 (2015).
7. Miron, B., *et al.* Defects in DNA Repair Genes Confer Improved Long-term Survival after Cisplatin-based Neoadjuvant Chemotherapy for Muscle-invasive Bladder Cancer. *Eur Urol Oncol* 3, 544-547 (2020).
8. Becker, R.E.N., *et al.* Clinical Restaging and Tumor Sequencing are Inaccurate Indicators of Response to Neoadjuvant Chemotherapy for Muscle-invasive Bladder Cancer. *Eur Urol* 79, 364-371 (2021).
9. Ravi, P., *et al.* Optimal pathological response after neoadjuvant chemotherapy for muscle-invasive bladder cancer: results from a global, multicentre collaboration. *BJU Int* 128, 607-614 (2021).
10. Teo, M.Y., *et al.* Fibroblast Growth Factor Receptor 3 Alteration Status is Associated with Differential Sensitivity to Platinum-based Chemotherapy in Locally Advanced and Metastatic Urothelial Carcinoma. *Eur Urol* 78, 907-915 (2020).

SUPPLEMENTARY METHODS

STUDY POPULATION / TREATMENT

The full cohort consisted of 165 prospectively collected samples from five different centers. All patients had muscle-invasive bladder cancer (MIBC, cT2-4aN0M0 and/or cT1-4aN1-3M0) diagnosed by transurethral resection (TUR) and treated with at least two cycles of neoadjuvant chemotherapy (NAC) followed by radical cystectomy. NAC consisted of either cisplatin + gemcitabine (cis/gem), methotrexate + vinblastine + doxorubicin + cisplatin (MVAC) or cisplatin + methotrexate + vinblastine (CMV).

Patient cohorts were named after the center in which the sequencing was performed. The NKI cohort (n=117) included retrospectively collected MIBC samples from three centers: Amsterdam (The Netherlands), Rotterdam (The Netherlands), and Barcelona (Spain) and the Vancouver cohort (n=48) included MIBC samples from two institutions: Bern (Switzerland) and Seattle (Washington, USA) were compiled (Supplementary Figure 1, Table 1).

TARGETED DNA SEQUENCING (NKI COHORT)

Formalin-fixed paraffin-embedded (FFPE) tumor blocks from TUR material were collected from the different hospitals and centrally reviewed by an experienced pathologist. Tumor area was marked for every tumor block and DNA was collected from subsequent FFPE slides (10 µm). NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was used to quantify the total amount of DNA, and Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was used to quantify the amount of double stranded DNA. Covaris Focused ultrasonicator was used to fragment up to 2000 ng of double stranded genomic DNA to get fragment sizes of 200-300 bp (Covaris, Woburn, Massachusetts, USA). 2X AMPure XP PCR purification beads were used to purify samples following the manufacturer's instructions (Beckman Coulter, Indianapolis, Indiana, USA). The sheared DNA samples were qualified and quantified using a BioAnalyzer DNA Analysis 7500 Kit (Agilent Technologies, Santa Clara, California, USA). Library preparation for Illumina sequencing was done using the KAPA HTP Library Preparation Kit (Kapa Biosystems, Roche Sequencing and Life Science, Wilmington, Massachusetts, USA) with an input of maximum 1 µg sheared DNA. Four PCR cycles were done during library amplification to obtain sufficient yield for the exome capture. Libraries were cleaned up using 1X AMPure XP PCR purification beads (Beckman Coulter, Indianapolis, Indiana, USA). The DNA libraries were analyzed on a Bioanalyzer system using the BioAnalyzer DNA Analysis 7500 Kit chips to determine the concentration. With 150 ng of each indexed sample three pools of eight samples were prepared, and 2 µl of Universal blockers - TS Mix (Integrated DNA Technologies, Coralville, Iowa, USA) and 5 µl Human Cot-1 DNA (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was added to the pools. Then, a concentrator was used to dry the pool, and to rehydrate 8.5 µl of Hybridization buffer, 3.4 µl Hybridization component A (SeqCap Hybridization and Wash Kit, Roche, Basel, Switzerland) and 1.1 µl nuclease-free water was added. The pool underwent incubation at room

temperature for 10 minutes, and at 96 °C for 10 minutes. The samples were hybridized with 4 µl of the custom 44 gene bladder cancer panel (which included muscle-invasive bladder cancer driver genes, clinically relevant genes, and frequently mutated genes) at 65 °C for 24 hours. The hybridized sample pool was captured according to the Rapid protocol for DNA probe Hybridization and Target Capture using an Illumina TruSeq or Ion Torrent Library and subsequently amplified using 10 PCR cycles (Integrated DNA Technologies, Coralville, Iowa, USA). The amplified pool was purified using AMPure XP PCR purification beads (Beckman Coulter, Indianapolis, Indiana, USA). The purified pools were quantified on the Agilent BioAnalyzer DNA Analysis 7500 system and one sequence pool was made by equimolar pooling (Agilent Technologies, Santa Clara, California, USA). The sequence pool was diluted to a final concentration of 10 nM and subjected to sequencing with a MID 150 cycle kit for a paired end 75 bp run following manufacturer's instructions on an Illlumina Nextseq 550 machine (Illumina, San Diego, California, USA).

Sequencing reads were aligned against the Human Reference Genome Ghr38 using Burrows-Wheeler aligner v0.7.17-r1188 (Li H. and Durbin R. (2010), Bioinformatics, 25:1754-60). Duplicated reads were marked and removed using GATK MarkDuplicates v4.1.1.0 (Broad Institute, Cambridge, Massachusetts, USA), and base quality score recalibration was done using GATK ApplyBQSR v4.1.1.0 (Broad Institute, Cambridge, Massachusetts, USA). Indel realignment was not performed as per current GATK best practices (June 2021) it is not recommended when performing variant calling with Mutect2.

EXOME DNA SEQUENCING (VANCOUVER COHORT)

To extract DNA from FFPE tumor samples, two FFPE cores per case were used to prepare hematoxylin and eosin-stained slides and a certified pathologist reviewed them for tumor content. Macro-dissection on tumor regions was done to enrich for tumor content. For paraffin removal, tissue re-hydration, tissue digestion and DNA extraction a M220 Focused ultrasonicator and a truXTRAC FFPE DNA microTUBE Kit (both from Covaris, Woburn, Massachusetts, USA) wered used. To quantify DNA the Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was used. The DNA quality was assessed by multiplex PCR assay with usable DNA in samples with >400 bp PCR products.

To sequence the samples, 150-200 bp fragments were generated by fragmenting 1 µg of genomic DNA by hydrodynamic shearing using a E210 Focused ultrasonicator (Covaris, Woburn, Massachusetts, USA). DNA fragments were ligated to Illumina barcoded adapters. PCR amplification was used to enrich for adapter ligated fragments. To quantify DNA the Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was used. The library was enriched by liquid phase hybridization using Agilent SureSelect XT Human All Exon v6 (Agilent Technologies, Santa Clara, California, USA) according to the manufacturer's recommendations, and amplified by PCR using indexing primers. Captured libraries were cleaned and controlled for quality using the Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). For clustering

of index-coded samples and sequencing of PE100 libraries an Illlumina HiSeq 4000 machine (Illumina, San Diego, California, USA) was used.

Sequencing reads were aligned against the Human Reference Genome Ghr38 using Burrows-Wheeler aligner v0.7.17-r1188. Duplicated reads were marked and removed using MarkDuplicates v2.23.8, and base quality score recalibration was done using GATK ApplyBQSR v4.1.1.0.

VARIANT CALLING AND INFERENCE OF SOMATIC AND FUNCTIONAL VARIANTS

GATK Mutect2 v4.1.9.0 (Broad Institute, Cambridge, Massachusetts, USA) was used to call single-nucleotide variants (SNVs) and short insertions and deletions (indels) on tumor samples using the 'Tumor-only' mode. Variants were further filtered using GATK FilterMutectCalls v4.1.9.0 (Broad Institute, Cambridge, Massachusetts, USA), and variants with an allele frequency below 5% or an alternate number or reads below 3 were filtered out.

Because the sequencing depth between samples and genes was variable, we annotated regions with a low sequencing coverage. First, we computed the sequencing depth using Samtools v1.9 (Sanger Institute, Hinxton, UK) and estimated the average coverage per-gene. For each sample, we annotated each gene from the oncoplots (i.e. Figure 1A) as '*Not available*' when the average number of gene sequencing coverage was <20 reads. As an exception to this rule, when a relevant (non-germline and pathogenic) variant and with at least 3 alternate reads was detected for a low coverage gene, the variant was subjected to downstream analysis and the '*Not available*' annotation was removed.

For downstream analysis, samples showing low coverage (average <20 reads) in >40% of the studied genes were filtered out, which affected 4 patients from the NKI cohort, and 2 patients from the Vancouver cohort. Variants were processed in R 3.6.0 using the packages VariantAnnotation v1.24.5 (Obenchain, V. et al. (2014), Bioinformatics, 30(14), 2076-2078), ComplexHeatmap v1.17.1 (Gu, Z. et al. (2016), Bioinformatics), tidyverse 1.2.1 (Wickham, H. et al. (2019), Journal of Open Source Software, 4(43)), and ggpubr 0.2.1 (Kassambara, A. (2020), CRAN).

Variants were annotated using SnpEff 4.3t (Cingolani, P. (2012), Fly, 6:80-92) against the Human Reference Genome Ghr38, and using SnpSift (Cingolani, P. (2012), Frontiers in Genetics, 3) against the databases dbSNP (build 151), COSMIC (v85), ClinVar, dbNSFP, GNOMAD (version exomes. r2.1.1.sites), and OncoKB (version MVL2.sorted). To filter out germline variants, only variants annotated as COMMON != 1, CAF <0.05, TOPMED <0.05, GNOMAD dbNSFP_gnomAD_exomes_NFE_AF <0.05. and dbNSFP_gnomAD_exomes_AF <0.05 were retrieved. To further filter out potential germline variants, we retrieved the germline DNA sequencing data collected from NABUCCO cohort 1, and identified the germline variants FOXQ1 T60P, KMT2D T4629P (mutated in >70% of the patients), which were filtered out from our list.

To retrieve functional variants, we filtered out variants annotated as introns, non-coding, synonymous, downstream gene variant, 3' UTR variant, 5' UTR variant, t' UTR premature start codon gain, sequence feature. Then, we only retrieved variants being annotated with at least one of the following annotations: Pathogenic or Likely Pathogenic on CLNSIG, High Impact, Deleterious on SIFT or FATHMM, Damaging or Probably Damaging on HVAR, or as Oncogenic or Likely Oncogenic on OncoKB. For ERBB2, only variants annotated as Gain of Function by OncoKB were retrieved and reported in the manuscript.

SHALLOW WHOLE-GENOME DNA SEQUENCING FOR COPY NUMBER (NKI COHORT)

For shallow genome sequencing, the protocol up to PCR amplification was analogous to the one indicated in 'DNA sequencing'. The uniquely indexed samples were mixed together by equimolar pooling. Different pools of samples were prepared, consisting of 5 batches of 12 (1 pool, 1 lane), and one batch of 34 samples (1 pool, 3 lanes). The pools were analyzed on the Agilent Technologies 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). The pools were diluted to 10 nM, and measured using qPCR. Finally, the pools were sequenced in a single-end 65 bp run, following the manufacturer's instructions on an Illumina HiSeq2500 machine (Illumina, San Diego, California, USA).

The low coverage whole genome samples were aligned to GRCh38.78 using the Burrows-Wheeler Aligner mem algorithm (bwa version 0.7.17) (Li H. and Durbin R. (2010), Bioinformatics, 25:1754-60). For every sample, and on segments of 20 kb on the genome, the mapping quality read counts were rated and tiled for 65 base pairs against a similar mapping of all known sequences for genome version GRCh38.78. A non-linear loess fit of mappabilities over 0.8 on autosomes was used to correct per 20 kb for local GC effects. Then a scaling to the slope of a linear fit of the reference mappabilities after GC correction was done on the reference values. where the intercept was forced at the origin. We filtered out the ratios of corrected sample counts and the reference values left out bins with a mappability below 0.2 or overlapped with ENCODE blacklisted regions (ENCODE Project Consortium, Nature, 2012). The pipeline used in the count and log2ratio corrections is available at https://github.com/NKI-GCF/SeqCNV. For male patients, the $\log_2$ ratio was increased by one in genes from Chromosome X. Copy number ratios (CNR) at a gene level (CNR-gene) were estimated using a weighted average copy number ratio per gene. We defined deletions as $\log_2$(CNR-gene) <−0.7, and amplifications as $\log_2$(CNR-gene) >1.

TCGA COHORT

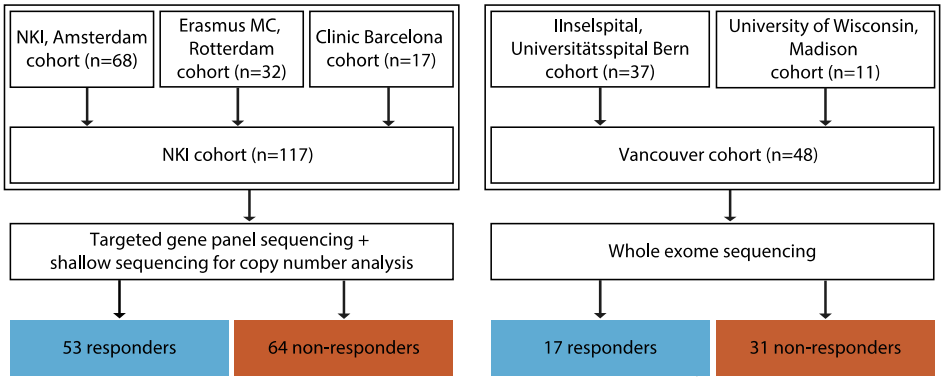
Mutation data from Muscle-invasive bladder cancer patients from the TCGA cohort (n=412, Robertson *et al*., Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer, Cell, 2017) was downloaded as provided on cBioportal on 8th April 2022. Mutation data was aggregated by patient to compute mutation rates, and compared with mutation rates from our cohort by a two-sided Fisher's exact test.

STATISTICAL ANALYSES

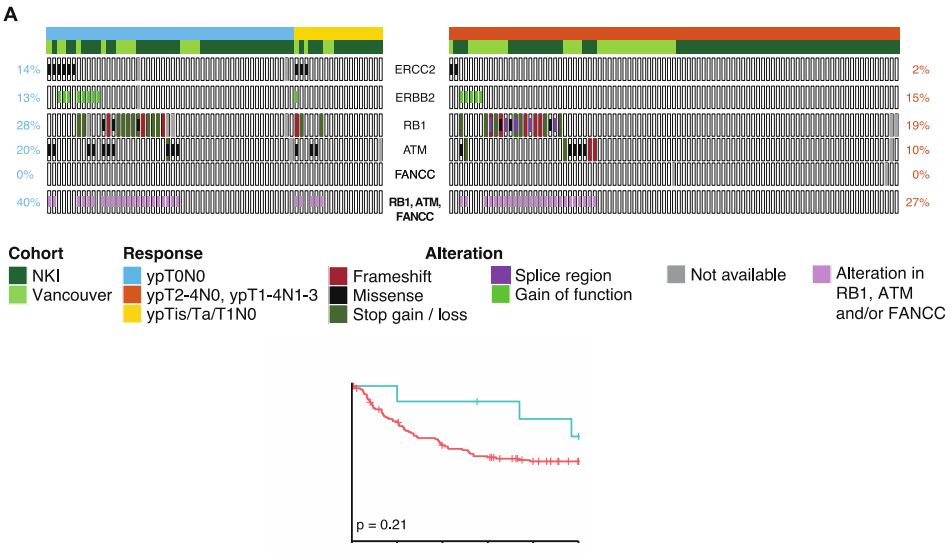
Associations between genomic mutations and clinical response were tested using a two-sided Fisher's exact test. We performed multiple hypothesis testing on our original set of 3 hypotheses (*ERCC2*, *ERBB2* and *ATM*, *RB1* or *FANCC*) using the Benjamini-Hochberg method. A 0.05 threshold for both the unadjusted p-values and the false discovery rates was used to define significance. Associations between baseline clinical characteristics and response (**Table 1**) were sought using a Fisher's exact test for binary predictors, and a t-test for numerical predictors. Unless otherwise stated, all the statistical tests were two-sided. The statical analysis was performed using R 3.6.0.

Survival analysis with time to event was performed with a Kaplan-Meier analysis, and statistical significance was tested with a log-rank test. Time to event was computed as the temporal window between the day of cystectomy and either time to death (overall survival, OS) or time to recurrence (recurrence-free survival RFS). For median follow-up estimation, reverse censoring was implemented using the reverse Kaplan-Meier method by reversing the event and censoring labels. For OS and RFS analysis, censoring was implemented for patients lost to follow-up. Hazard ratio testing with a Cox proportional hazards regression model was not performed, due to the proportional hazard assumption being violated for a Cox model of the 5-year OS association with *ERCC2* mutations, in which a significant association between the Cox model residuals and time was identified (p=0.037). Analyses were performed using the R packages survival version 2.44 (Therneau, T. M, CRAN), prodlim version 2019.11.13 (Gerds, T. A., CRAN), and survminer version 0.4.6 (Kassambara, A., CRAN).

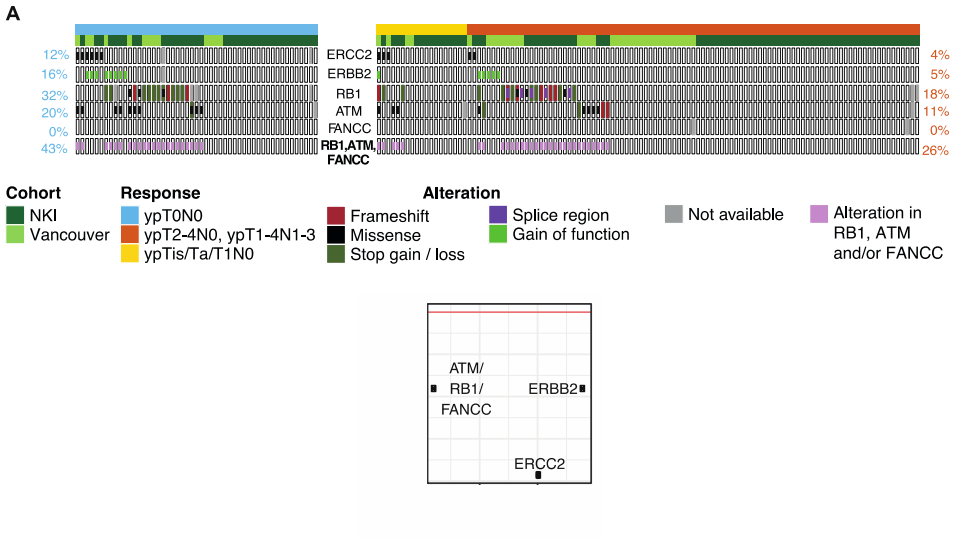
SUPPLEMENTARY DATA



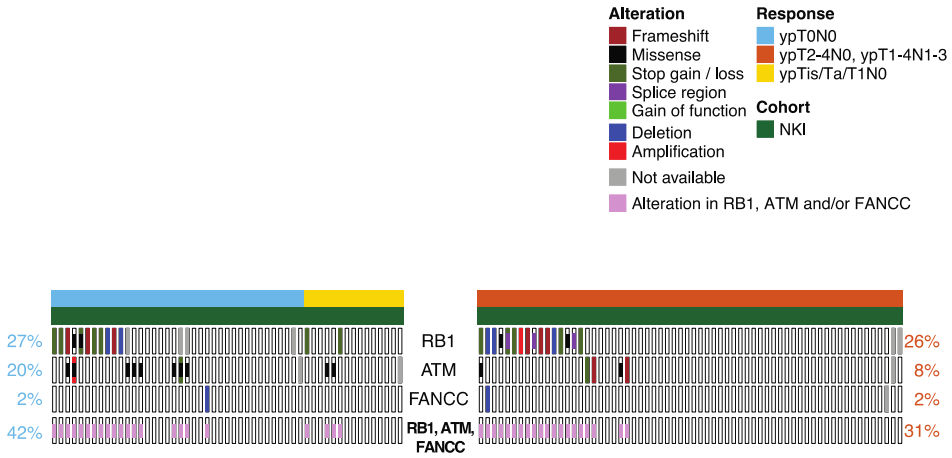
Supplementary Figure 1 | Overview of patient cohorts from 5 different centers. The patients from Amsterdam (The Netherlands), Rotterdam (The Netherlands) and Barcelona (Spain) (n=117) were sequenced by the NKI using targeted DNA sequencing and shallow whole genome sequencing. Whole-exome sequencing was performed in Vancouver on tumor material from Bern (Switzerland) and Seattle (Washington) (n=48).



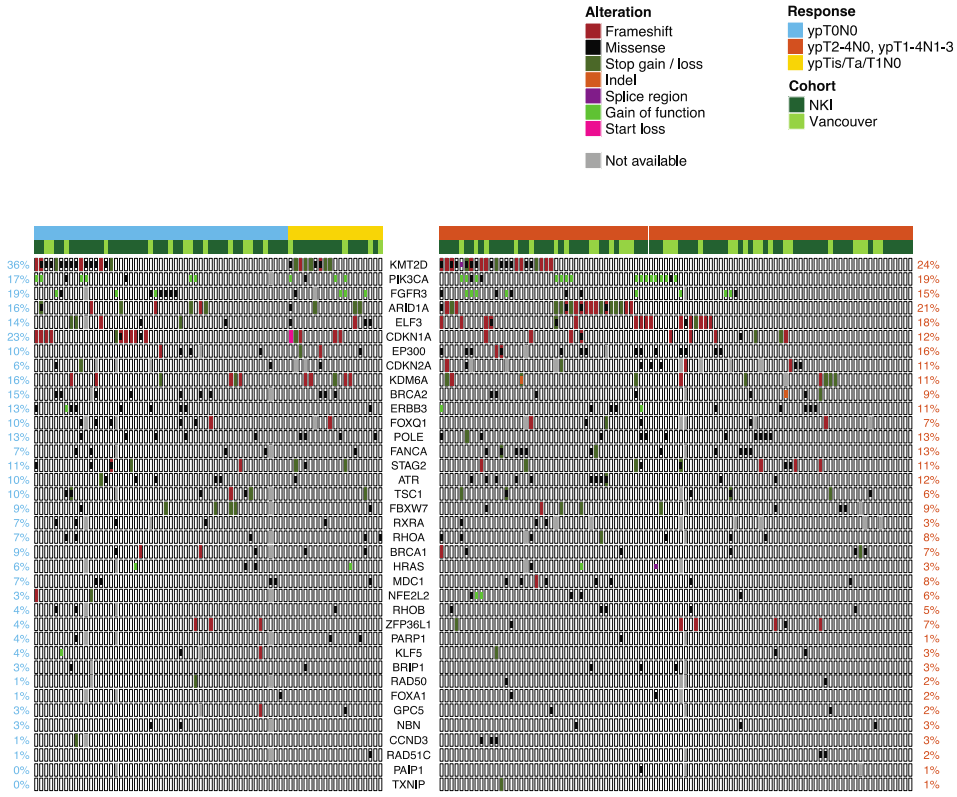
Supplementary Figure 2 | Somatic mutations in the genes *ERCC2*, *ERBB2, ATM, RB1 and FANCC* in patients with muscle-invasive bladder cancer treated with neoadjuvant chemotherapy with at least 3 NAC cycles. Six patients that received 2 cycles of chemotherapy were excluded from the analysis. **A,** Overview of relevant mutations for each individual patient. Left panel shows patients with pathological complete response (ypT0N0, light blue, n=50), and right panel shows patients with ypTis/Ta/T1N0 (yellow, n=18) and ≥ypT2N0 (orange, n=91). Percentages represent the number of patients per cohort with a relevant mutation relative to the total number of eligible patients for that specific gene for ypT0N0 (left) and >ypT0N0 (right) patients. Patients with an alteration in any one of *ATM*, *RB1* or *FANCC* are indicated in the last row. **B,** Adjusted p-value versus odds ratio for *ERCC2*, *ERBB2* and any one of *ATM*, *RB1* or *FANCC* between response groups. P-values were calculated by a two-sided Fisher's exact test, and adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure. The 5-year (**C**) overall survival and (**D**) recurrence-free survival for patients with (blue) and without (red) a mutation in *ERCC2*. The p values indicate statistical significance for a log-rank test. *NKI = patients from the NKI-cohort; Not available = gene-coverage below 20 reads.*



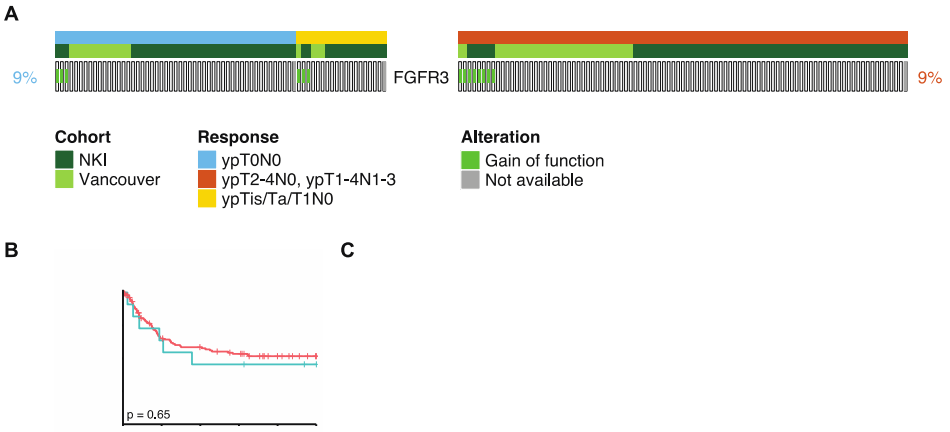
Supplementary Figure 3 | Somatic mutations in the genes *ERCC2*, *ERBB2, ATM, RB1 and FANCC* in patients with muscle-invasive bladder cancer treated with neoadjuvant chemotherapy related to complete pathological response. A, Overview of relevant mutations for each individual patient. Left panel shows patients with pathological complete response (ypT0N0, light blue, n=51), and right panel shows patients with ypTis/Ta/T1N0 (yellow, n=19) and ≥ypT2N0 (orange, n=95). Percentages represent the number of patients per cohort with a relevant mutation relative to the total number of eligible patients for that specific gene for ypT0N0 (left) and >ypT0N0 (right) patients. Patients with an alteration in any one of *ATM*, *RB1* or *FANCC* are indicated in the last row. **B,** Adjusted p-value versus odds ratio for *ERCC2*, *ERBB2* and any one of *ATM*, *RB1* or *FANCC* between response groups. P-values were calculated by a two-sided Fisher's exact test, and adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure. *NKI = patients from the NKI-cohort; Not available = gene-coverage below 20 reads.*



Supplementary Figure 4 | Somatic mutations and copy number alterations in *ERCC2*, *ERBB2, ATM*, *RB1* and *FANCC* in patients from the NKI cohort. Overview of relevant mutations and copy number alterations for all patients for whom mutation data and copy number alterations were available (NKI cohort, n=117). Left panel shows responders (ypT0/Tis/Ta/T1N0), and the right panel shows non-responders. Percentages represent the number of patients with a relevant alteration relative to the number of eligible patients for that specific gene for responders (left) and non-responder (right) patients. Patients with an alterations in any one of *ATM*, *RB1* or *FANCC* are shown separately in the last row. *NKI = patients from the NKI-cohort; Not available = gene-coverage below 20 reads.*



Supplementary Figure 5 | Somatic mutations in oncogenic driver genes and genes frequently-mutated in muscle-invasive bladder cancer. Overview of somatic mutations in driver genes and genes frequently altered in muscle-invasive bladder cancer in responders (ypT0/Tis/Ta/T1N0, left) and in non-responders (right). Adjusted p-value versus odds ratio for the alterations of each gene between response groups. P-values were calculated by a two-sided Fisher's exact test, and adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure. *NKI = patients from the NKI-cohort; Not available = gene-coverage below 20 reads.*



Supplementary Figure 6 | Somatic mutations in the *FGFR3* gene in patients with muscle-invasive bladder cancer treated with neoadjuvant chemotherapy. A, Overview of relevant *FGFR3* mutations for each patient. Left panel shows patients with ypT0N0 (light blue, n=51) or ypTis/Ta/T1N0 (yellow, n=19) after neoadjuvant chemotherapy (responders), and right panel shows non-responders (orange, n=95). Percentages represent the number of patients with a relevant mutation relative to the total number of eligible patients for that specific gene for responders (left) and non-responders (right). **B,** Five-year overall survival (OS) for patients with (blue) and without (red) somatic mutations in *FGFR3*. The p-value indicates statistical significance by a log-rank test. **C,** 5-year recurrence-free survival (RFS) for patients with (blue) and without (red) somatic mutations in *FGFR3*. The p-value inside the survival plot indicates statistical significance by a log-rank test. *NKI = patients from the NKI-cohort; Not available = gene-coverage below 20 reads*

Supplementary Table 1 | Baseline characteristics and response of 165 patients with muscle-invasive bladder cancer treated with neoadjuvant chemotherapy and radical cystectomy and comparison between cohorts.

	NKI Cohort	Vancouver	Significance NKI vs Vancouver[a]
Number of patients	117	48	-
Median age in years (IQR)	71.0 (61.0, 75.0)	62.1 (57.3, 72.0)	0.003
Male sex (% of pts)	78 (67%)	39 (81%)	0.09
cT-stage (% of pts)			0.002
cT1	1 (1%)	0 (0%)	
cT2	46 (39%)	6 (13%)	
cT3	47 (40%)	27 (56%)	
cT4	23 (20%)	15 (31%)	
cN-stage (% of pts)			0.09
cN0	71 (61%)	22 (46%)	
cN+	46 (39%)	26 (54%)	
Chemotherapy regimen (% of pts)			0.03
Cis/Gem	81 (69%)	42 (88%)	
MVAC	34 (29%)	6 (13%)	
CMV	2 (2%)	0 (0%)	
Chemotherapy cycles received (% of pts)			0.15
2	4 (3%)	2 (4%)	
3	23 (20%)	13 (27%)	
4	85 (73%)	27 (56%)	
>4	5 (4%)	6 (13%)	
Pathological response (% of pts)			0.5
ypT0N0 (complete response)	38 (32%)	13 (27%)	
ypTis/Ta/T1N0	15 (13%)	4 (8%)	
≥ypT2N0 (non-response)	64 (55%)	31 (65%)	
Nodal response (% of pts)			0.7
ypN0	81 (69%)	32 (67%)	
≥ypN1	34 (29%)	16 (33%)	
ypN unavailable	2 (2%)	0 (0%)	

[a] Fisher's exact test for binary predictors, t-test for numerical predictors. All the statistical tests were two-sided. No adjustments were made for multiple hypothesis testing. Significant associations are highlighted on bold.
Cis/Gem = cisplatin + gemcitabine; MVAC = methotrexate + vinblastine + docorubicin + cisplatin; CMV = cisplatin + methotrexate + vinblastine; IQR = Interquartile range; pts = patients.

Supplementary Table 2 | Comparison of mutation rates against TCGA Bladder Urothelial Carcinoma (n=412).

Gene	NKI/Vancouver mutation rate (%)	TCGA mutation rate (%)	Significance TCGA vs NKI/Vancouver[a]	Significance (false-discovery rate)[a]
KDM6A	13	33	0.0005	0.02
FANCA	10	4	0.01	0.16
CDKN1A	16	9	0.01	0.16
POLE	13	7	0.02	0.16
TXNIP	1	5	0.02	0.16
FOXQ1	8	3	0.02	0.17
MDC1	8	4	0.06	0.4
ARID1A	19	30	0.15	0.6
ELF3	16	13	0.18	0.6
ATR	11	8	0.19	0.6
FANCC	0	1	0.19	0.6
PAIP1	1	3	0.19	0.6
RHOA	7	5	0.22	0.6
RB1	22	19	0.23	0.6
BRCA1	8	5	0.24	0.6
ERBB2	9	13	0.24	0.6
CDKN2A	9	7	0.28	0.7
CCND3	2	1	0.28	0.7
ERCC2	7	10	0.33	0.7
PIK3CA	18	26	0.36	0.8
KLF5	4	6	0.41	0.8
STAG2	11	15	0.41	0.8
FOXA1	2	3	0.42	0.8
FGFR3	17	15	0.44	0.8
GPC5	2	4	0.46	0.8
BRIP1	3	5	0.50	0.8
BRCA2	11	10	0.54	0.8
NFE2L2	5	7	0.56	0.9
RHOB	5	7	0.69	0.9
EP300	14	17	0.70	0.9
RAD51C	2	1	0.72	0.9
FBXW7	9	8	0.73	0.9
ERBB3	12	11	0.77	0.9
RAD50	2	3	0.77	0.9
NBN	3	3	0.78	0.9
RXRA	5	6	0.84	1
KMT2D	29	34	0.92	1
ATM	14	15	1	1
HRAS	4	4	1	1
PARP1	2	3	1	1
TSC1	8	9	1	1
ZFP36L1	6	6	1	1

[a] Fisher's exact test for frequency of somatic mutations between TCGA and NKI/Vancouver cohorts. P-values were adjusted for multiple hypothesis testing by the Benjamini-Hochberg method.

Supplementary Table 3 | Baseline characteristics and response rates of 163 patients with muscle-invasive bladder cancer treated with neoadjuvant chemotherapy and radical cystectomy and comparison between *ERCC2* mutation status.

	ERCC2$_{mut}$	*ERCC2*$_{wt}$	Significance *ERCC2*$_{mut}$ vs *ERCC2*$_{wt}$ [a]
Number of patients	11	152	
Median age in years (IQR)	57.0 (53.2, 74.0)	69.0 (60.9, 75.0)	0.12
Male sex (% of pts)	8 (72%)	107 (70%)	1
cT-stage (% of pts)			0.7
cT1	0 (0%)	1 (1%)	
cT2	2 (18%)	49 (32%)	
cT3	6 (55%)	67 (44%)	
cT4	3 (27%)	35 (23%)	
cN-stage (% of pts)			0.4
cN0	8 (73%)	84 (55%)	
cN+	3 (27%)	68 (45%)	
Chemotherapy regimen (% of pts)			0.8
Cis/Gem	9 (82%)	112 (74%)	
MVAC	2 (18%)	38 (25%)	
CMV	0 (0%)	2 (1%)	
Chemotherapy cycles received (% of pts)			0.5
2	0 (0%)	6 (4%)	
3	4 (36%)	32 (21%)	
4	6 (54%)	104 (68%)	
>4	1 (9%)	10 (7%)	
Pathological response (% of pts)			0.009
ypT0N0 (complete response)	6 (54%)	43 (28%)	
ypTis/Ta/T1N0	3 (27%)	16 (11%)	
≥ypT2N0 (non-response)	2 (18%)	93 (61%)	

[a] Fisher's exact test for binary predictors. t-test for numerical predictors. All the statistical tests were two-sided. No adjustments were made for multiple hypothesis testing. Significant associations are highlighted on bold.
Cis/Gem = cisplatin + gemcitabine; MVAC = methotrexate + vinblastine + docorubicin + cisplatin; CMV = cisplatin + methotrexate + vinblastine; mut = mutant; wt = wild-type; IQR = Interquartile range; pts = patients.