Supplementary Material

Methods

Clinical trial design and participants

Study participants were recruited from 33 sites across the USA between 2016 and 2019 as part of a phase 3 multicenter, single-arm, open-label, repeat-dose study of intravesical nadofaragene firadenovec therapy for patients with bacillus Calmette-Guérin (BCG)-unresponsive high-risk non–muscle-invasive bladder cancer (NMIBC) [1]. Eligible patients were aged 18 yr or older at the time of providing written informed consent. All patients met the definition of BCG-unresponsive NMIBC [2]. This included patients with persistent carcinoma in situ (CIS) or high-grade Ta/T1 tumors at 6 mo despite receiving adequate BCG therapy, with the latter being defined as at least five of the six induction doses and two of the three maintenance doses of BCG, or at least two of six instillations of a second induction course in which maintenance BCG was not given (previously termed BCG-refractory disease). It also included patients with recurrent high-grade Ta/T1 NMIBC within 6 mo, those with CIS within 12 mo of disease-free state after BCG (previously termed BCG relapse), and those with persistent high-grade Ta or CIS or progression to T1 disease after BCG. Trial patients included in the safety population were categorized into two subcohorts based on the initial diagnosis. The CIS subcohort included patients with CIS with or without concomitant high-grade Ta/T1 NMIBC, and the high-grade Ta/T1 subcohort included patients with high-grade Ta/T1 tumors without concomitant CIS. Patients had an Eastern Cooperative Oncology Group performance status of ≤2 and life expectancy of at least 2 yr.

Detailed study criteria and protocol have been reported previously [1]. Briefly, patients were excluded if they had evidence of upper urinary tract malignancy, urothelial carcinoma within the prostatic urethra, lymphovascular invasion, micropapillary disease, or tumor-associated hydronephrosis. Patients on current systemic therapy or those who had received pelvic external beam radiotherapy within the previous 5 yr for bladder cancer were also excluded. Patients who received immunosuppressive therapy, investigational drugs, or intravesical therapy were subject to a prestudy washout period. The study was done in accordance with the Declaration of Helsinki, in compliance with Good Clinical Practice Guidelines. All study participants provided written informed consent. The study protocol and amendments were approved by an institutional review board at each participating center.

Procedures, sample acquisition, and study endpoint

Detailed on-protocol procedures have been described previously [1]. Briefly, all visible bladder tumors were resected at enrolment, and patients with T1 tumors on transurethral resection underwent a re-resection 14–60 d before first treatment with intravesical nadofaragene firadenovec. Obvious areas of CIS were also fulgurated before beginning study treatment, which was standard practice before BCG administration at all centers.

In addition to standard screening procedures, 6 ml whole blood was withdrawn through a peripheral venipuncture from enrolled patients 1–24 h prior to administration of the first therapy dose (at month 0) for the determination of baseline anti–human adenovirus type-5 (anti–HAdV-5) antibody levels. All patients then received 75 ml nadofaragene firadenovec (3×10^{11} VP/ml) intravesically with a dwell time of 1 h. Appropriate supportive care including the use of pretreatment anticholinergic agents to minimize bladder irritation was permitted as necessary.

Patients were evaluated for recurrence with urine cytology and cystoscopy (with biopsy if warranted) at efficacy assessment visits every 3 mo (Supplementary Fig. 1). Investigators were not required to use advanced cystoscopic modalities for evaluation, but, if used, they were advised to use the same modality consistently at screening and at each efficacy assessment for an individual patient. In the absence of high-grade recurrence, nadofaragene firadenovec administration was repeated at month 3 (day 90), month 6 (day 180), and month 9 (day 270), with a peripheral blood sample withdrawn prior to initiating each treatment. Patients with recurrence of high-grade disease at month 3, 6, or 9 were removed from the study. Patients who experienced recurrence of high-grade disease at month 3 were not included for this analysis as they a priori did not meet the definition of any durable response. All patients who were not withdrawn from treatment at earlier time points had an efficacy assessment at 12 mo (day 365) after the first dose of treatment. Anti–HAdV-5 antibody levels were also determined for those patients who remained free of high-grade recurrence at 12 mo. In addition to serum anti–HAdV-5 assessment, urine samples were also collected for exploratory assessment of cytokine levels (TRAIL, interleukin [IL]-6, and IL-17); the latter analyses are currently underway.

The primary endpoint for this analysis was response to intravesical nadofaragene firadenovec at 12 mo. This was defined as absence of high-grade disease recurrence by cystoscopy, cytology, and biopsy examination (if clinically indicated or mandated) at all evaluation visits until 12 mo after the administration of the first dose. Specifically, this either included no evidence of progression to CIS with Ta or T1 lesions, or showed evidence of Ta or T1 lesions without CIS that were low grade.

Antibody titer determination

Elevated anti–HAdV-5 antibody levels have been associated with a response to adenovirus-mediated gene therapy in high-grade gliomas [3]. To determine any such association in the setting of intravesical nadofaragene firadenovec therapy for BCG-unresponsive NMIBC, peripheral blood samples were obtained at each time point and allowed to clot at room temperature for 30 min followed by centrifugation at 1500*g* for 15 min. The resulting serum was frozen at –70°C and shipped from individual sites to a central laboratory for analysis.

Detection and titration of anti–HAdV-5 antibodies in serum were performed using a quantitative enzyme-linked immunosorbent assay. The 96-well microtiter plates were coated with virions of rAd-IFNα2b that comprised nucleocapsids of HAdV-5 as the antigen-containing moiety. After washing, the coated plate was blocked with casein in phosphate-buffered saline to minimize nonspecific binding. Samples were serially diluted in blank matrix and added to the plates. Incubation was performed to allow binding of any antibodies present in the samples to the immobilized antigen. Unbound material was washed away and bound anti–HAdV-5 was detected with protein A/G-horseradish peroxidase conjugate, which was visualized by color development with 3,3′,5,5′-tetramethylbenzidine substrate. Rabbit-derived antiadenovirus type-5 antibody (Abcam, Cambridge, UK) served as positive control.

Antibody levels were estimated by a titration assay. Since endogenous anti–HAdV-5 activity in the normal population is too high for creation of a matrix negative control, a normal pool was created and titrated to a dilution that crossed the cut point determined using depleted serum spiked with the positive control at the dilution determined for the cut point control (CPC; 1:75 000). Briefly, a plate-based cut point corresponding to the average optical density of the CPCs was used for an individual plate. A CPC that used pooled IgG-stripped human serum spiked with positive

control to 1:75 000 dilution generated a “low responding control,” which generated a response three to four times lower than that of the low positive control, but still within a reproducible range of the instrument.

Cut point = average signal of CPCs on that plate

The rationale for assigning plate-specific cut point was to ensure that any fluctuations in responses across the plate due to assay or analytical variation would be observed in the CPCs, positive controls, and validation samples. Titers were determined as a reciprocal of the value calculated by multiplying the minimum required dilution (1 in 10) by the highest serial dilution at which the corresponding response was greater than or equal to the plate-specific cut point.

Statistical analysis

Overall analyses were performed on the entire study population, and in the CIS and high-grade Ta/T1 subcohorts where appropriate. Antibody fold change levels for individual patients were calculated as the ratio of titers at a post-treatment time point to baseline. The peak titer was defined as the highest post-treatment anti–HAdV-5 antibody level for individual patients during the course of their therapy. Higher antibody titers and fold change levels were correlated with durable treatment response. Test performance metrics included sensitivity, specificity, accuracy, and positive and negative predictive values. Optimal antibody titer and fold change cutoffs were determined by maximizing the area under the receiver operating characteristic curve.

Differences between continuous variables were assessed by the Mann-Whitney U test and visualized as their $\log_{10}$-transformed values, where appropriate. Differences between categorical variables were assessed by Pearson’s chi-square test, except when any expected cell value was <5, where Fisher’s exact test was used instead. Relative risk was assessed by Cox regression.

Nonlinear modeling was used to assess the relationship between time from the last transurethral bladder tumor resection to the first intravesical nadofaragene firadenovec dose in study population patients and 3-mo post-treatment antibody titers. Two-sided $p \leq 0.050$ was considered statistically significant. Analyses were done using SPSS statistics version 24 (IBM Corp., Armonk, NY, USA). The nadofaragene firadenovec phase 3 trial is registered with ClinicalTrials.gov (identifier NCT02773849) [4].

Results

Antibody assay validation

Antibody detection and titration assay validation were conducted across 37 runs using negative and positive control samples before the assay was applied to unseen patient serum samples from the phase 3 study. This was done in accordance with practices commensurate with International Standards of Good Laboratory Practice and Good Clinical Practice guidelines (Covance Laboratories, Chantilly, VA, USA). Assay quality control and validation metrics are described below.

Intra/interassay screening precision was assessed using six sets of controls on each plate analyzed by at least two analysts over 2 d. Coefficient of variation calculated within each analytical run achieved target criteria at the CPC, and low and high positive controls confirming appropriate intra-assay precision. Coefficient of variation of the mean responses over six analytical runs was used to assess interassay precision. All performance control levels were acceptable and within the target criteria of ≤25%.

System suitability control ranges were assigned from all validation runs. Assay acceptance ranges provide a snapshot of the variability of assay control responses and were as follows:

1. Negative control response ≤0.32
2. CPC response of ≥0.25 and ≤ 0.77
3. Low positive control response of ≥0.85 and ≤2.25
4. High positive control response of ≥2.86 and ≤4.25

Intra/interassay titer precision was deemed acceptable when individual titer determinants were ±1 dilutions of the median target titer (calculated as 40 000), allowing a dilution range of

1:20 000 to 1:80 000. All titer determinants were within the range of the median target titer, which supported intra- and interassay precision criteria.

During sample analysis, false negative results are possible if antidrug antibodies are exposed and bound to any nadofaragene firadenovec present systemically. To investigate the concentration of vector that could interfere with the detection of anti–HAdV-5 antibodies, *vector tolerance* was assessed by spiking test clinical samples into matrix samples containing various concentrations of positive control antibodies. This demonstrated that the assay could detect a positive response at up to 1:5000 dilution factor of anti–HAdV-5 antibodies in the presence of 10 000 000 000 VLP/ml of vector, which was within the expectations for this type of assay.

Normal selectivity was tested using ten healthy control samples that were unspiked and spiked at high positive control level and serially diluted per protocol. No difference in titer result was noted between the spiked and unspiked samples due to the presence of pre-existing antibodies.

Diseased selectivity was assessed using ten anonymized patient samples that were serially diluted three times to show reproducibility of results. All cases generated the final result within one titer.

An ultrahigh positive control was prepared to assess any possible *hook effect* at concentrations above the high positive control. An increased response was observed in relation to the increase in concentration, demonstrating that saturation of any binding sites by the reference material did not affect the formation of complexes required to measure a qualitative result.

Stability assessments were considered valid when precision was acceptable at CPC, and low and high positive controls for all cycles, and responses were within system suitability ranges.

Freeze/thaw stability was assessed by subjecting control samples to four, six, and eight freeze/thaw cycles. Stability was retained for up to eight freeze/thaw cycles when stored at –80°C. *Benchtop stability* was assessed by subjecting control samples to 24 ± 4 h at room

temperature. Stability was retained for up to 24 h when stored on the bench at room temperature. *Refrigerator stability* was assessed by subjecting control samples to 74 ± 4 h at 2–8°C. Refrigerator stability was retained for up to 68 h and 24 min when stored at this temperature. *Long-term stability* was assessed at 40, 92, and 193 d. Data indicated that the assay was stable for up to 193 d when stored at –60 to –80 °C.

Marker associations and performance metrics in the subcohorts

Similar to observations in the overall study population, baseline antibody titer cutoff of 200 was unable to differentiate between responders and nonresponders in the CIS and high-grade Ta/T1 subcohorts (Supplementary Tables 2 and 3). Sensitivity of baseline antibody titer to predict durable response was 62%, 58%, and 65% in the overall study population, CIS subcohort, and high-grade Ta/T1 subcohort, respectively.

Sensitivity of 3-mo antibody titer for predicting durable response was 67% for the overall study population and both subcohorts (Fig. 1H–J). Three-month antibody fold change >8 from baseline was sensitive for predicting durable response in the overall study population (82%; Fig. 1H) and CIS subcohort (85%; Fig. 1I).

As a majority of responders and nonresponders achieved peak antibody titers within 6 mo following the first treatment dose (Fig. 1F and 1G), an exploratory analysis was performed to evaluate this assay's performance. For patients who achieved post-treatment antibody titer levels >800 at or before 6 mo following the first dose, the assay sensitivity for predicting durable response was 88% in the overall study population, and 95% and 81% in the CIS and high-grade Ta/T1 subcohorts, respectively (data not shown).

Twenty-two (92%) responders in the CIS subcohort had peak post-treatment antibody titers >800 compared with 20 (61%) nonresponders (Supplementary Table 2). Lower peak titers were associated with a higher relative risk of nondurable treatment response in the CIS and high-grade Ta/T1 subcohorts (Supplementary Tables 2 and 3). Sensitivity, negative predictive value, and accuracy of this assay were superior in the CIS (92%, 87%, and 61%, respectively) and high-grade Ta/T1 (87%, 63%, and 74%, respectively) subcohorts (Fig. 1I and 1J). Further, in the CIS subcohort, 15 (63%) responders had peak antibody fold change >8 from baseline compared with eight (24%) nonresponders (Supplementary Table 2). The relative risk of nondurable treatment response among CIS patients with peak fold change ≤8 was also higher (Supplementary Table 2).

Fifteen (63%) responders in the CIS subcohort had a favorable marker combination compared with six (18%) nonresponders (Supplementary Table 2). Among patients with CIS, unfavorable marker combination was associated with a nearly four-time increased risk of nondurable response (Supplementary Table 2). Specificity and positive predictive value of the marker combination were also superior in CIS (82% and 71%, respectively) and high-grade Ta/T1 (82% and 78%, respectively) subcohorts (Fig. 1I and 1J).

References

[1] Boorjian SA, Alemozaffar M, Konety BR, et al. Intravesical nadofaragene firadenovec gene therapy for BCG-unresponsive non-muscle-invasive bladder cancer: a single-arm, open-label, repeat-dose clinical trial. Lancet Oncol 2021;22:107–17.

[2] U.S. Food and Drug Administration. BCG-unresponsive nonmuscle invasive bladder cancer: developing drugs and biologics for treatment. Guidance for industry. Silver Spring, MD: U.S. Department of Health and Human Services; 2018.www.fda.gov/media/101468/download

[3] Westphal M, Ylä-Herttuala S, Martin J, et al. Adenovirus-mediated gene therapy with sitimagene ceradenovec followed by intravenous ganciclovir for patients with operable high-grade glioma (ASPECT): a randomised, open-label, phase 3 trial. Lancet Oncol 2013;14:823–33.

[4] ClinicalTrials.gov. Instiladrin in patients with bacillus Calmette-Guerin (BCG) unresponsive non-muscle invasive bladder cancer (NMIBC). Bethesda, MD: U.S. National Library of Medicine; 2016. clinicaltrials.gov/ct2/show/NCT02773849

Supplementary Table 1 – Distribution of antiadenoviral antibody titers and fold changes from baseline for all patients based on response to nadofaragene firadenovec treatment, as determined on a continuous scale

	Responders Median (IQR)	Nonresponders Median (IQR)	*p* value
Titer at baseline	400 (200–800)	300 (200–700)	0.48
Titer at 3 mo	1600 (800–3200)	800 (500–2800)	0.071
Fold change at 3 mo	4 (2–8)	3 (2–8)	0.47
Peak titer	6400 (1600–6400)	1600 (800–3200)	<0.001
Peak fold change	16 (4–32)	4 (2–16)	0.009

IQR = interquartile range.

p value was based on Mann-Whitney U test.

Supplementary Table 2 – Distribution of antiadenoviral antibody titers and fold changes from baseline, and relative risk of nondurable response to nadofaragene firadenovec treatment in the carcinoma in situ subcohort (*n* = 57)

	Distribution		Relative risk of nondurable response
	Responders *n* (%) [a]	Nonresponders *n* (%) [a]	Hazard ratio (95% CI)
Subcohort, *n* (row %)	24 (42)	33 (58)	
Titer at baseline			
>200	14 (58)	16 (48)	1.00 (Reference)
≤200	10 (42)	17 (52)	1.18 (0.60–2.34)
Titer at 3 mo			
>800	16 (67)	16 (48)	1.00 (Reference)
≤800	8 (33)	17 (52)	1.46 (0.74–2.90)
Fold change at 3 mo			
>8	7 (29)	5 (15)	1.00 (Reference)
≤8	17 (71)	28 (85)	1.75 (0.68–4.55)
Peak titer			
>800	22 (92)	20 (61)	1.00 (Reference)
≤800	2 (8)	13 (39)	2.80 (1.37–5.74)
Peak fold change			
>8	15 (63)	8 (24)	1.00 (Reference)
≤8	9 (37)	25 (76)	2.92 (1.31–6.50)
Marker combination [b]			
Favorable	15 (63)	6 (18)	1.00 (Reference)
Unfavorable	9 (37)	27 (82)	3.98 (1.63–9.72)

CI = confidence interval.

[a] Column % unless otherwise indicated.

[b] Favorable defined as a combination of peak antibody titer >800 and peak antibody fold change level >8. Patients not meeting both criteria were designated as unfavorable.

Supplementary Table 3 – Distribution of antiadenoviral antibody titers and fold changes from baseline, and relative risk of nondurable response to nadofaragene firadenovec treatment in the high-grade Ta/T1 subcohort (*n* = 34)

	Distribution		Relative risk of nondurable response
	Responders *n* (%) [a]	Nonresponders *n* (%) [a]	Hazard ratio (95% CI)
Subcohort, *n* (row %)	23 (68)	11 (32)	
Titer at baseline			
>200	15 (65)	6 (55)	1.00 (Reference)
≤200	8 (35)	5 (45)	1.34 (0.41–4.37)
Titer at 3 mo			
>800	14 (67)	3 (27)	1.00 (Reference)
≤800	7 (33)	8 (73)	3.39 (0.90–12.80)
Fold change at 3 mo			
>8	3 (14)	3 (27)	1.00 (Reference)
≤8	18 (86)	8 (73)	0.57 (0.15–2.13)
Peak titer			
>800	20 (87)	6 (55)	1.00 (Reference)
≤800	3 (13)	5 (45)	3.63 (1.10–12.01)
Peak fold change			
>8	9 (39)	4 (36)	1.00 (Reference)
≤8	14 (61)	7 (64)	1.18 (0.34–4.02)
Marker combination [b]			
Favorable	7 (30)	2 (18)	1.00 (Reference)
Unfavorable	16 (70)	9 (82)	2.00 (0.43–9.30)

CI = confidence interval.

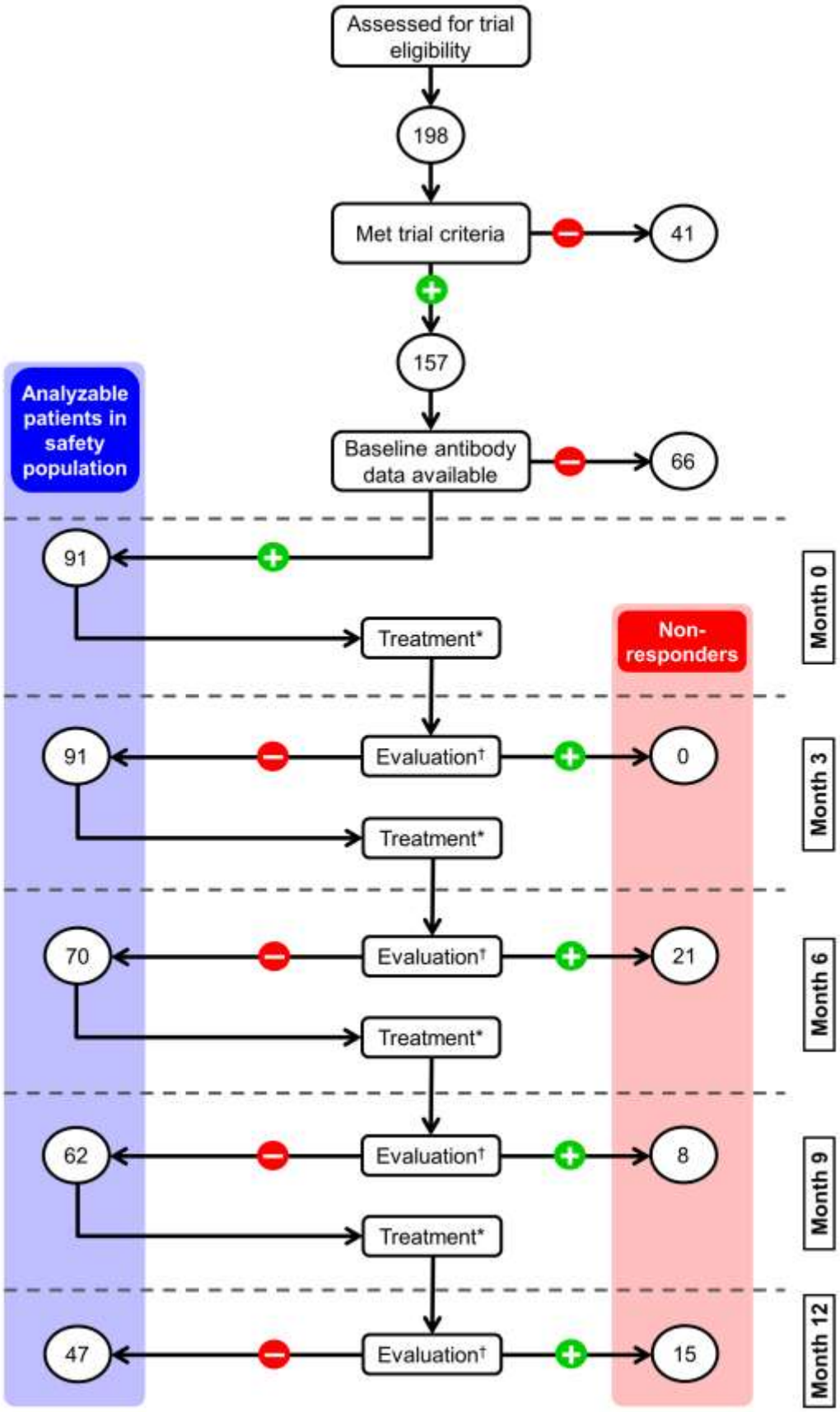
[a] Column % unless otherwise indicated.

[b] Favorable defined as a combination of peak antibody titer >800 and peak antibody fold change level >8. Patients not meeting both criteria were designated as unfavorable.

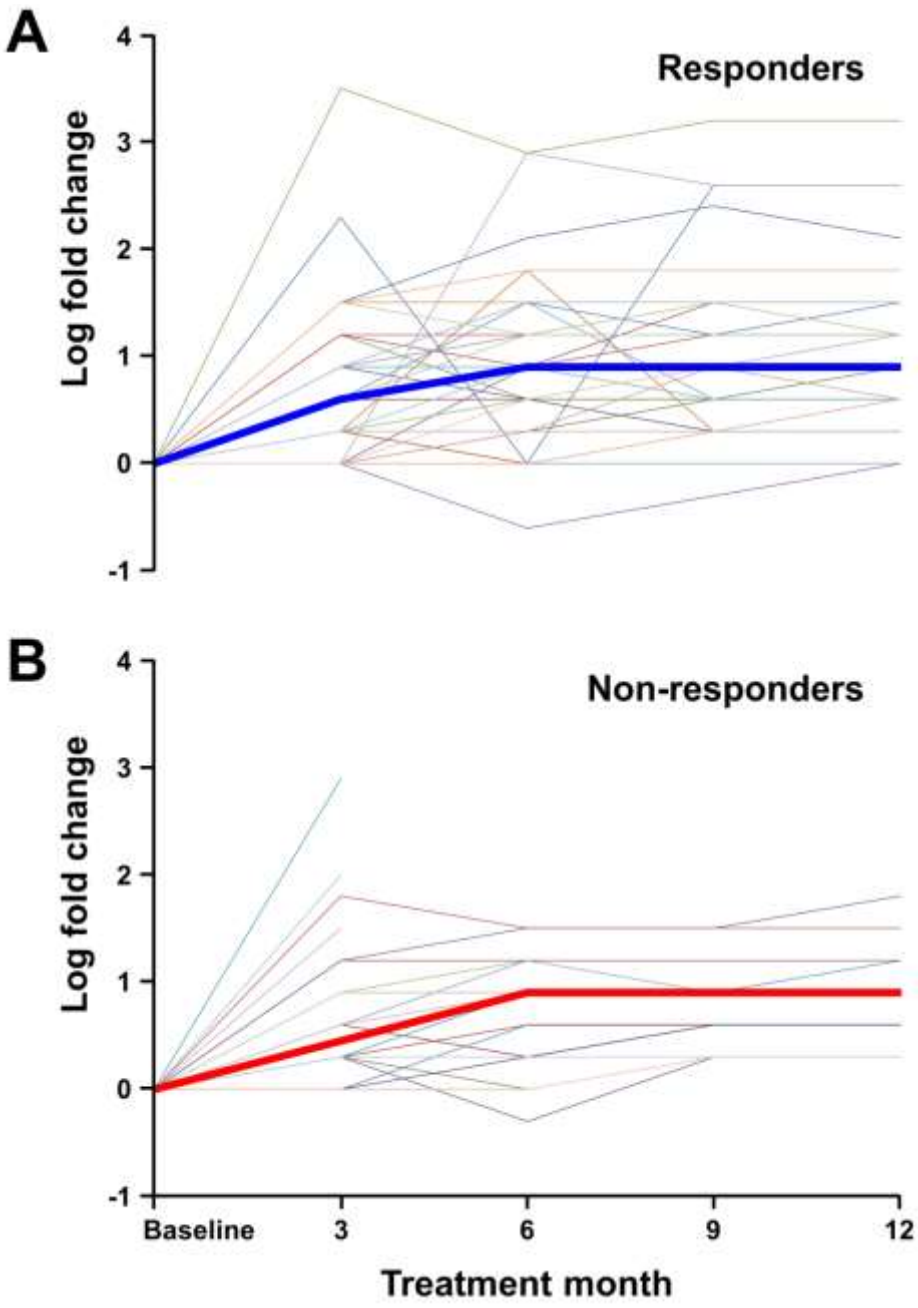
Supplementary Table 4 – Differences in baseline antiadenoviral antibody titers based on post-treatment titer groups at 3 mo and peak antibody response

	Baseline antibody level Median (IQR)		*p* value
	Post-treatment titer group >800	Post-treatment titer group ≤800	
At 3 mo	400 (200–1600)	200 (125–400)	<0.001
At peak	400 (200–800)	200 (1–400)	<0.001

IQR = interquartile range.
p value is based on Mann-Whitney U test.

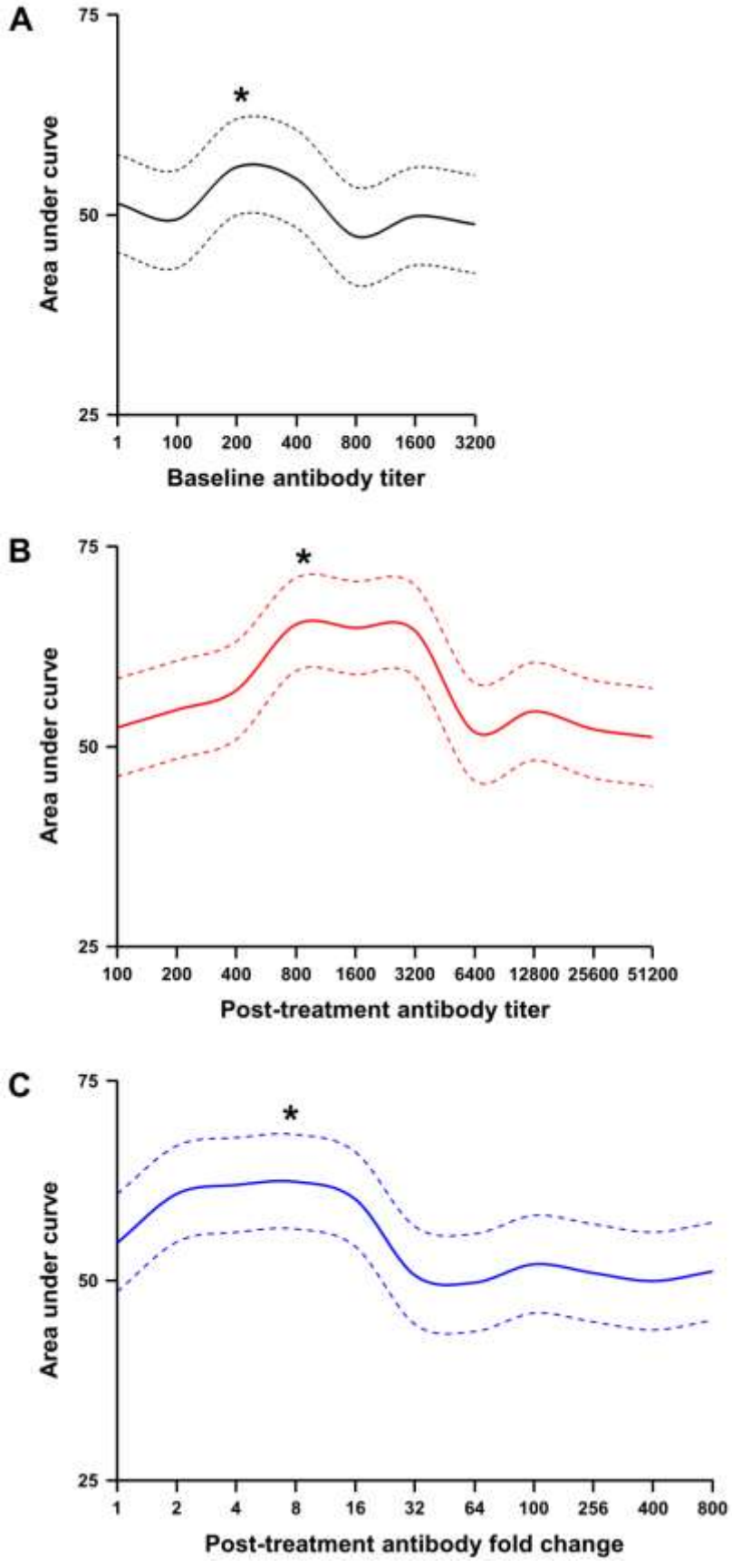

Assessed for trial eligibility
198
Met trial criteria
41
157
Baseline antibody data available
66
Analyzable patients in safety population
Non-responders
91
Treatment*
Month 0
91
Evaluation†
0
Treatment*
Month 3
70
Evaluation†
21
Treatment*
Month 6
62
Evaluation†
8
Treatment*
Month 9
47
Evaluation†
15
Month 12

Supplementary Fig. 1 – Summary of patients analyzed for the study. Of the 157 patients in the safety population who were enrolled in the phase 3 trial and received at least one dose of nadofaragene firadenovec, 91 underwent evaluation of serum anti–human adenovirus type-5 antibody titers. Titer measurements were performed prior to each treatment cycle. Measurements prior to first treatment (month 0) were considered baseline titers. Patients included in the blue area underwent pretreatment antibody testing, including those who were determined to be responders at month 12. * Treatment with nadofaragene firadenovec. † Clinical evaluation for recurrence of high-grade disease.

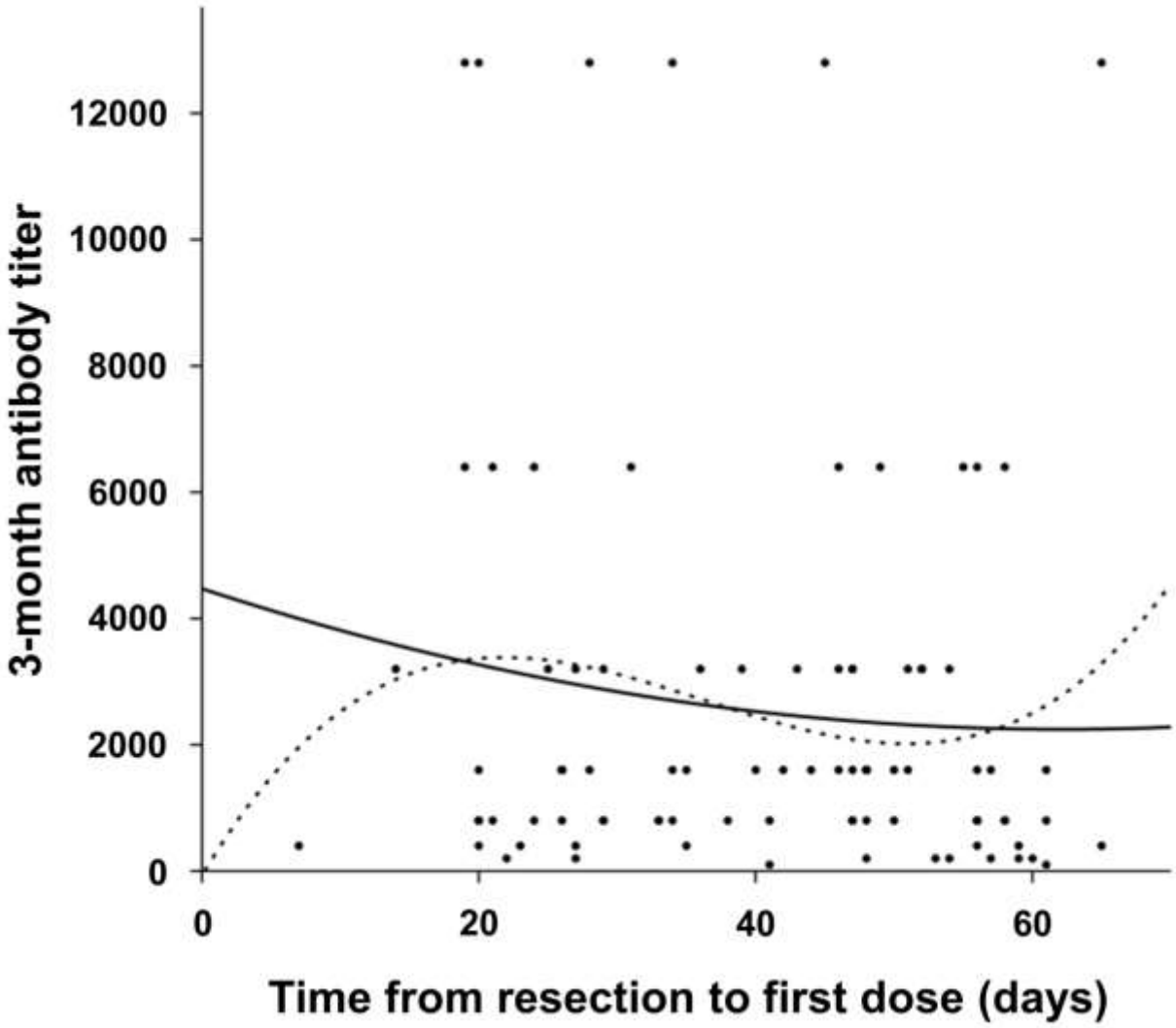


Supplementary Fig. 2 – Antibody fold change trends during treatment course. Trends in log_{10}-transformed antibody fold change levels for individual patients compared with baseline among (A) responders and (B) nonresponders during their treatment course. Thick blue and

red lines indicate median $\log_{10}$ fold change levels for responders and nonresponders, respectively.


A
75
50
25
Area under curve
*
1
100
200
400
800
1600
3200
Baseline antibody titer
B
75
50
25
Area under curve
*
100
200
400
800
1600
3200
6400
12800
25600
51200
Post-treatment antibody titer
C
75
50
25
Area under curve
*
1
2
4
8
16
32
64
100
256
400
800
Post-treatment antibody fold change

Supplementary Fig. 3 – Determining optimal antibody titer and fold change cutoffs. Incremental cutoff values of (A) baseline (black curve) and (B) post-treatment (red curve) anti–human adenovirus type-5 antibody titers, and (C) post-treatment antibody fold change from baseline were assessed for their ability to predict durable response to intravesical nadofaragene firadenovec therapy, as determined by the area under the receiver operating characteristic curve. Solid and dotted lines indicate area under the curve and corresponding standard error values, respectively. The symbol * indicates maximum area under the curve value.



Supplementary Fig. 4 – Association of the time from the last tumor resection with 3-mo antibody titers. Cubic (dotted curve) as well as quadratic (solid curve) modeling was used to assess the distribution of time from the last transurethral bladder tumor resection (in days) to the first intravesical nadofaragene firadenovec dose in study population patients with respective 3-mo post-treatment antibody titers. R^2 values for these models (0.026 and 0.014, respectively) indicated no significant association.