Dose painting of prostate cancer based on Gleason score correlations with apparent diffusion coefficients

Eric Grönlund, Silvia Johansson, Tufve Nyholm, Camilla Thellenberg & Anders Ahnesjö

To cite this article: Eric Grönlund, Silvia Johansson, Tufve Nyholm, Camilla Thellenberg & Anders Ahnesjö (2018) Dose painting of prostate cancer based on Gleason score correlations with apparent diffusion coefficients, Acta Oncologica, 57:5, 574-581, DOI: 10.1080/0284186X.2017.1415457

To link to this article: https://doi.org/10.1080/0284186X.2017.1415457

Published online: 20 Dec 2017.

Article views: 90
Dose painting of prostate cancer based on Gleason score correlations with apparent diffusion coefficients

Eric Grönlund, Silvia Johansson, Tufve Nyholm, Camilla Thellenberg and Anders Ahnesjö

Department of Immunology, Genetics and Pathology, Medical Radiation Sciences, Uppsala University, Uppsala, Sweden;
Department of Immunology, Genetics and Pathology, Experimental and Clinical Oncology, Uppsala University, Uppsala, Sweden;
Uppsala University Hospital, Uppsala, Sweden; Department of Radiation Sciences, Umeå University, Umeå, Sweden

ABSTRACT

Background: Gleason scores for prostate cancer correlates with an increased recurrence risk after radiotherapy (RT). Furthermore, higher Gleason scores correlates with decreasing apparent diffusion coefficient (ADC) data from diffusion weighted MRI (DWI-MRI). Based on these observations, we present a formalism for dose painting prescriptions of prostate volumes based on ADC images mapped to Gleason score driven dose-responses.

Methods: The Gleason score driven dose-responses were derived from a learning data set consisting of pre-RT biopsy data and post-RT outcomes for 122 patients treated with a homogeneous dose to the prostate. For a test data set of 18 prostate cancer patients with pre-RT ADC images, we mapped the ADC data to the Gleason driven dose-responses by using probability distributions constructed from published Gleason score correlations with ADC data. We used the Gleason driven dose-responses to optimize dose painting prescriptions that maximize the tumor control probability (TCP) with equal average dose as for the learning sets homogeneous treatment dose.

Results: The dose painting prescriptions increased the estimated TCP compared to the homogeneous dose by 0–51% for the learning set and by 4–30% for the test set. The potential for individual TCP gains with dose painting correlated with increasing Gleason score spread and larger prostate volumes. The TCP gains were also found to be larger for patients with a low expected TCP for the homogeneous dose prescription.

Conclusions: We have from retrospective treatment data demonstrated a formalism that yield ADC driven dose painting prescriptions for prostate volumes that potentially can yield significant TCP increases without increasing dose burdens as compared to a homogeneous treatment dose. This motivates further development of the approach to consider more accurate ADC to Gleason mappings, issues with delivery robustness of heterogeneous dose distributions, and patient selection criteria for design of clinical trials.

Introduction

The main objective in radiotherapy (RT) of cancer is to prescribe and deliver doses to the tumor target volume that yield a sufficient tradeoff between achieving a high tumor control probability (TCP) with an acceptable normal tissue complication probability (NTCP). Several studies demonstrate that dose escalation to the prostate volume decrease the post-RT risk for biochemical recurrences (BCR) [1–4]. However, homogeneous dose escalations come with the expense of also increasing the NTCP. To overcome such problems ‘dose painting’ has been proposed [5], a concept that explore the use of functional imaging for dose-response differentiation to yield individually optimized dose prescriptions that increase the TCP to NTCP ratio.

A spatially varying factor that yields important prognostic information for prostate cancer is the Gleason score. It is determined from biopsy samples of prostate cancer tissue, where increasing Gleason scores correlate with an increased BCR risk both after RT and prostatectomy [6–9]. Ghobadi et al. [10] explored the use of Gleason scores determined from prostatectomy specimens as a basis to differentiate dose-responses for individual patients. They found through implementing the observed Gleason scores in the linear-quadratic model, that dose differentiations up to 10 Gy could be achieved for individual prostate volumes without compromising the TCP compared to homogeneous dose treatments. Furthermore, it has been shown that increasing Gleason scores correlates with decreasing apparent diffusion coefficients (ADC) from diffusion weighted MRI (DWI-MRI) [11–17]. The study from Casares-Magaz et al. [18] used ADC images of prostate cancers to estimate individual TCP levels. They translated ADC data into varying cell densities which...
Dose painting by numbers [23]) were optimized by redis-

The FLAME trial [19] is an example that tests the use of image driven dose differentiation for prostate cancer. They used multiparametric MRI to determine primary tumor regions which for 35 fractions were treated with a boost dose of 95 Gy and with 77 Gy to the rest of the prostate volumes. The final results of the FLAME trial [19] are yet to be concluded, since long follow-up times are needed to acquire the treatments results of the included patients. In a review, Bentzen and Grégoire [20] stated that dose painting prescriptions ideally should be based on empirically observed dose-responses correlated to pretreatment image data. Such observation driven approaches have been presented by Vogelius et al. and Grönlund et al. [21,22]. In both these studies, pre-RT FDG-PET/CT image data were related to observed post-RT dose-responses, which yielded dose-response driven dose painting prescriptions that indicated significant TCP increases for patients with head and neck cancer as compared to homogeneous dose prescriptions.

In this study, we combine the ideas presented by Ghobadi et al. [10] and Casares-Magaz et al. [18] to design a formalism for Gleason score driven dose painting prescriptions that can be acquired from ADC images correlating with Gleason scores [11–17]. The aim is to present a formalism for dose painting prescriptions of prostate volumes by means of mapping ADC image data to Gleason driven dose-responses determined from retrospectively observed follow-up data.

Material and methods

In this study, dose painting prescriptions on a voxel level (i.e., dose painting by numbers [23]) were optimized by redistributing a conventional RT homogeneous dose so that individual TCP levels were maximized. The optimization process was driven by voxel specific dose-response functions derived from retrospectively observed pre-RT Gleason scores and post-RT outcomes from a learning data set of patients treated with a homogeneous dose to the prostate. To enable the ADC driven dose painting optimization, we translated ADC data to the Gleason driven dose-responses by means of published correlations between Gleason scores and ADC data. As end-point for TCP, we used freedom from BCR at 5 years post-RT.

We used two different sets of patient data, a learning set to derive Gleason score driven dose-response functions, and a test set with ADC images to test the formalism of acquiring image driven dose painting prescriptions. Both these sets of patient data are described below, followed by the derivation of Gleason driven dose response functions and the methodology for mapping ADC values into Gleason scores. Finally, we describe the optimization to achieve the dose painting prescriptions.

Learning set from retrospective treatment data

The learning data set consisted of pre-RT Gleason scores and post-RT outcomes from 122 high-risk patients (risk criteria according to NCCN guidelines [24]), treated between 2002 and 2015 at Uppsala University hospital (Uppsala board ethical approval reference number 2009/229). Patients with PSA >60 µg/l were not included to decrease the risk for pre-RT spread to confound the outcome. The patients were treated with dose-escalated RT as described by Johansson et al. [25], where all patients were treated with a proton boost 5 Gy × 4 fractions to the prostate volume (CTVT) only, followed by external X-ray RT 2 Gy × 25 fractions to both the CTVT and lymph node regions. With RBE = 1.1 for protons and α/β = 1.93 Gy [26] this yield the homogeneous CTVT treatment dose D h = 91.6 Gy EQD2. All patients underwent androgen deprivation therapy. The follow-up time ranged between 6 and 157 months where the time to a BCR varied between 2 and 114 months, and the time to a local recurrence 27–58 months. In total, 32 patients got a BCR and four a verified local recurrence. To determine the TCP of the learning set with the 5 year freedom from a BCR as end-point, we used the Kaplan–Meier product limit estimator and acquired TCP_{BCR} = 74% (see Figure 1). From these data, we parameterized the TCP of the learning set population (index ‘ls’) with the logistic function [27] as

\[
\text{TCP}_{ls}(D) = \frac{1}{1 + \left(\frac{D - D_{50,ls}}{\Delta D}\right)^{4\gamma_{50,ls}}},
\]

where γ_{50,ls} was set to 2.01 according to Dasu et al. [28], and D_{50,ls} set to 80.6 Gy, so that TCP_{ls} at the dose D = D_h equals the observed TCP_{BCR} = 74%.

Test set with ADC images of prostate cancer

To test ADC driven dose painting, we used image data from 18 high-risk prostate cancer patients (same risk criteria as for the learning set [24]) that had underwent diffusion weighted MR-imaging at Umeå University hospital. The used images were acquired both before RT and androgen deprivation therapy, to decrease the risk for hormonal effects on the prostate tissue and hence influence the ADC images. For some patients, the ADC images were acquired with a 1.5 Tesla MRI camera from Siemens, where the ADC image data were resolved with three values of the diffusion weighting parameter (‘b-value’), 0, 160 and 800 s/mm², and with voxel size of 1.63 × 1.63 × 3.60 mm³. For the remaining patients a 3.0 Tesla PET/MRI camera from GE Healthcare was used, with two b-values of 200 and 800 s/mm² and voxel size of 0.94 × 0.94 × 4.00 mm³. To have equal voxel sizes for all ADC images, we resampled them by trilinear interpolation into sizes of 3.00 × 3.00 × 3.00 mm³. These 18 patients constituted a subset of the patients included in the PARAPLY study (ClinicalTrials.gov Identifier NCT01962324) where patients with more than 15% risk of lymph node metastasis were treated to the prostate, seminal vesicles and pelvic lymph nodes (Umeå board ethical approval reference numbers 2013/154-31 and 2015/75-32).
Derivation of Gleason score driven dose response functions from the learning set

We used the pre-RT biopsy records of the learning set patients to construct virtual prostate volumes with Gleason scores voxelized into a grid of voxels sized \(3.00 \times 3.00 \times 3.00\, \text{mm}^3\). For this voxelization, we assumed that the taken biopsy samples were unbiased regarding the volumetric presence of Gleason scores within the prostate. Hence, we used the relative frequencies of Gleason scores for the dose-response derivation without using specified localizations of the Gleason scores. A detailed description is given in the Supplementary material, where we also describe a robustness test of the final dose painting prescriptions regarding possible biases in biopsy sample localization.

To derive dose-response functions driven by voxelized Gleason scores \(G\), we assumed for the homogeneous dose \(D_h\) a constant voxel specific TCP level \(D\) and \(G\), giving for \(b\) and \(m\) were determined from the voxelized Gleason scores of the learning data set, such that the tumor control of the learnings set TCP_{BCR} was reconstructed (see Supplementary material for details).

The assumed voxel specific TCP variation (Equation (2)) is only valid at the dose \(D_h\). To extend the dose-responses to any dose level (as in our earlier study [22]), we used the logistic sigmoidal function as

\[
\text{TCP}_{\text{vox}}(D, G) = \frac{1}{1 + \left(\frac{D_h(G)}{D}\right)^{\gamma_{50}}}.
\]

where \(D\) and \(G\) is voxel specific dose and Gleason score, respectively. Moreover, \(\gamma_{50}\) was set constant as suggested by Vogelius et al. [21] and in accordance with our previous work [22], and \(D_{50}(G)\) was set as a \(G\) dependent parameter. To determine \(\gamma_{50}\) and \(D_{50}(G)\) we assumed that tumor voxels comprising a patient’s prostate (CTVT) respond independently of each other, i.e., the TCP for a patient is given by

\[
\text{TCP}_{\text{pat}} = \prod_{\text{vox \in CTVT}} \text{TCP}_{\text{vox}}(D, G) \tag{4}
\]

and then required two conditions to be fulfilled for the homogeneous dose \(D_h = 91.6\, \text{Gy}\). First, that the patient specific TCP (calculated with Equation (4)) averaged over all learning set patients equals the observed TCP_{BCR} = 74\%. Second, that the tumor control gradient (i.e., the dose derivative of Equation (4)) averaged over all learning set patients equals the dose-response gradient of the whole learning set population (given by the dose derivative of Equation (1)). These two requirements yielded the determination of \(\gamma_{50}\) and \(D_{50}(G)\) as

\[
\gamma_{50} = \frac{N \cdot D_h/4}{\sum_{\text{pat}} \left[ \prod_{\text{vox \in CTVT}} \text{TCP}_{\text{vox}}(D_h, G) \cdot \left( \sum_{\text{vox \in CTVT}} (1 - \text{TCP}_{\text{vox}}(D_h, G)) \right) \right]}
\]

and

\[
D_{50}(G) = D_h \cdot \left( \frac{1}{\text{TCP}_{\text{vox}}(D_h, G)} - 1 \right)^{1/(\gamma_{50})} \tag{6}
\]

where \(N\) is the number of learning set patients, TCP_{vox}(D_h, G) the assumed TCP variation at the homogeneous dose \(D_h = 91.6\, \text{Gy}\) (Equation (2)), and \(\frac{d\text{TCP}_{\text{pat}}}{dD}_{D=D_h}\) the dose-response derivative for the learning set population at \(D_h = 91.6\, \text{Gy}\) (acquired by the dose derivative of Equation (1)). See the Supplementary material for further details.

Figure 1. Kaplan–Meier plot for the learning set, showing the local recurrence-free and biochemical recurrence-free ratios vs. time after RT.
Translation of ADC images to Gleason score driven dose-response functions

Decreasing ADC values are found to correlate with increasing Gleason scores $G$ [11–17]. However, while these observations correlate $G$ with ADC, we instead need the mapping of ADC to $G$. However, a direct one-to-one translation of ADC-to-$G$ would ignore the inherent uncertainties described for the correlations of $G$ with ADC. Hence, based on the uncertainty ranges given by Turkbey et al. [11], we constructed a probability density function (PDF) for each $G$ from 6 to 10 that described the related distribution of ADC. For this construction, we used lognormal distributions as these both restrict to positive ADC, and allow for variable skewness (see Figure 2 Left). For $G < 6$ explicit data was lacking and hence the distribution for $G = 6$ were duplicated. The required mapping of ADC to $G$ was achieved through renormalizing the constructed PDFs which yielded the conditional probability map $P(G|\text{ADC})$ that fulfills

$$
\int_{0}^{10} P(G|\text{ADC})\text{d}G = 1 \quad (7)
$$

for any ADC value. Hence, the thus renormalized map (see Figure 2 Right) gives the Gleason score probability distribution for any given ADC. For very low ADC values, a low-level background was added so that low ADC yielded the probability of 0.95 for $G = 10$ and 0.05 for $G \leq 6$. Small regions of high Gleason scores, resulting from ADC image noise, were removed through an image erosion operation with a spherical kernel of radius 0.45 cm. The kernel size was chosen so that the overall Gleason score distribution for the whole test set became similar to the observed Gleason score distribution of the learning set. Finally, the voxel specific ADC driven dose-response functions is expressed as the (geometric) average over the corresponding distribution of $G$ as

$$
\text{TCP}_{\text{vox}}(D, \text{ADC}) = \prod_{G=1}^{10} \text{TCP}_{\text{vox}}(D, G)^{P(G|\text{ADC})} \Delta G \quad (8)
$$

where we inherently assumed the TCP for voxels to be composed of independent sub compartments of different $G$.

Dose painting optimization

To acquire dose painting prescriptions on a voxel basis we maximized the TCP product for individual patients (with Equation (4)). However, to avoid excessively high doses, we constrained the average dose over each CTVT to equal the learning sets homogeneous treatment dose $D_h = 91.6$ Gy. For the learning set patients, we did not have ADC images and hence dose optimized with the derived Gleason driven dose-response functions (from Equation (3)). Moreover, for the test set patients we used the ADC driven dose response functions (acquired with Equation (8)) for dose optimization. The optimization was performed with a Lagrange optimization procedure as described by Ebert and Hoban [31], which we implemented into Mathematica® version 10.0 (Wolfram Research, Champaign, IL, USA, 2014).

Results

We derived the Gleason driven dose-response functions from the voxelized $(3 \times 3 \times 3\, \text{mm}^3)$ Gleason scores $G$ of the learning set and the assumed linear TCP variation for different $G$ at the treatment dose $D_h = 91.6$ Gy (given by Equation (2) and shown in Figure 3 Left). For this TCP variation, the value of $b$ for $G \leq 6$ and the linear gradient $m$ for $G > 6$ were determined so that the TCP averaged over all learning set patients equaled the observed TCP_{\text{BOR}} = 74\% yielding $b = 0.9999843$ and $m = -0.0004032$ (Figure 3 Left). Furthermore, the derived dose-response functions demonstrated an increasing radio-resistance with increasing $G$ (Figure 3 Middle), which was a result from the $D_{50}$ parameter that depended on varying $G$ (given by Equation (6) and shown in Figure 3 Right). The constant $G_{50}$ parameter for the dose-response functions (given by Equation (5) and shown in Figure 3 Right) equaled 2.289.

With the homogeneous dose prescription $D_h$, the calculated TCP per patient of the learning set ranged between 84% and 99% with the average 74%, and for the test set between 36% and 92% with the average 69%. Furthermore, by using the derived dose-response functions for dose optimization...
we acquired voxel specific dose levels that for both the patients of the learning and the test set increased with increasing values of $G$ (Figure 4).

The optimized dose painting prescriptions yielded TCP increases vs. the homogeneous dose prescription with $D_h$ that varied between 0 and 51% for the learning set patients and between 4 and 30% for the test set patients (Figure 5 Left). The average TCP increase for the learning set was 13% for the learning set, and 17% for the test set. It was also found that the potential for individual TCP increases
correlates well with the product of the CTVT volumes by the Gleason score spread (Figure 5 Left). The patients with the worst prognosis with low expected TCP for the homogeneous dose prescription had a higher potential for TCP increases with the dose painting prescriptions (Figure 5 Middle). Also, the product of the CTVT volumes by the Gleason score spread correlates with a decreasing TCP prediction for the homogeneous dose prescription with $D_h$ (Figure 5 Right).

For the dose painting prescriptions of the test set patients, it was found that regions with lower ADC values acquired higher doses and vice versa. This is demonstrated (in Figure 6) for centrally located prostate sections of the two tests set patients with the lowest and highest TCP increases, 4 and 30%.

**Discussion**

Our study is based on the assumption that Gleason scores can form a basis for dose painting prescriptions of prostate volumes, which is substantiated with observations that higher Gleason scores increases the post-RT risk for a BCR [6,9], and that higher doses decrease tumor cell survival. We used the 5 year freedom from BCR as tumor control endpoint; however, it is likely that a few patients of the learning set had some pre-RT micro spread that would result in a BCR.
independent of the dose given to the prostate. This is not taken into consideration in the dose-response derivation, which else would reduce the dose sensitivity with a less inclined gradient $m$ (see Equation (2) and Figure 3). Furthermore, (as shown in Figure 3) the TCP levels for voxelized Gleason scores at the dose $D_h$ are close to unity, which is a result of the small voxel volume. The TCP values for voxels comprising a target volume needs to be multiplied (as in Equation (4)), which hence as expected results in a lower TCP for a whole target volume. We chose to use a voxel size of $3 \times 3 \times 3 \, \text{mm}^3$, since this size often is used in dose grids for dose calculations in treatment planning systems.

Our dose-response derivation was based on the logistic dose-response function. This function is parameterized with the $\gamma_{50}$ and $D_{50}(G)$ parameters (determined by Equations (5–6)) that not bring mechanistic interpretations about the dose-response of cell survival. By instead implementing, e.g., a Poisson statistics based dose-response function as in Ghobadi et al. and Casares-Magaz et al. [10,18], mechanistic interpretations linking to the linear-quadratic model with respect to Gleason dependent dose-responses would be possible. However, we pragmatically chose the same non-mechanistic dose-response function as in our earlier study [22] and did not further explore any particular mechanisms for Gleason dependent dose-responses.

The dose-response functions were derived from $G$ values voxelized with the biopsy data of the learning set patients. Potential biases regarding the biopsy sample locations, such as avoidance of the central region to protect the less cancerous urethra, could for the prostate volumes result in a voxel distribution biased toward higher $G$ values. We tested this potential bias in voxelization of $G$ values (described in the Supplementary material) by replacing high $G$ values with zero samples. From this test we found that the remaining high $G$ values represented a worse voxel specific TCP prediction since the dose-response assumption (given by Equation (2)) still needed to reconstruct the observed TCP$_{SCR}$ of the learning set. However, for both routes to voxelize $G$ values we found the optimized dose painting prescriptions to be very similar due to the average dose constraint for dose optimization. Hence, dose-response uncertainties resulting from potential voxelization biases of $G$ values will have minimal impact on the resulting dose painting prescriptions, but rather translate into uncertainties regarding the prediction of absolute TCP values.

The ADC images of the test set were acquired with different MRI scanners as compared to the scanner used by Turkbey et al. [11], and may hence not be not fully comparable with the ADC data presented by Turkbey et al. [11]. This potential bias will lead to similar differences in TCP predictions as biases in sampling of $G$, but with small impact on the resulting dose painting prescriptions. This is consistent with our earlier study for head & neck cancers [22], were we found the resulting dose painting prescriptions to be robust by using absolute or relative FDG-PET standardized uptake values.

The ADC to $G$ translation is based on data from Turkbey et al. [11]. Their analysis was designed to investigate how specific $G$ correlate with ADC data, rather than how ADC data relate to different $G$. We renormalized their results to instead get the mapping of ADC values to probabilities for different $G$. Potential renormalization uncertainties could be avoided if studies were performed that directly scored ADC-to-Gleason distributions instead of Gleason-to-ADC distributions. Furthermore, we found that small noisy regions with low ADC resulted in the presence of small high $G$ regions. In the work by Ghobadi et al. [10], Gleason score regions smaller than 0.5 cm$^3$ were set to constitute nonmalignant regions, since such small regions likely are not detectable from MR-imaging. Hence, similar to Ghobadi et al. [10], we transformed small high Gleason score regions (resulting from noise of low ADC regions) into the surrounding region of lower Gleason scores by means of an erosion kernel. This erosion operation also resulted that a set of solitary voxels with arbitrary ADC data were optimized to the minimum dose prescription per patient (shown in Figure 6).

The potential TCP increases for the test set resulted in a slightly lower trend compared to the learning set (shown in Figure 5). This is explained by that the ADC driven dose-response functions represent geometric averages over probabilities of $G$ (Equation (8)). For voxels with low ADC values, low Gleason scores are still probable and vice versa for high ADC values. Hence, the TCP optimization for the test set gives a slight reduction in the dose-response differentiation compared to the learning set, which also yielded smoother dose variations (as shown in Figure 4).

Our results demonstrate the potential for significant TCP increases based on Gleason score driven dose painting prescriptions (Figure 5). The patients estimated with poor prognosis demonstrate a higher potential for TCP increases. Furthermore, the TCP increases correlates with increasing prostate volume and Gleason score heterogeneity, which can be used as a prognostic factor for selection of patients to be treated with dose painting. However, clinical trials are needed to confirm the feasibility of Gleason score driven painting approaches by means of ADC imaging. For such trials, consideration is needed both for which minimum and maximum dose levels that can be given safely. In the presented formalism, these levels varied between 63–90 and 100–137 Gy, respectively, for different patients (Figure 4). These dose levels are determined for an assumed ideal dose delivery with no delivery dependent dose correlations between neighboring voxels. The optimization of dose painting prescriptions does not include any lower or upper dose constraints, but such constraints could simply be included. However, our resulting dose levels seems to be in agreement with the study from Azzeroni et al. [32] that found through biological optimization based on Poisson dose-response modeling, dose levels up to 120 Gy for the most malignant subvolumes in prostates. Lastly, the range of heterogeneous dose levels that can be given is also limited by the physical realizability to deliver such dose levels, although we leave such issues to future studies.

Conclusions

Based on retrospective data of prostate cancer patients treated with RT, we have demonstrated a working formalism
for dose painting prescriptions of prostate volumes that potentially can yield significant TCP increases without extra dose burdens. The formalism is driven by Gleason scores that we probabilistically map from ADC images, which further argues for the inclusion of MRI within the preparatory workflow for RT. The potential for individual TCP increases correlated both to increasing prostate volumes and Gleason score heterogeneities, which may guide the patient selection for clinical trials of dose painting for prostate cancer.

Acknowledgments

We are grateful to Associate Professor Alexandru Dasu at the Skandion clinic in Uppsala, and Associate Professor Jan Weis at Uppsala University for helpful discussions.

Disclosure statement

We declare no conflict of interests.

Funding

This work was supported by the Swedish Cancer Society under grant number [130632].

References
