Molecular Marker Profiles Predict Locoregional Control of Head and Neck Squamous Cell Carcinoma in a Randomized Trial of Continuous Hyperfractionated Accelerated Radiotherapy

Francesca M. Buffa,[1] Søren M. Bentzen,[1] Frances M. Daley,[1] Stanley Dische,[1] Michele I. Saunders,[1,3] Paul I. Richman,[2] and George D. Wilson[4]

[1]Gray Cancer Institute and [2]Department of Pathology, Mount Vernon Hospital, London, United Kingdom; [3]Department of Oncology, University College London, London, United Kingdom; and [4]Barbara Ann Karmanos Cancer Institute, Detroit, Michigan

ABSTRACT

Purpose: **Identification of factors that assist prediction of tumor response to radiotherapy may aid in refining treatment strategies and improving outcome. Possible association of molecular marker expression profiles with locoregional control of head and neck squamous cell carcinoma was investigated in a randomized trial of conventional *versus* continuous hyperfractionated accelerated radiotherapy (CHART).**

Experimental Design: **Tumor material was obtained from 402 patients. Immunohistochemistry was used to assess Ki-67, CD31, p53, Bcl-2, and cyclin D1 expression. A hierarchical clustering algorithm with a Bayesian information criterion was used to group tumors with similar marker expression; resulting expression profiles were then compared in terms of their difference in outcome after CHART and conventionally fractionated radiotherapy.**

Results: **Molecular marker profile was an independent prognostic factor for locoregional control. This was confirmed in multivariate analysis, including clinical variables such as tumor and nodal status, primary site, histological grade, age, and gender ($P < 0.001$ and $P = 0.006$ for local and nodal relapse, respectively). In particular, Bcl-2-positive tumors responded significantly better than average in both arms of the trial. Tumors negative for p53- and Bcl-2, with high and randomly patterned Ki-67 expression, responded worse than average with no benefit from CHART. Tumors with similarly negative p53 and Bcl-2, but low Ki-67 staining, with an organized pattern, benefit significantly from CHART schedule.**

Conclusions: **This study demonstrates the potential of molecular profiles to predict radiotherapy response of head and neck squamous cell carcinoma and for treatment stratification. Distinct expression profiles correlate with three distinct clinical phenotypes, including good locoregional control, poor locoregional control, and an outcome strongly dependent upon fractionation schedule.**

INTRODUCTION

Radiotherapy, alone or combined with chemotherapy, plays a critical role in the management of head and neck squamous cell carcinoma (HNSCC; Ref. 1). Identification of factors that help to predict the response of HNSCC to radiotherapy and, in particular, assist identification of patients at risk of relapse, may be useful in refining treatment strategies and improving overall clinical outcome. The primary aim of this study was to determine a molecular marker "signature" capable of identifying a subgroup of patients who might benefit from being randomized to the strongly accelerated continuous hyperfractionated accelerated radiotherapy (CHART) treatment schedule rather than conventionally fractionated radiotherapy (CRT). Data were obtained from a group of 402 HNSCC patients participating in the Medical Research Council trial of CHART *versus* CRT (2). The primary clinical endpoint was locoregional control, with patient survival and occurrence of distant metastases as secondary end points. The panel of markers was chosen to represent biological factors known to affect tumor response to fractionated radiotherapy, such as vascular density and expression of proteins involved in the regulation of cell cycle progression, proliferation, or apoptosis. Immunohistochemistry was used to detect the following: (*a*) Ki-67, an antigen that is expressed by proliferating cells in all phases of the active cell cycle; (*b*) p53, a transcription factor that regulates cell cycle progression and apoptosis; (*c*) Bcl-2, an antiapoptotic molecule; (*d*) cyclin D1, which is required for cell cycle progression at the G_1-S transition; and (*e*) the cell adhesion molecule CD31, to assess vascular density.

The secondary aim of this study was to investigate the potential and flexibility of hierarchical clustering for clinical outcome analysis of data sets with a high number of factors, or covariates, measured for each case. Cluster analysis is an unsupervised classification method for data mining, directed toward identifying tumors with a similar molecular "fingerprint." Here, the clusters are built in the covariate space, and it is only after the clusters have been identified that they are compared in terms of outcome after a specific therapy, both in uni- and multivariate analysis. This strategy overcomes some of the limitations of survival analysis methods that select the covariates in the model based on their association with outcome in the same series, for

Received 9/17/03; revised 2/6/04; accepted 2/13/04.
Grant support: The Gray Cancer Institute Trust and the Translational Research Funding, Marie Curie Research Wing, Mount Vernon Hospital.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Requests for reprints: Francesca M. Buffa, Gray Cancer Institute, P. O. Box 100, Northwood, London, HA6 2JR, United Kingdom. E-mail: buffa@gci.ac.uk.

example Cox proportional hazards model. This model has been the mainstay of multivariate analysis of the influence of prognostic factors on end points requiring prolonged observation of the patient, such as survival or locoregional control, and it has been used already to analyze the impact of gene expression profiles on the outcome of breast carcinoma (3) and lung adenocarcinoma (4). However, *P*-values of the covariates included in the model are biased, and they appear to be more significant than they would in an independent confirmatory study. Also, there is a tendency to overfit the data, particularly when the number of covariates increases. As a result, the parameter estimates are often highly sensitive to the detailed composition of the patient population analyzed. Furthermore, this model is linear in the parameters, and it is not suited for the study of complex association patterns between many covariates and outcome. The use of clustering is based on the biological assumption that the overall biomolecular profile of a tumor will be indicative of its response to a specific treatment; thus, considering recurring biomolecular patterns, rather than the expression of single markers, may be a more powerful approach to identifying specific groups that benefit from a particular treatment.

MATERIALS AND METHODS

The Medical Research Council CHART Head and Neck Trial. Randomization to CHART *versus* CRT was performed with a 3:2 allocation ratio in favor of CHART. Treatment planning was identical for all of the patients and consisted of two irradiation phases. During the first phase, a large volume was irradiated, including primary tumor, clinically involved lymph nodes, and the relevant area of lymphatic drainage (large volume); and during the second phase, only the primary tumor and any involved lymph nodes, with a margin, were irradiated (small volume). The trial protocol consisted of 54 Gy in 36 fractions over 12 consecutive days or conventional therapy delivering 66 Gy in 33 fractions over 6.6 weeks. The large volume received only 37.5 Gy in the first 25 fractions of the trial protocol and 44 Gy in the first 22 fractions of the conventional treatment (2). Details on the tumor sites, tumor T and nodal N class, immunohistochemical data, patient population, and treatment in this study are given in Table 1.

Clinical Material and Immunohistochemical Staining. Of the 918 patients who were enrolled in the trial, it was possible to obtain adequate material for histological slides in 402 patients; the samples were taken before radiotherapy. The pathology department of each referring hospital was requested to either provide the original blocks or to cut up to 12 4-μm sections mounted onto poly-L-lysine coated slides. Each specimen was examined by a pathologist (P. I. R.) to confirm the presence of squamous cell carcinoma tumor and that the specimen was assessable. Specimens were immunostained to assess the expression of Ki-67, p53, CD31, Bcl-2, and cyclin D1. Sections were dried overnight at 37°C. Before antibody staining, the slides for each marker were blocked for endogenous peroxidase activity with a 3% solution of hydrogen peroxide in methanol for 30 min. Microwave irradiation was used to unmask the binding epitopes together with either 10 mM citric acid (pH 6.0) for Ki-67, p53, Bcl-2, and CD31 or 1 mM EDTA (pH 8.0) for cyclin D1. A cycle of three 5-min irradiations was used for all of the antibodies except Bcl-2, which required only two cycles. Slides were then left to stand for 10 min in buffer at room temperature before being washed thoroughly in tap water. After three washes in Tris-buffered saline (TBS), slides were incubated in the different antibodies in TBS containing 1 drop/ml of Dako Serum Free Protein Block (X0909; Dako Ltd., High Wycombe, United Kingdom) for 1 h at room temperature. Monoclonal antibodies were used at the following dilutions: p53 (clone DO-7; Dako Ltd.) at 1:75; CD31 (platelet/endothelial cell adhesion molecule-1; Dako Ltd.) at 1:30; Bcl-2 (clone 124; Dako Ltd.) at 1:40; and cyclin D1 (clone P2D11F11; Novacastra, Peterborough, United Kingdom) at 1:75. The antibody against Ki-67 was a rabbit polyclonal (clone MIB-1; Dako Ltd.) and used at 1:200 dilution. After three additional washes in TBS, biotinylated rabbit antimouse antibody (E0354; Dako Ltd.) or biotinylated swine antirabbit antibody (E0353; Dako Ltd.) diluted 1:400 in TBS was applied for 1 h at room temperature. After three additional washes, avidin-biotin complex (K0355; Dako Ltd.) was added for 1 h at room temperature. The staining was visualized by adding diaminobenzidine (DAB kit SK 4100; Vector Laboratories) for 5 min at room temperature. For cyclin D1, Dako Envision polymer kit (K4006; Dako Ltd.) was added for 30 min. After three additional washes in TBS, Envision DAB (Dako Ltd.) was added for 5 min. All of the slides were washed well in tap water and counterstained with Mayer's hematoxylin for 10 s to 1 min and then dehydrated, cleared, and counted in DPX.

Qualitative and Quantitative Assessment of Staining. All of the slides were qualitatively or semiquantitatively analyzed by a consultant pathologist, except in the case of cyclin D1. Ki-67 slides were visually scanned and ascribed to one of three scores, 1 = <20% positive cells, 2 = 20–40% positive, and 3 = >40% positivity. Ki-67 proliferative pattern was also assessed, as described previously (5), by scoring, 1 = marginal (mostly organized), 2 = intermediate (mainly organized), 3 = mixed (more than one pattern), and 4 = random (diffuse, disorganized staining). Bcl-2 was scored as negative if <5% of cells were stained and positive otherwise. The p53 staining was ascribed to one of three scores, 1 = negative (<5% positive cells), 2 = sporadic (5–75% cells positive), and 3 = all (>75% cells). The staining intensity also classified by scores, 1 = weak, 2 = intermediate, and 3 = strong. The microvessel count identified by CD31 staining was classified in 10 high-power fields, 1 = ≤35 microvessels, 2 = 35–55 microvessels, and 3 = >55 microvessels. Cyclin D1 was assessed by manual counting aided by an in-house image acquisition system incorporating a CCIR-format 3-CCD chip color camera coupled to a PC-based video-rate frame grabber. Software routines for image acquisition, normalization, and storage have been developed as have grids and manual counting data recording and export to spreadsheets. At least 10 high-power fields (×40 objective) were counted for each specimen.

Clinical End Points. All of the patients were observed weekly for a period of 6 weeks after the start of treatment. Additional follow-ups were at 8 weeks and 3 months after the first day of treatment, subsequently 3 monthly to 2 years, 6 monthly to 5 years, and annually thereafter. Local or nodal failures were recorded when clinically definitive tumor growth was detected at the primary site or in the nodes, respectively.

Table 1 Clinical and immunohistochemical data

Variable	Category	Number of patients	% of total patients
Age	Continuous	402	100
Gender	Male[a]	295	73.4
	Female	107	26.6
Treatment	CHART[b]	240	59.7
	Conventional[a]	162	40.3
T class	T1	12	3.0
	T2	185	46.0
	T3	131	32.6
	T4[a]	74	18.4
N class	N0[a]	255	63.4
	N+	147	36.6
Histological grade	Well differentiated[a]	85	21.1
	Moderately differentiated	183	45.5
	Poorly differentiated	79	19.7
	Squamous, not specified	55	13.7
Site	Oropharynx[a]	100	24.9
	Hypopharynx	41	10.2
	Larynx	190	47.3
	Oral cavity	56	13.9
	Nasal sinuses	2	0.5
	Nasopharynx	13	3.2
% of Ki-67 stained cells	Low (0–20%)	186	46.3
	Intermediate (20–40%)	129	32.1
	High (>40%)	87	21.6
Ki-67 staining pattern	Mostly organized	70	17.4
	Mainly organized	134	33.3
	Mixed	52	12.9
	Random[a]	146	36.3
% of p53 stained cells	Negative	208	51.7
	Sporadic (5–75% positive)	100	24.9
	All (>75% positive)	92	22.9
p53 staining intensity	Weak	222	55.2
	Intermediate	68	16.9
	Strong[a]	103	25.6
CD31 stained vessels	Low (up to 35 vessels)	166	41.3
	Intermediate (35–55 vessels)	151	37.6
	High (>55 vessels)	83	20.6
Bcl-2 expression	Negative (0–5% positive)[a]	349	86.8
	Positive (>5% positive)	51	12.7
Cyclin D1 staining	Continuous (% of staining)	322	80.1

[a] Marked categories are reference categories used in the outcome analyses.
[b] CHART, continuous hyperfractionated accelerated radiotherapy.

Distant failure was defined as the appearance of metastases outside of the irradiated volume. For each patient, the time to first failure and its location were recorded. Patients without clinical progression were censored at the time of the last follow-up. The local and nodal failures considered in this study were those within the small volume; that is the volume receiving the full dose and including the tumor and involved nodes with a margin. Locoregional control is defined as local control without any nodal recurrence.

Survival Analysis Methods. The statistical analyses were performed using the SPSS (SPSS Inc., Chicago, IL), S-PLUS 6.1 (Insightful, Basingstoke, United Kingdom), and R packages.[5] Backward stepwise elimination based on the likelihood ratio test was used to identify reduced Cox proportional hazards models including only covariates significant at the 0.05 level. For nominal variables, deviation contrasts were used (each category of the predictor variable except the reference category was compared with the overall effect), and for ordinal variables, indicator contrasts were used (each category was compared with the reference category). The significance of the interaction terms between covariates and nonlinear effects of scale covariates were assessed using the interaction and spline algorithms in the S-PLUS 6.1 function *coxphi*. The proportional hazards hypothesis was tested by estimating the Schoenfeld residuals correlation with the survival time (cox.zph; S-PLUS 6.1). To evaluate the predictive performance of different survival models, receiving operating characteristic curves and relative areas were calculated, using bootstrapping with replacement.

Hierarchical Clustering Methods. Cluster analysis was performed using SPSS implementation (TwoStep function) of the balanced iterative reducing and clustering using hierarchies method (6). This consists of two iterative steps. In the precluster step, the data are scanned individually, and a modified cluster

[5] Internet address: http://www.r-project.org.

features tree is constructed. Each new case, starting from the root node, descends along the tree based on its similarity to existing nodes, and it is either added to an existing leaf node or it can form a new leaf node. In the cluster step, the leaf nodes of the cluster feature tree are grouped into the desired number of clusters using an agglomerative clustering algorithm. This produces a range of solutions that are compared to estimate the optimal number of clusters. This method has two advantages compared with other K-mean or hierarchical clustering methods: it has the ability to cluster in a space formed by both scale and categorical variables, and it allows optimization of the number of clusters. It can also handle large data sets more efficiently than conventional hierarchical clustering algorithms. Here, the likelihood distance measure was used as the similarity criterion, because both scale and continuous variables were considered. In particular, the percentage of Ki-67/p53/cyclin D1-stained cells and the number of CD31-stained microvessels were entered as scale variables, because they are semiquantitative variables. An analysis where the percentage of Ki-67/p53-stained cells was entered as categorical variables were also performed, and the main results did not vary significantly from the results shown in this article (data not shown). Ki-67 staining pattern and p53 staining were entered as categorical variables, because they are qualitative parameters. Bcl-2 staining was entered as a binary variable. Schwarz's Bayesian criterion was used as the clustering optimization criterion. In a test on simulated data sets, the balanced iterative reducing and clustering using hierarchies algorithm was found superior to other K-mean and hierarchical clustering algorithms in identifying the simulated clusters, with a very low misclassification frequency (data not shown). In-house routines for bootstrapping with replacement were developed within the SPSS software to assess the robustness of cluster structure obtained with the balanced iterative reducing and clustering using hierarchies algorithm and the strength of the correlation of the resulting profiles with clinical end points in Cox regression analysis.

RESULTS

Molecular Marker Expression Profiling. Eight profiles were identified by cluster analysis (Fig. 1). CD31 and cyclin D1 values are shown for reference, because they were not statistically significant for formation of the cluster structure. Cluster dimensions varied between 24 and 68 patients (cluster 6 and cluster 1, respectively). Bcl-2-negative patients showed several patterns for the Ki-67 and p53 staining, while cluster 4 was fully characterized by Bcl-2-positive patients (Fig. 1). This result can be explained by considering that the number of Bcl-2-positive patients is very small ($n = 49$), and additional division of these patients into subclusters would worsen the Bayesian information criterion score of the cluster model. A bivariate correlation analysis of marker expression showed that the percentage and pattern of Ki-67 staining were positively correlated (Spearman two-tailed test, $P < 0.001$), the percentage and intensity of p53 staining were positively correlated ($P < 0.001$), Bcl-2 was positively correlated with the percentage ($P = 0.01$) and pattern ($P < 0.001$) of Ki-67 staining, and CD31 correlated with the percentage ($P = 0.005$) and intensity ($P < 0.001$) of p53 staining. Although the use of the likelihood distance measure assumes that the variables are independent, the balanced iterative reducing and clustering using hierarchies algorithm has been shown to be robust to violations of this assumption (6). However, because the percentage of Ki-67 and p53 staining showed a very high correlation to Ki-67 staining pattern and p53 staining intensity, respectively, and they were not scored in independent staining trials, a cluster analysis without including

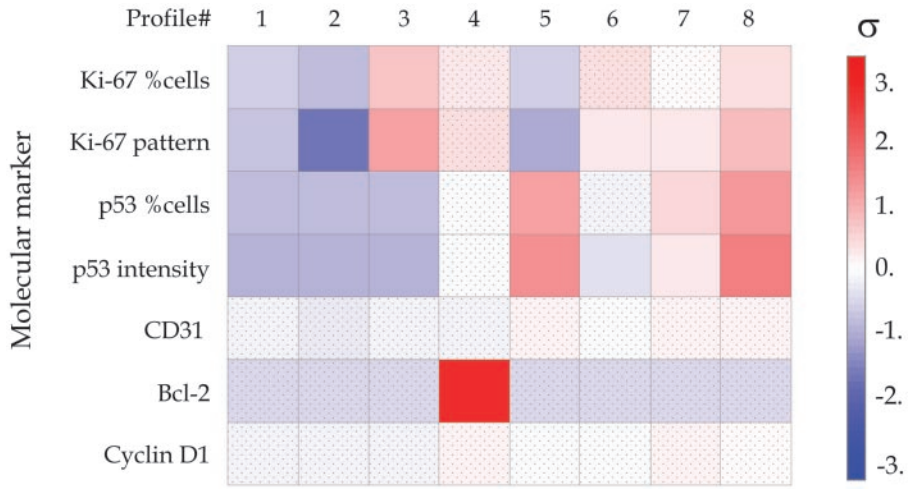

Fig. 1 Molecular marker expression profiles in head and neck squamous cell carcinoma of 402 patients. Results from the reduced balanced iterative reducing and clustering using hierarchies cluster model are shown; CD31 and cyclin D1 profiles are shown for reference only, because these two variables did not contribute significantly to the formation of any of the clusters. Expression values from the different markers were standardized and profiles normalized to the overall average over the clusters. The color coding is shown at *right*; *blue* corresponds to a value < overall average, *white* corresponds to average value, and *red* corresponds to a value > average. *Stippled squares* indicate that the variable was not significant ($P > 0.05$) for the formation of that particular cluster; *t* tests and χ^2 tests, with Bonferroni adjustment, were used to assess the significance for cluster formation of scale and categorical variables, respectively.

the percentage of stained cells was performed. Some changes in cluster membership were observed, but clusters 1–4 were unchanged (results not shown), and the main results of the treatment outcome analysis (see next section) also did not change. Bootstrapping on the cluster formation with 1000 samples showed that clusters 1–3 were very robust, with no misclassifications after bootstrapping; the average misclassification yield in cluster numbers 4–8 ranged between 21% and 29% of the cases, respectively, with no misclassifications at all in only 17–43% of the bootstrap runs, respectively.

Site and histological grade varied significantly among the clusters (Fig. 2), whereas T or N status, gender, and age did not.

Molecular Marker Profiles Correlate with Radiotherapy Locoregional Control and Survival. For all of the end points considered in this study, Kaplan-Meier plots showed that the hazard varied significantly among the molecular marker profiles (Fig. 3). A multivariate analysis incorporating the profiles into a Cox model together with the clinical variables of tumor T and N category, histological grade, tumor site, treatment, patient age, and gender, confirmed that profile membership was an independent prognostic factor for survival and locoregional control (Fig. 4) but not for distant metastases. Cluster 4, which is characterized by Bcl-2-positive patients (Fig. 1), showed a significantly better prognosis than any other cluster, with a higher survival and locoregional control (Figs. 3 and 4). Conversely, cluster 3 showed a significantly worse prognosis than any other cluster (Figs. 3 and 4); this cluster is characterized by a significantly higher than average percentage of Ki-67 stained cells, weak or absent p53 staining, and a random proliferative pattern (Fig. 1). After bootstrapping the reduced Cox model in Fig. 4 with 1000 samples these results were retained in >95% of the samples.

When the molecular markers were included also in the analysis, together with cluster membership and the other clinical variables, the clusters were still significant both for locoregional control and survival, although a high correlation was observed between the molecular markers and cluster membership, as expected. In this analysis (data not shown), cyclin D1 showed independent significance for survival and borderline significance for nodal relapse ($P = 0.010$ and $P = 0.066$, respectively), and p53 staining intensity was significant for survival ($P = 0.032$).

Cox Proportional Hazards Analysis of Molecular Markers and Clinical Variables. In a conventional Cox proportional hazards model, including molecular markers, tumor T and N class, histological grade, tumor site, treatment received, patient age, and gender, Bcl-2 status tended to dominate the results (data not shown). Bcl-2-positive patients showed lower hazard of local failure (0.416; $P = 0.008$) and nodal relapse (0.551; $P = 0.006$) and higher survival ($P < 0.001$). Cyclin D1 was positively correlated with nodal relapse hazard ($P = 0.027$) and negatively correlated with survival rate ($P = 0.012$). The p53 showed borderline correlation with local failure ($P = 0.085$) and significant correlation with survival ($P = 0.009$); tumors with low and high intensities had worse local control and survival than those with intermediate staining. Ki-67 staining pattern was positively correlated with the risk of distant failure ($P = 0.001$), with a 6.45 hazard risk of mostly organized *versus* random pattern. Interaction terms between the different markers

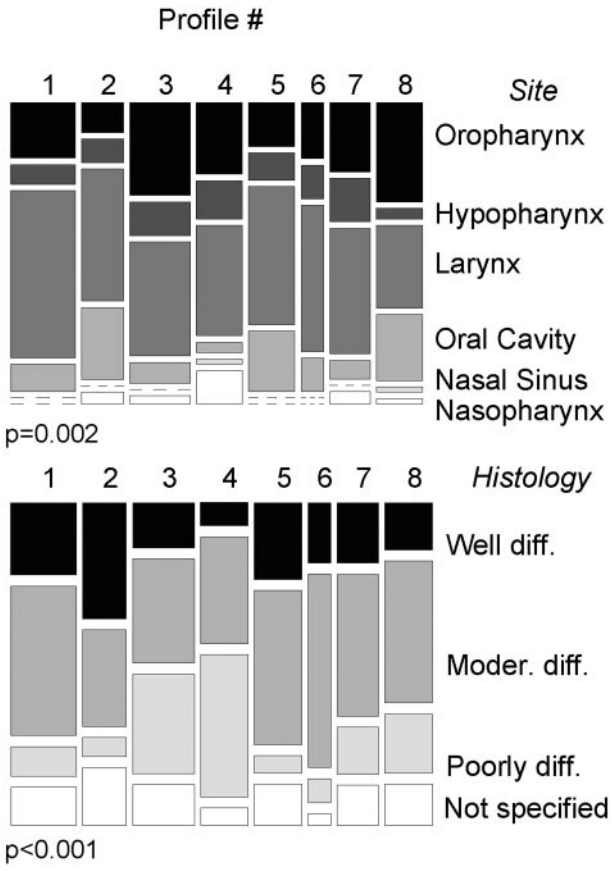

Fig. 2 Mosaic plots (the plotting surface is divided recursively according to the proportion of each of the two factors) are shown with the respective χ^2 test for independence of the cluster profiles (Fig. 1) as a function of histological grade (*bottom*) and site (*top*).

and quadratic terms did not show significance in these Cox models. The Cox proportional hazards hypothesis was statistically met in all of the cases, apart from the local control model ($P = 0.035$). However, when single covariates in this model were considered, T class only did not meet the hypothesis ($P = 0.003$). The area under the receiving operating characteristic curves was higher for Cox models including the clusters than for Cox models including separate biological or clinical characteristics only, but this was not significant.

Distinct Molecular Marker Profiles Are Associated with a Benefit from CHART. Treatment was found to be a significant prognostic factor for local control only (Fig. 4). When a Cox analysis for local control, including T and N class, treatment, histological grade, gender, and age, was stratified by cluster membership, treatment and T class were significant ($P = 0.045$ and 0.001, respectively). Log-rank test in each cluster (Fig. 5) showed that patients in clusters 2 and 7 had a significant benefit in local control from CHART ($P = 0.002$ and $P = 0.044$, respectively). After bootstrapping with 1000 samples, these results were retained in >95% of the samples. Similar results were obtained for nodal control and survival, although in this case, the significance was seen only for cluster 2 (data not shown). Clearly, comparing the effect of randomization to

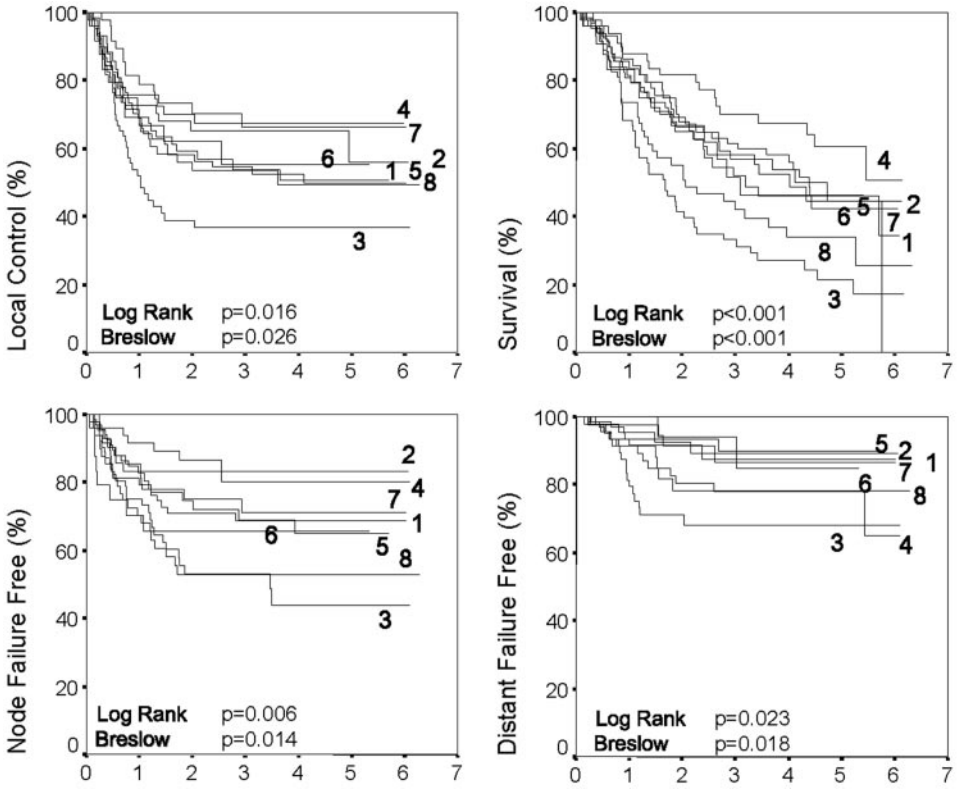

Fig. 3 Kaplan-Meier plots of survival, locoregional control, and local control factorized by molecular marker expression profile (see Fig. 1); *log rank* and *Breslow* statistics and their respective level of significance are reported in each graph.

CHART in each of the eight clusters involves a risk of false-positive findings because of multiple comparisons. However, after application of the Bonferroni correction for multiple tests, the difference in local control between patients in cluster 2 receiving CHART *versus* CRT was still significant (P = 0.0136).

DISCUSSION

With the help of new strategies for data analysis and new techniques that allow molecular and genetic characterization of tumors before treatment, individualization is becoming a realistic goal for cancer management. In particular, selection of patients who are likely to perform particularly badly and would benefit from a particular treatment schedule or modality could improve overall treatment outcome and efficacy.

Cluster analysis is an unsupervised classification method for data mining; it has been used recently in the analysis of gene expression data sets, showing the importance of gene expression profiles in relation to survival and treatment outcome (7, 8). However, there are fundamental biological differences in the information provided by molecular markers and gene expression microarrays, and histological data often include very different types of data, from quantitative data (such as the percentage of staining) to purely qualitative data (such as staining pattern); thus, algorithms allowing for different data types are needed. The balanced iterative reducing and clustering using hierarchies algorithm (6), with a Bayesian information criterion, was used in this study. This hierarchical clustering method allows for the presence of different types of data and can handle large data sets. Furthermore, the Bayesian information criterion selects the most informative number of clusters by finding an optimum balance between the two extreme scenarios, the few poorly defined clusters with many patients in each cluster and the many very well-defined clusters with few patients in each cluster. In particular, the latter case tends toward the extreme scenario of grouping patients on the basis of all of the possible marker expression combinations. This approach has been used in multiple marker studies, but it is mostly ineffective, because when the number of markers increases, the groups contain too few patients to allow a powerful comparison in terms of treatment outcome. Another important aspect of the strategy used in this study is that the profiles are clustered on their molecular similarity, and only after the formation process has ended are the clusters compared in terms of treatment outcome. This overcomes problems with biased significance of the parameters and overfitting, which are typical of survival analysis methods, where the covariates are selected based on their association with treatment outcome in the same series, such as Cox proportional hazards model.

This study focused on biological factors known to affect tumor response to radiotherapy and to fractionation schedule, such as proliferation (Ki-67), cell cycle deregulation (p53) and progression (cyclin D1), apoptosis (Bcl-2 and p53), and vascularity (CD31). Profile membership showed significant correlation with tumor site and histological grade (Fig. 2), and with the considered clinical end points (Figs. 3 and 4). In particular, patients in cluster 4 performed significantly better than patients in the other clusters. This cluster was fully characterized by

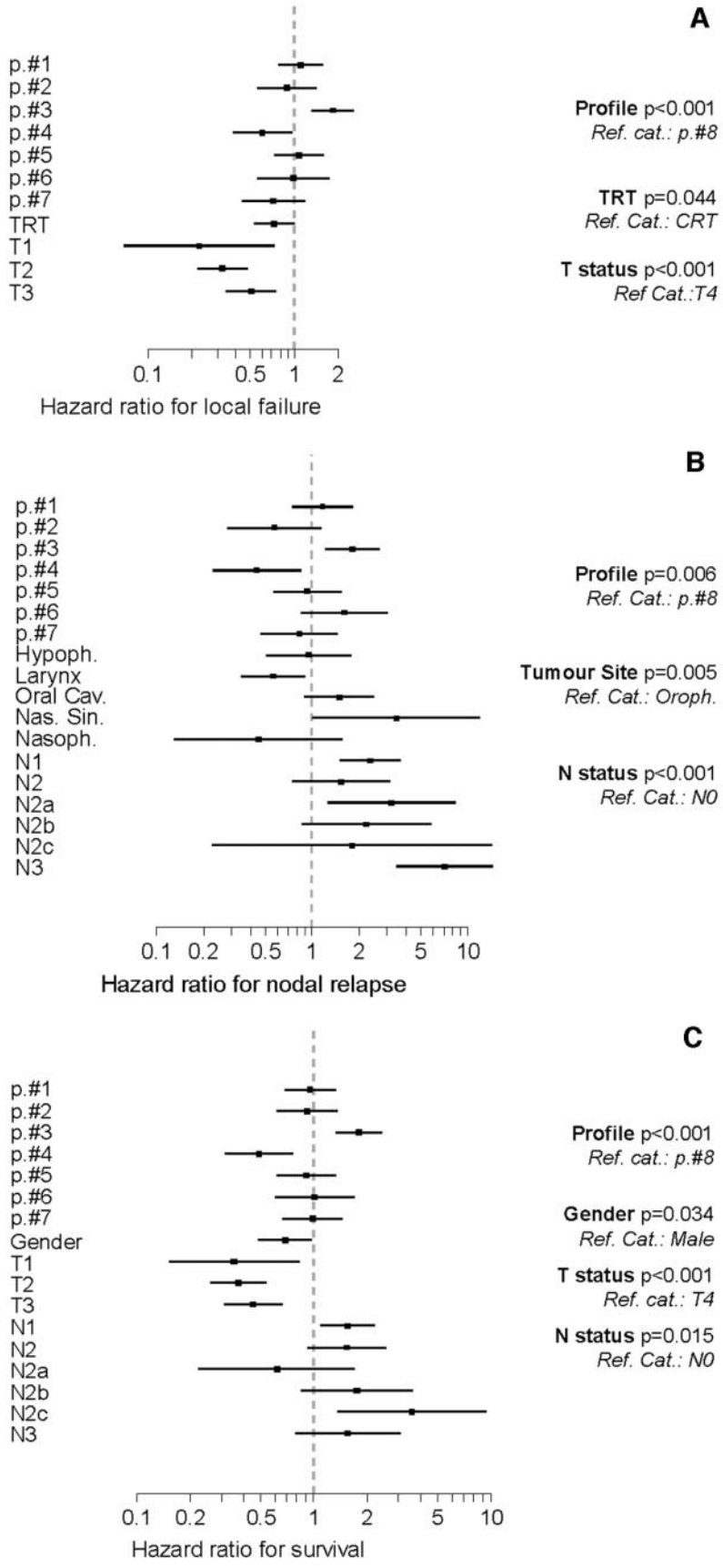

Fig. 4 Survival analysis including the molecular expression profiles, tumor T and N class, histological grade, tumor site, treatment received, patient age, and gender. The reduced Cox proportional hazards model is shown for survival (*A*), local (*B*), and nodal control (*C*). For each of the covariates, the hazards are shown with 95% confidence limits. The unity line is shown (*dotted gray line*). *P*s are indicated to the right of each variable. Deviation contrasts were used for site and cluster membership; indicator contrasts were used for the other categorical variables (see "Materials and Methods").

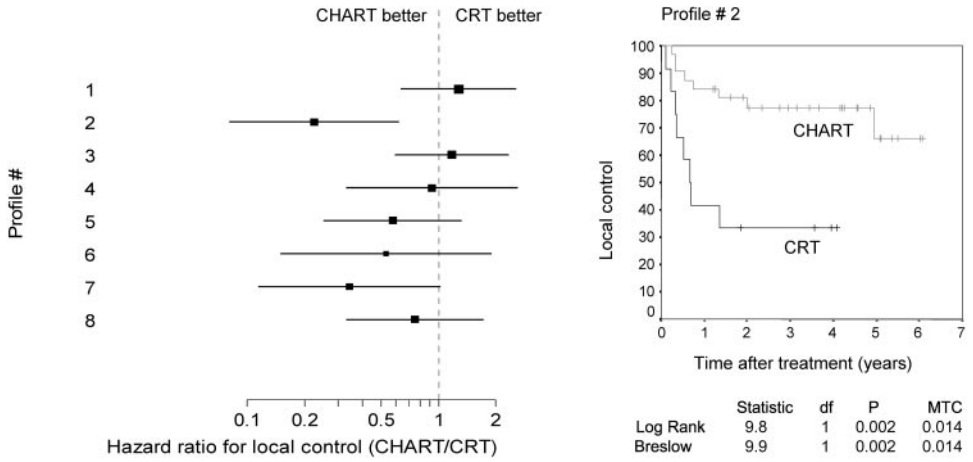

Fig. 5 The local control hazard ratios of continuous hyperfractionated accelerated radiotherapy (*CHART*) *versus* conventionally fractionated radiotherapy (*CRT*) and 95% confidence limits are shown for the molecular marker expression profiles (Fig. 1); the unity line is shown (*dotted gray line*). The *dots* are proportional to the profile size (*i.e.,* number of patients in the cluster). Kaplan-Meier local control curves are shown for the expression profile that showed the greatest local control hazard ratios of control hazard ratios of continuous hyperfractionated accelerated radiotherapy *versus* conventionally fractionated radiotherapy. The *log rank* and *Breslow* statistics are also shown, together with Bonferroni multiple test correction (*MTC*).

Bcl-2-positive tumors, and, accordingly, Bcl-2 expression has been shown previously to be associated with higher locoregional control in these patients (9). Although Bcl-2 appears to have a specific role in promoting the initiation and progression of numerous cancers, it has shown very controversial results in relation to radiotherapy outcome (10). In these patients, Bcl-2-negativity within tumors may simply reflect deregulation of this pathway (9). Furthermore, there is evidence that the expression ratio of Bcl-2 to Bax could be more informative than Bcl-2 absolute value (10). However, few patients (Table 1) had Bcl-2-positive tumors, and, thus, the Bcl-2:Bax ratio may be less informative here than in other tumor types.

Patients in cluster 3 showed a significantly worse response to treatment than patients in any other clusters (Figs. 3 and 4), with significantly lower local control ($P < 0.001$) and survival ($P = 0.001$). Tumors in this cluster were Bcl-2- and p53-negative and showed a high percentage of Ki-67 stained cells with a random proliferation pattern. Although Bcl-2 negativity is one of the possible causes for the poor outcome, it was not a unique feature of this cluster (Figs. 1 and 2). The high percentage of Ki-67-stained cells with a random pattern of staining implies a group of highly proliferative tumors, which might be more aggressive and, hence, have a worse overall treatment outcome. However, clusters 6–8 show a moderate to high percentage of Ki-67-stained tumor cells, with mixed to random patterns (Figs. 1 and 2), and they do not show poorer than average treatment outcome (Figs. 3 and 4). Like Bcl-2, Ki-67 is a marker that has generated studies also with conflicting results. For example, a study of larynx tumors from a large cohort of patients concluded that Ki-67 expression was not predictive of prognosis (11), whereas a recent study of oral cavity tumors has argued that a high percentage of Ki-67-stained cells may correlate with an improved radiotherapy response (12). However, in contrast to our study, these oral cavity tumors were mainly T_1 tumors. The p53 status of tumors in cluster 3 is difficult to establish, because the antibody used was not specific for wild-type or mutated p53. It is likely that high p53 staining intensity is attributable to the accumulation of mutated p53, whereas intermediate p53 staining intensity might be attributable to environmental stress-related induction of wild-type p53. Negative staining could indicate wild-type, deleted, or truncated p53 not recognized by the antibody. However, tumors with a negative and high p53 staining intensity experienced significantly poorer prognosis than tumors with intermediate intensity ($P = 0.039$, log-rank test), suggesting that tumors with negative or high level of intensity might have similar p53 status. This hypothesis could be tested by staining for target molecules of p53 such as p21, although this has the limitation that it will suggest only some of the possible mutations or, ideally, by looking for p53 mutations using DNA sequencing in a series where frozen material is available. Mutated p53 has been shown to correlate with increased resistance to radiotherapy in animal models (13, 14) and in humans (12) and could contribute to the poor prognosis in this cluster. However, low p53 intensity is not a unique feature of this cluster. All of the above considerations indicate that the particular combination of all of these factors, rather than one specific factor alone, determines the poor prognosis of patients in cluster 3.

Although cyclin D1 had no major role in the formation of the clusters, it was still a significant variable for outcome in a Cox model, independently from the observed profiles, and correlated negatively with nodal failure and survival (see "Results"). This is in agreement with the previous finding that cyclin D1 overexpression is an indicator of poor prognosis in HNSCC (15, 16).

Randomization to CHART was a significant factor for only local control (Fig. 4), with a 0.7 hazard ratio of CHART *versus* CRT. Kaplan-Meier analyses in the various clusters showed that treatment was a significant variable for local control in cluster 2 and cluster 7 (Fig. 5). Multiple paired comparisons of CHART

versus CRT in each of the eight clusters may lead to spuriously significant difference. However, the difference in local control in cluster 2 between patients receiving CHART *versus* CRT was still significant after a Bonferroni correction for multiple testing ($P = 0.0136$). Bcl-2 and p53 negativity and low Ki-67 expression, with mostly organized staining pattern, characterized these tumors (Fig. 1). Cluster 2 shows also a higher percentage of well-differentiated tumors with respect to the other clusters (Fig. 2); this agrees with a previous study suggesting that well-differentiated tumors have a more pronounced treatment-time dependency (17). Profile 7 showed benefit also from CHART, but this result was not significant after a multiple test correction. This cluster is characterized by intermediate values of p53, Bcl-2 negativity, and low organization of the Ki-67 staining, suggesting that highly and randomly proliferative tumors might benefit also from CHART in the case of intermediate p53 levels.

Although these results are interesting, additional investigation is needed, because identification of tumor characteristics that are predictive for a benefit from acceleration would allow both selection of patients for a specific fractionation schedule and development of radiation modifiers targeted at specific pathways or molecules. When comparing CHART with CRT, the benefit from CHART will be a trade-off between a decreased likelihood of locoregional control because of a lower dose and an increased likelihood of locoregional control because of a shorter treatment time. However, the CHART trial has found no overall benefit from accelerated schedule with respect to CRT (2); this is thought to reflect either a lower-than-expected overall accelerated repopulation response in these tumors, or, perhaps, that the advantage from a short schedule has been counteracted by a shorter time (12 days) to reoxygenate the most hypoxic of these tumors (18). Thus far, attempts to relate cellular proliferation directly to the benefit from accelerated fractionation have failed, and it is likely that the putative link involves multiple biological factors (19). Here, we have shown that tumors with negative p53 and Bcl-2, and with a mostly organized Ki-67 pattern, did benefit from CHART. These tumors could have a greater tendency to respond to radiation damage with accelerated repopulation, as seen in normal squamous epithelium (20, 21), and this type of response could occur less frequently or be less pronounced in tumors showing a proliferation pattern with lower organization, or tumors with a different p53 or Bcl-2 status. This would make these tumors most sensitive to changes in fractionation schedule. This hypothesis is under study, both using biomodeling simulations and by assessing the expression of molecules believed to be involved or associated with the occurrence of accelerated repopulation after irradiation, such as epithelial growth factor receptor.

An alternative hypothesis is that these tumors have a more pronounced or faster reoxygenation, which would overcome the disadvantage of the shortened time for reoxygenation of very hypoxic tumors undergoing CHART. The expression of CD31 before therapy (and thus microvessel density) in these tumors is not different from other clusters (Fig. 1), suggesting that the initial hypoxic status is not different. However, a Kaplan-Meier analysis of local control, comparing patients according to treatment and stratifying them by CD31 expression (dichotomized to gain some statistical power), showed a significant benefit from CHART both for the low and the high CD31 tumors ($P = 0.01$), with a very similar hazard for CHART *versus* CRT between the low and high CD31 expression groups (0.28 and 0.26, respectively). This supports the hypothesis that the hypoxic tumors in this cluster are able to reoxygenate during treatment. In contrast, when the same analysis was performed in cluster 7 (the other cluster showing a benefit from CHART), significance was seen only for the high CD31 group (CHART/CRT hazard = 0.14; $P = 0.004$); no benefit was seen for the low CD31 group. This suggests that hypoxic tumors limit the benefit from CHART for patients in cluster 7, or, in other words, that for these tumors, 12 days are not sufficient for reoxygenation. Clinical support for the hypothesis that impeded reoxygenation may play a role in the CHART regimen comes from a randomized controlled trial showing a benefit when the bioreductive drug mitomycin C was added to CHART radiotherapy (22). The validation of both the repopulation and reoxygenation hypotheses would benefit from an independent prospective series of patients, where proliferation and oxygenation could be measured before, during, and after radiotherapy, using either histological molecular markers or molecular imaging techniques such as positron emission tomography, where new biomarkers are being found and tested for targeting hypoxia and proliferation (23).

In conclusion, this study identifies the potential of molecular marker expression profiles to predict radiotherapy response of HNSSC and treatment stratification. In particular, three distinct expression profiles were identified that correlate with three distinct clinical phenotypes; one group of patients responded particularly well to radiotherapy, a second group showed poor radiotherapy response, and a third group showed a significant benefit from a strongly accelerated radiotherapy. Additional staining of these tumors, to study additional markers, may help to identify patients who are at risk of locoregional HNSCC failure and particularly sensitive to changes in fractionation schedules. Prospective studies, where immunohistochemical staining, molecular imaging, and gene expression data can be collected at different time points, may help to correlate distinct radiation responses with specific aberrant proliferation, apoptotic and/or hypoxic pathways, and to refine radiotherapy fractionation and chemoradiation strategies in HNSCC.

ACKNOWLEDGMENTS

We acknowledge the vital work of the CHART Steering Committee including A. Barret (Chairman), D. Coyle, B. Cottier, A. Crellin, P. Dawes, S. Dische, M. Drummond, C. Gaffney, D. Gibson, A. Harvey, J. Henk, T. Herrmann, B. Littbrand, J. Littler, F. Macbeth, D. Morgan, H. Newman, M. Parmar, A. Robertson, M. Robinson, R. Rothwell, M. Saunders, R. Symonds, J. Tobias, M. Whipp, and H. Yosef, and D. Gascoyne for manuscript comments.

REFERENCES

1. Bernier J, Bentzen SM. Altered fractionation and combined radiochemotherapy approaches: pioneering new opportunities in head and neck oncology. Eur J Cancer 2003;39:560–71.

2. Dische S, Saunders M, Barret A, et al. A randomised multicentre trail of CHART versus conventional radiotherapy in head and neck cancer. Radiother Oncol 1997;44:123–36.

3. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.

4. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002; 8:816–24.

5. Bennett MH, Wilson GD, Dische S, et al. Tumour proliferation assessed by combined histological and flow cytometric analysis: implications for therapy in squamous cell carcinoma in the head and neck. Br J Cancer 1992;65:870–8.

6. Chiu T, Fang D, Chen J, Wang Y, Jeris C. A robust and scalable clustering algorithm for mixed type attributes in large database environment. 7th ACM SIGKDD Conference Proceedings 2001;263.

7. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 2003;100:8418–23.

8. Weiss MM, Kuipers EJ, Postma C, et al. Genomic profiling of gastric cancer predicts lymph node status and survival. Oncogene 2003;22: 1872–9.

9. Wilson GD, Saunders MI, Dische S, et al. Bcl-2 expression in head and neck cancer: an enigmatic prognostic marker. Int J Radiat Oncol Biol Phys 2001;49:435–41.

10. Haffty BG, Glazer PM. Molecular markers in clinical radiation oncology. Oncogene 2003;22:5915–25.

11. Hirvikoski P, Kumpulainen E, Virtaniemi J, et al. Enhanced apoptosis correlates with poor survival in patients with laryngeal cancer but not with cell proliferation, Bcl-2 or p53 expression. Eur J Cancer 1999;35:231–7.

12. Couture C, Raybaud-Diogene H, Tetu B, et al. p53 and Ki-67 as markers of radioresistance in head and neck carcinoma. Cancer (Phila) 2002;94:713–22.

13. Lowe SW, Bodis S, McClatchey A, et al. p53 status and the efficacy of cancer therapy in vivo. Science (Wash DC) 1994;4:807–10.

14. Pirollo KF, Hao Z, Rait A, et al. p53 mediated sensitization of squamous cell carcinoma of the head and neck to radiotherapy. Oncogene 1997;14:1735–46.

15. Michalides RJ, van Veelen NM, Kristel PM, et al. Overexpression of cyclin D1 indicates a poor prognosis in squamous cell carcinoma of the head and neck. Arch Otolaryngol Head Neck Surg 1997;123:497–502.

16. Michalides R, van Veelen N, Hart A, et al. Overexpression of cyclin D1 correlates with recurrence in a group of forty-seven operable squamous cell carcinomas of the head and neck. Cancer Res 1995;55:975–8.

17. Hansen O, Overgaard J, Hansen HS, et al. Importance of overall treatment time for the outcome of radiotherapy of advanced head and neck carcinoma: dependency on tumour differentiation. Radiother Oncol 1997;43:47–51.

18. Kaanders JH, Bussink J, van der Kogel AJ. ARCON: a novel biology-based approach in radiotherapy. Lancet Oncol 2002;3:728–37.

19. Bentzen SM. Repopulation in radiation oncology: perspectives of clinical research. Int J Radiat Biol 2003;79:581–5.

20. Dorr W. Modulation of repopulation processes in oral mucosa: experimental results. Int J Radiat Biol 2003;79:531–7.

21. Trott KR. Perspectives of experimental research on repopulation during radiotherapy. Int J Radiat Biol 2003;79:577–80.

22. Dobrowsky W, Naude J. Continuous hyperfractionated accelerated radiotherapy with/without mitomycin C in head and neck cancers. Radiother Oncol 2000;57:119–24.

23. Van de Wiele C, Lahorte C, Oyen W, et al. Nuclear medicine imaging to predict response to radiotherapy: a review. Int J Radiat Oncol Biol Phys 2003;55:5–15.