

Impact of supervised gene signatures of early hypoxia on patient survival.

Citation for published version (APA):

Seigneuric, R., Starmans, M. H. W., Fung, G., Krishnapuram, B., Nuyten, D. S. A., van Erk, A. A., Magagnin, M. G., Rouschop, K. M. A., Krishnan, S., Rao, R. B., Evelo, C. T. A., Beqq, A. C., Wouters, B. G., & Lambin, P. (2007). Impact of supervised gene signatures of early hypoxia on patient survival. *Radiotherapy and Oncology*, 83(3), 374-382. <https://doi.org/10.1016/j.radonc.2007.05.002>

Document status and date:

Published: 01/01/2007

DOI:

[10.1016/j.radonc.2007.05.002](https://doi.org/10.1016/j.radonc.2007.05.002)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Clinical radiobiology

Impact of supervised gene signatures of early hypoxia on patient survival

Renaud Seigneuric^{a,*}, Maud H.W. Starmans^a, Glenn Fung^b, Balaji Krishnapuram^b,
Dimitry S.A. Nuyten^c, Arie van Erk^d, Michael G. Magagnin^a, Kasper M. Rouschop^a,
Sriram Krishnan^b, R. Bharat Rao^b, Chris T.A. Evelo^d, Adrian C. Begg^e,
Bradly G. Wouters^a, Philippe Lambin^a

^aMaastricht Radiation Oncology (Maastricht), GROW Research Institute, Maastricht University, The Netherlands, ^bSiemens Medical Solutions, CAD & Knowledge Solutions, Malvern, PA, USA, ^cDepartment of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands, ^dBiGCaT Bioinformatics, University of Maastricht, The Netherlands, ^eDivision of Experimental Therapy, The Netherlands Cancer Institute, Amsterdam, The Netherlands

Abstract

Background and purpose: Hypoxia is a common feature of solid tumors associated with therapy resistance, increased malignancy and poor prognosis. Several approaches have been developed with the hope of identifying patients harboring hypoxic tumors including the use of microarray based gene signatures. However, studies to date have largely ignored the strong time dependency of hypoxia-regulated gene expression. We hypothesized that use of time-dependent patterns of gene expression during hypoxia would enable development of superior prognostic expression signatures.

Materials and methods: Using published data from the microarray study of Chi et al., we extracted gene signatures correlating with induction during either early or late hypoxic exposure. Gene signatures were derived from *in vitro* exposed human mammary epithelial cell line (HMEC) under 0% or 2% oxygen. Gene signatures correlating with early and late up-regulation were tested by means of Kaplan–Meier survival, univariate, and multivariate analysis on a patient data set with primary breast cancer treated conventionally (surgery plus on indication radiotherapy and systemic therapy).

Results: We found that the two early hypoxia gene signatures extracted from 0% and 2% hypoxia showed significant prognostic power (log-rank test: $p = 0.004$ at 0%, $p = 0.034$ at 2%) in contrast to the late hypoxia signatures. Both early gene signatures were linked to the insulin pathway. From the multivariate Cox-regression analysis, the early hypoxia signature ($p = 0.254$) was found to be the 4th best prognostic factor after lymph node status ($p = 0.002$), tumor size ($p = 0.016$) and Elston grade ($p = 0.111$). On this data set it indeed provided more information than ER status or p53 status.

Conclusions: The hypoxic stress elicits a wide panel of temporal responses corresponding to different biological pathways. Early hypoxia signatures were shown to have a significant prognostic power. These data suggest that gene signatures identified from *in vitro* experiments could contribute to individualized medicine.

© 2007 Elsevier Ireland Ltd. All rights reserved. Radiotherapy and Oncology 83 (2007) 374–382.

Keywords: Gene Expression Signature; Molecular medicine; Time series; Hypoxia; Insulin

Maintenance of physiological oxygen concentrations is a highly dynamic process vital to many aspects of cell biology. Lack of oxygen (hypoxia) is a frequently encountered situation in solid tumors and is associated with increased radiotherapy and chemotherapy resistance, malignancy and poor prognosis [2,8,18,19,22,23]. Hypoxia is known to be highly heterogeneous within tumors in terms of its spatial distribution, severity and kinetics. Hypoxia arises through different mechanisms associated primarily with limits in oxygen diffusion (chronic hypoxia) and blood perfusion (acute hypoxia). In addition, hypoxia regulates several dif-

ferent cellular pathways that have unique activation kinetics and sensitivity to oxygen concentration. As a consequence, hypoxia-regulated gene expression is complex and displays large temporal characteristics.

With DNA micro-arrays, it is now possible to monitor the expression of several tens of thousands of genes at once. In oncology, this ability is exploited to extract lists of genes (or gene signatures) rather than to rely on a few clinical variables for diagnosis [5,14] or prognosis. For the latter, these gene sets include those derived from clinical data, in which correlation with a supervised classifier identifies the clinical

group with a better or worse prognosis [13,20,21]. More recently, *in vitro* derived gene sets have been described containing genes associated with a particular phenotype hypothesized to be clinically important [1,3,4,9]. This allows an unbiased test of such a hypothesis, by applying the *in vitro* derived signature to a separate patient microarray study. This latter type of study recently demonstrated that a gene signature for hypoxia could act as a prognostic factor in a range of different tumor types. In this latter study, Chi et al. [4] also measured the temporal gene expression programs under hypoxia for several primary cell lines *in vitro*. We hypothesized that because of the large dependency of hypoxic gene expression on time, this data set might be used to extract hypoxic gene signatures that reflect differences between slow and fast hypoxia kinetic responses and their contribution to prognosis. We therefore isolated gene sets correlating with early induction during hypoxia and compared these with genes up-regulated after long hypoxia. Our data indicate that the early hypoxia gene signatures may be useful prognostic tools.

Materials and methods

Data set

The starting material was the data set provided by the study of Chi et al., which represents, to our knowledge, the largest set of time series under hypoxia with 2.4 million of gene expression measurements. Four normal cell lines were used: human coronary artery endothelial cells (ECs), smooth muscle cells (SMCs), human mammary epithelial cells (HMECs), and renal proximal tubule epithelial cells (RPTECs 1 and 2) under two oxygen concentrations (less than 0.02% and 2%). Using cDNA microarrays of 42,000 reporters, gene expression was monitored under hypoxia resulting in 10 time series with at most six time points for each. After they filtered for noise and intensity, the time series of the remaining 4333 reporters was plotted as Fig. 1a [4]. This data set was downloaded from <http://microarray-pubs.stanford.edu/hypoxia/index.htm>. SM and HMEC both had the longest time series, i.e. six time points: 0, 1, 3, 6, 12 and 24 h of hypoxia, each for two oxygen concentrations (less than 0.02% and 2%). The absence of early time points did not allow us to include the other cell lines. As also found by Chi et al., it was not possible to clearly define differential expression for the SM cell line due to a low range of expression between up- and down-regulation. Our analysis thus focused only on the two time series provided for the HMEC cell line: one with an oxygen concentration less than 0.02% (HMEC0), the other under 2% (HMEC2).

A batch query was made with SOURCE (<http://smd.stanford.edu/cgi-bin/source/sourceSearch>) to translate the 4333 CLONEIDs into UniGenes (last accession: February 2007, Build 199). A reporter (probe) was removed if at least one of the following criteria was met: no UniGene identifier was found (328 such reporters); one reporter corresponded to multiple UniGene identifiers (257 such reporters); the Gene Symbol was corresponding to mitochondria (28 such reporters); the UniGene was not for Human (1 such reporter).

For each time series independently, the remaining 3719 reporters were further filtered out if at least one time point was missing. For HMEC0 this resulted in a time series of 1196 reporters represented by 1082 unique genes. For the HMEC2 series, it consisted of 1047 reporters representing 955 unique genes.

Gene expression profiling

Extraction of genes with an up-regulation in early time points was assessed in a supervised way (Fig. 1a). A Pearson correlation was selected as a similarity distance to select profiles based on their time-dependent shape but regardless of their magnitude changes (i.e. an early low or an early high up-regulation is found to be equivalent in these conditions). A curve of interest representing the pattern of gene expression as a function of time was user-defined as a sequence of zeros and ones. The time points 1, 3 and 6 h were considered early time points in the hypoxia response whereas 12 and 24 h were reckoned late time points. The curve of interest for selecting genes with an early up-regulation that return back to basal level for late time points was thus set to be 0-111-00. This template was used to select each gene whose temporal profile was similar to this *a priori* determined pattern of expression. A filtering step was included to require at least a 2-fold induction (with respect to expression under control condition). This procedure was run for each cell line independently (Fig. 1b). A correlation coefficient of 0.6 was selected to provide gene lists of manageable sizes.

We compared the early hypoxia signature (curve of interest: 0-111-00 for control–early–late time points) to a late hypoxia pattern with a curve of interest being: 0-000-11, meaning that gene expression had to be constant and equal to control value during the three early time points, then up-regulated at 12 and 24 h of hypoxia (Fig. 1c). For each time series, three types of gene signatures were derived: early hypoxia, late hypoxia (matching the correlation coefficient (long version), late hypoxia matching the number of reporters (short version)).

Statistical analyses

The gene expression based signatures extracted *in vitro* were evaluated *in vivo* on a large cancer study providing microarray data (downloaded from <http://www.ncbi.nlm.nih.gov/projects/geo/>, Accession No. GSE3494). Clinical annotations for its 251 patients (a subset of the Uppsala cohort) with primary breast cancer (Table 1) treated with surgery plus on indication adjuvant radiotherapy and systemic therapy [13,16]. Expression data were log-transformed and multiple reporters for the same gene symbol were averaged. Depending on the overall expression of the genes selected in the signature, a patient was assigned to either the high expression or the low expression group. Outcome in the two groups was analyzed and compared by the Kaplan–Meier method. Log-rank tests were computed to assess survival differences between the two groups. Univariate and multivariate Cox regression analyses were run with SPSS (SPSS, Chicago, IL.) including all the clinical variables and the early signature under 0%. A machine learning approach was also used to assess the impact of the early signature. It was com-

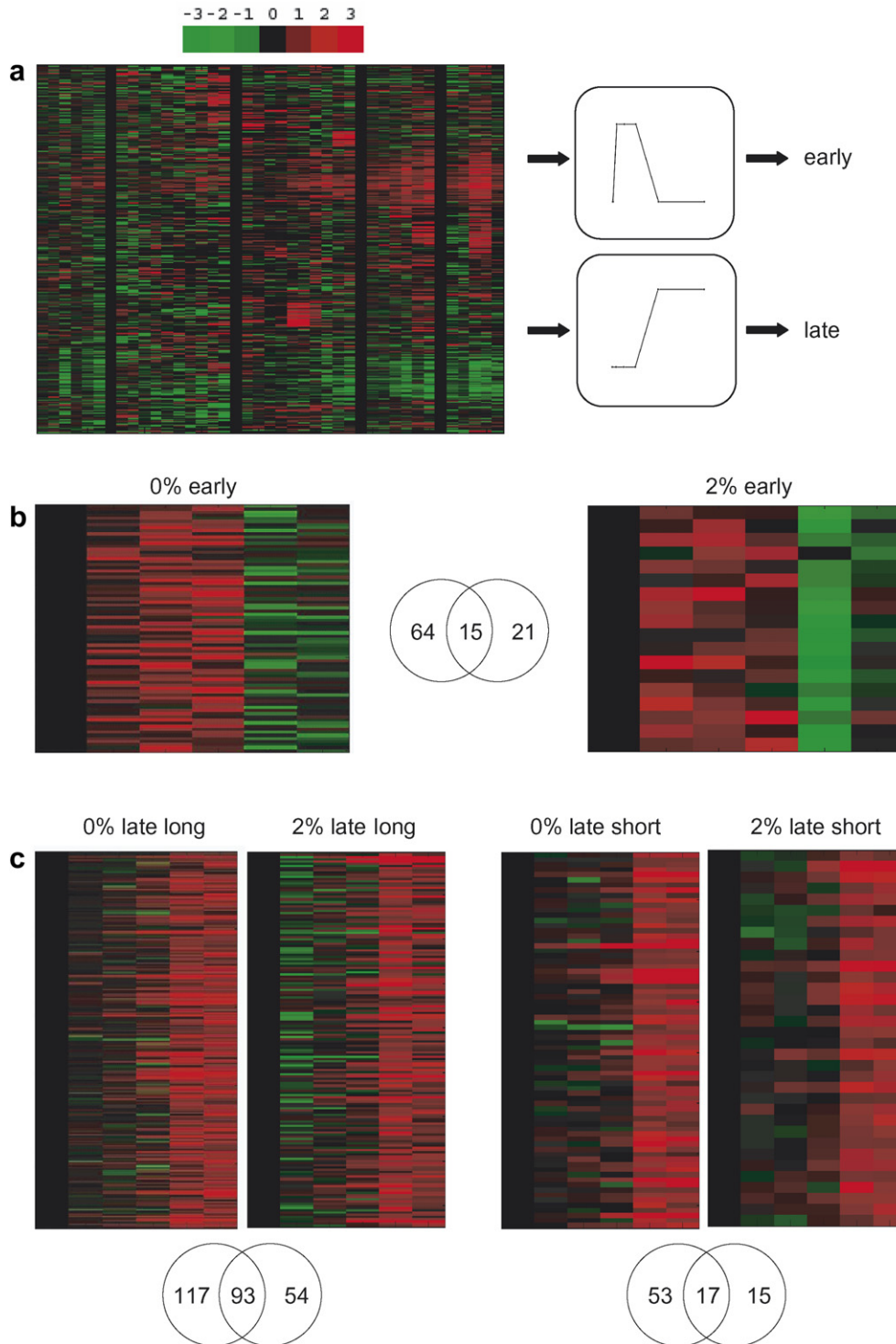


Fig. 1. Log base 2 values of induction (red) and repression (green) with respect to control level (in black, missing data in (a) also): (a) Pipeline of the analysis: from the microarray data from the Chi study. Temporal filters were used to profile: early and late gene signatures for HMEC cell line under 0% and 2% of oxygen concentration. (b) Log induction of the 80 reporters (0%) and 36 reporters (2%) matching the curve of interest for the early hypoxia signature. The UniGene overlap is included. (c) Log induction of the late signatures. For each early signature, two late versions were derived: one matching the correlation coefficient, the other with a similar size. Venn diagrams are used to compare their content.

posed of (1) a feature selection (by leave-one-out cross-validation) followed by (2) a validation stage consisting of a bootstrapping procedure (resampling with replacement)

where a randomly picked 70% of the data was used for training and the remaining 30% of the data was used for testing. This procedure was repeated 100 times and the means and

Table 1
Clinical information for the 251 patients composing the validation set

Characteristic	
No. of patients	251
Age (years)	62.1 ± 13.9
Tumor stage – No. of patients (%)	
Stage I	67 (26.7)
Stage II	128 (51.0)
Stage III	54 (21.5)
Tumor size (mm)	22.4 ± 12.5
ER status – No. of patients (%)	
Positive	213 (84.9)
Negative	34 (13.5)
Lymph node status – No. of patients (%)	
Positive	84 (33.5)
Negative	158 (62.9)
Progesterone receptor status – No. of patients (%)	
Positive	190 (75.7)
Negative	61 (24.3)
p53 status – No. of patients (%)	
Positive	58 (23.1)
Negative	193 (76.9)

standard deviations of the results were reported. Also, multivariate models representing all the possible variable combinations ($2^8 - 1 = 255$) were tested and the top five performing models were retained.

Results

Correlating genes with a predefined pattern of interest was used to derive early hypoxia gene signatures from the HMEC cell lines under 0% and 2% oxygen. The genes reaching a correlation with the required temporal profile greater than 0.6 and a 2-fold induction or more were selected (Table 2). Results from log-rank tests on the Miller data set are reported in Table 3.

HMEC0%

This number of reporters correlating with early induction (early hypoxia gene signature) following exposure to 0% oxygen was equal to 80 (79 unique UniGenes). For the same level of correlation, 241 reporters (210 unique UniGenes) were found for the late hypoxic signature (late long hypoxic signature). For comparison, we also increased the required correlation to reduce the number of reporters to match a size similar to the early signature. This late short hypoxic signature contained 74 reporters (70 unique UniGenes). Of these three signatures, only the early hypoxia signature was significant on the breast cancer data set ($p = 0.004$ for difference in survival). Surprisingly, the late signatures were not significant and showed no evidence of splitting the patients into different prognostic groups ($p = 0.118$ and 0.110 for the long and short versions, respectively) (Fig. 2).

HMEC2%

Under 2%, the number of selected reporters was smaller than under 0%. The early gene signature contained 36

reporters (36 unique UniGenes) exhibiting an up-regulation within the first hours (1–3–6 h) following hypoxia exposure. The late long hypoxia signature derived with a correlation threshold of 0.6 was made of 169 reporters (147 unique UniGenes). For the short version, the correlation threshold was raised to yield a size comparable to the early one. It was composed of 34 reporters (32 UniGenes). Similar to the results under 0%, the early hypoxia signature was the only significant signature ($p = 0.034$) while the long and short versions of the late hypoxia signature were not (p -value of 0.919 and 0.842, respectively).

Combination of signatures (0% and 2%)

Time series without missing data HMEC under 0% (1082 UniGenes) and 2% (955 UniGenes) had a large overlap: 793 UniGenes. Comparing the 2 early signatures revealed a small overlap (15 UniGenes). This core HMEC response was also tested and had a p -value of 0.005 (log-rank test). The overlap between the two long late signatures showed an overlap of 93 UniGenes. This core signature for late response was tested and found to be not significant ($p = 0.240$) on the breast cancer data set.

In order to link the genes from the early signatures to biological functions, Gene Ontologies analyses were run with Ingenuity Pathways Analysis (Ingenuity Systems, Redwood City, CA). The results provided for the top functions are shown in Table 4. For both early signatures, only two top functions were found. The genes selected for enrichment ('focus genes') are listed and mainly involved in proliferation, cell cycle and cancer. For the core early response (intersection of 0% and 2%), the top function was related to amino acid metabolism.

Prognostic power of the early signature (HMEC0)

A univariate Cox-regression analysis was run for the early signature (HMEC0) on the Miller data set with all the variables provided and the early hypoxia score. The signature score was found to be significant (p -value < 0.05), while PgR status, ER status and age were not.

From the multivariate analysis (with a stepbackward feature selection procedure), the early hypoxia signature ($p = 0.254$) was found to be the 4th best prognostic factor after lymph node status ($p = 0.002$), tumor size ($p = 0.016$) and Elston grade ($p = 0.111$). On this data set, it indeed provided more information than ER status or p53 status (both individually and when combined with other variables) for instance.

A multivariate analysis with a machine learning-based feature selection approach was also run. Providing all the variables to the algorithm (but age and the early signature score) gave an Area Under the Curve (AUC) of 62.3 with a standard deviation (std) of 0.07. Repeating the same analysis but adding age only gave an AUC of 62.2 with a std of 0.07. When age and the signature score were added together, the AUC was 62.5 (std of 0.07). When only the signature score was added, the AUC was 66.1 (std of 0.06). From the hundred runs computed for each analysis, a t -test was performed between the six variables and the signature score versus the six variables and the age. A p -value of $4.8e-6$ suggested that age is less effective than the signature score on this data set.

Table 2

Content of the early and late hypoxia gene signatures in terms of UniGene cluster IDs: late hypoxia signatures are represented by their short versions only

Early 0%				Late 0% short				Early 2%			
UniGene	Symbol	Corr.	Max	UniGene	Symbol	Corr.	Max	UniGene	Symbol	Corr.	Max
Hs.592692		0.95	1.59	Hs.102267	LOX	1.00	1.26	Hs.72550	HMMR	0.90	2.39
Hs.633514	TIMP2	0.92	3.49	Hs.435051	CDKN2D	1.00	1.16	Hs.528299	HTATIP	0.90	1.84
Hs.127126	CPEB4	0.91	1.80	Hs.520819	INSIG1	0.99	2.18	Hs.5 10078	SGK	0.86	1.58
Hs.334587	RBPMS	0.90	1.35	Hs.380906	MYADM	0.98	1.56	Hs.6 33514	TIMP2	0.86	3.83
Hs.648626		0.89	1.09	Hs.102267	LOX	0.98	2.38	Hs.6 02706	LOC6 45591	0.86	1.63
Hs.646346	GAS6	0.87	1.65	Hs.523012	DDIT4	0.98	1.69	Hs.642877	MALAT1	0.85	2.46
Hs.584803	ST3GAL1	0.87	1.45	Hs.465870	KEAP1	0.97	1.42	Hs.5 93232		0.84	3.25
Hs.567495	TRNT1	0.85	1.80	Hs.649390		0.97	1.20	Hs.5 96783		0.82	1.59
Hs.651126	DUSP3	0.85	1.50	Hs.173381	DPYSL2	0.97	2.70	Hs.441113	MAGEA6	0.80	1.06
Hs.478746	CENTB2	0.84	2.60	Hs.287659	STRBP	0.97	2.70	Hs.149983	PEX14	0.79	4.42
Hs.72550	HMMR	0.83	1.81	Hs.131433	ADAMT513	0.97	3.65	Hs.94542	ALKBH1	0.79	1.18
Hs.154276	BACH1	0.83	1.18	Hs.443976	CEP250	0.97	1.52	Hs.154276	BACH1	0.77	2.12
Hs.233568	HIST1H2AL	0.82	2.15	Hs.235782	SLCO4A1	0.96	2.24	Hs.1 89772	CCT2	0.77	4.09
Hs.106861	NSD1	0.81	1.51	Hs.405662	CRABP2	0.96	1.33	Hs.112432	AMH	0.76	2.54
Hs.414418	BET1L	0.80	3.41	Hs.34871	ZFHX1B	0.96	2.58	Hs.643599	PAPPA	0.74	1.51
Hs.593565		0.80	4.08	Hs.269722	TCBA1	0.96	1.65	Hs.1 26774	DTL	0.74	1.76
Hs.235116	GRK6	0.79	1.94	Hs.540696	SLC6A8	0.96	3.26	Hs.1 46406	NIT1	0.73	2.04
Hs.554791	TP53111	0.79	1.65	Hs.644065	TSC22D2	0.96	1.43	Hs.463838	LOC440459	0.73	1.25
Hs.226780	OSTM1	0.79	1.86	Hs.129003		0.95	2.49	Hs.5 23847	IFI6	0.72	4.45
Hs.525549	BTBD7	0.79	2.68	Hs.502116	NAV2	0.95	2.02	Hs.4 6700	ING1	0.72	1.08
Hs.536158	PARG	0.79	2.83	Hs.511915	ENO2	0.95	4.10	Hs.55131	C3orf23	0.71	1.21
Hs.436489	ATP5S	0.79	1.12	Hs.530381	PIM3	0.95	2.03	Hs.558396	SCD	0.70	1.64
Hs.643279	EIF4EBP2	0.79	2.96	Hs.530381	PIM3	0.95	2.03	Hs.148907	SLC5A12	0.70	4.49
Hs.189772	CCT2	0.77	3.24	Hs.379821	FAM83A	0.95	1.36	Hs.643920	C1GALT1C1	0.69	2.17
Hs.78977	PCSK1	0.76	2.18	Hs.96996	HNRPAO	0.95	2.50	Hs.1 60556	ACACA	0.67	1.78
Hs.155983	JMJD2A	0.76	2.03	Hs.26010	PFKP	0.94	1.56	Hs.5 62083	ICMT	0.66	2.27
Hs.612872	TLL5	0.76	2.23	Hs.75093	PLOD1	0.94	2.01	Hs.515383	LOC644242	0.66	4.93
Hs.435933	PHF10	0.76	1.55	Hs.525704	JUN	0.93	2.15	Hs.5 13430	CDR2	0.65	2.10
Hs.489603	ATXN7L1	0.76	3.60	Hs.446240	PRKCBP1	0.93	1.44	Hs.155983	JMJD2A	0.63	2.71
Hs.128959	PCF11	0.75	3.40	Hs.89387	CASC2	0.93	1.96	Hs.590575	GRM3	0.62	3.65
Hs.335205	SSH2	0.74	2.84	Hs.288232	SERINC5	0.93	1.03	Hs.632226	ITGB4	0.62	1.49
Hs.146406	NIT1	0.74	1.81	Hs.533887	SMEK1	0.93	1.39	Hs.1 65607	FLJ25416	0.62	1.02
Hs.596783		0.73	1.14	Hs.78771	PGK1	0.93	2.41	Hs.514033	SPAG5	0.61	1.92
Hs.512973	PTPLAD1	0.73	1.71	Hs.108106	UHRF1	0.93	1.46	Hs.6 32447	WDR42A	0.61	1.87
Hs.461030	GFOD2	0.73	1.99	Hs.644649		0.93	3.49	Hs.8 9603	MUC1	0.60	2.11
Hs.149983	PEX14	0.73	4.49	Hs.445402	PCTK3	0.93	3.23	Hs.4 34961	ATXN1	0.60	2.34
Hs.1310.4	ACOX1	0.73	2.21	Hs.535297	ARID5B	0.93	2.73				
Hs.292524	CCNH	0.72	3.27	Hs.515032	MKNK2	0.92	2.76				
Hs.283749	RNASE4	0.71	1.74	Hs.422113	ZNF511	0.92	1.80	Late 2% short			
Hs.287362	TLE3	0.71	4.38	Hs.133350	CNOT2	0.92	2.65	UniGene	Symbol	Corr.	Max
Hs.492203	TERT	0.71	1.20	Hs.146688	PTGES	0.92	1.15	Hs.446017	WSB1	1.00	1.77
Hs.250693	ZNF117	0.71	1.04	Hs.460355	PRKCB1	0.92	2.79	Hs.3 70365	HK1	1.00	1.13
Hs.593232		0.70	3.15	Hs.369520	SYTL2	0.92	4.32	Hs.5 20819	INSIG 1	0.97	1.51
Hs.590575	GRM3	0.70	3.26	Hs.647120	MLL3	0.92	4.44	Hs.405662	CRABP2	0.96	2.02
Hs.428214	MAML2	0.70	3.13	Hs.517145	ENO1	0.92	1.62	Hs.173705	LOC401152	0.94	2.24
Hs.523847	IFI6	0.70	4.68	Hs.533782	KRT8	0.91	1.83	Hs.644649		0.94	2.91
Hs.533712	RBM4	0.70	2.29	Hs.21691	GPR75	0.91	1.55	Hs.446017	WSB1	0.94	1.32
Hs.44067	C12orf53	0.69	1.69	Hs.458513	PPP1R3B	0.91	2.57	Hs.511915	ENO2	0.94	3.18
Hs.647072	PRKAG2	0.68	1.59	Hs.643452	DGCR8	0.91	2.82	Hs.4 74935	SEMA4B	0.94	2.40
Hs.606472		0.68	3.93	Hs.591443	RAVER2	0.91	1.14	Hs.497822	DUSP10	0.94	2.06
Hs.149032	PIK3R4	0.68	1.21	Hs.433146	SLC6A10P	0.91	3.62	Hs.3 35614	SEC14L2	0.93	2.59
Hs.436705	KIAA1219	0.68	2.10	Hs.429879	EHHADH	0.91	4.21	Hs.102267	LOX	0.93	2.47
Hs.631539	RAB4B	0.67	2.42	Hs.9613	ANGPTL4	0.91	4.95	Hs.287659	STRBP	0.93	1.78
Hs.529353	ACSS1	0.67	2.55	Hs.494529	FANCC	0.91	1.94	Hs.5 02116	NAV2	0.91	1.82
Hs.592020	IGF1R	0.67	4.02	Hs.523012	DDIT4	0.91	2.14	Hs.581021	SIR PA	0.91	1.08
Hs.642938	IGFBP1	0.67	1.63	Hs.97858	KIF1B	0.90	2.55	Hs.403933	FBXO32	0.91	2.85
Hs.631930	LOC64645 0	0.67	2.42	Hs.581355	EIF4E3	0.90	3.84	Hs.5 00047	P4HA1	0.91	2.97
Hs.148907	SLC5A12	0.66	5.14	Hs.530904	CSRP2	0.90	1.32	Hs.540696	SLC6A8	0.90	2.43
Hs.160556	ACACA	0.65	1.63	Hs.159430	FNDC3B	0.90	1.73	Hs.2795	LDHA	0.89	1.38
Hs.126891	IL22RA2	0.64	1.63	Hs.435667	USP34	0.90	1.93	Hs.9613	ANGPTL4	0.89	3.64

Table 2 (continued)

Early 0%				Late 0% short				Late 2% short			
UniGene	Symbol	Corr.	Max	UniGene	Symbol	Corr.	Max	UniGene	Symbol	Corr.	Max
Hs.124011		0.64	1.70	Hs.505172	LOC645619	0.90	1.93	Hs.535297	ARID5B	0.89	2.28
Hs.524828	ZNF664	0.64	3.12	Hs.446017	WSB1	0.90	2.09	Hs.173381	DPYSL2	0.89	2.19
Hs.4779	GATAD2B	0.64	2.28	Hs.470633	PDK1	0.89	2.22	Hs.429879	EHHADH	0.89	2.88
Hs.612872	TLL5	0.63	1.63	Hs.501023	MXI1	0.89	4.06	Hs.4 25144	MTMR11	0.89	1.74
Hs.233240	COL6A3	0.63	1.19	Hs.501023	MXI1	0.89	1.92	Hs.98643	RAP2B	0.89	1.44
Hs.445030	RHOBTB3	0.63	3.76	Hs.46423	HIST1H4C	0.89	1.75	Hs.34871	ZFH1B	0.88	2.08
Hs.460	ATF3	0.63	1.79	Hs.433146	SLC6A10P	0.89	3.61	Hs.591140	FOXK2	0.88	2.09
Hs.112432	AMH	0.63	2.24	Hs.591849	C8orf4	0.89	2.18	Hs.581355	EIF4E3	0.87	2.15
Hs.544738	LY86	0.63	3.32	Hs.159195	DOCK1	0.89	3.02	Hs.1 55247	ALDOC	0.87	1.38
Hs.530941	C12orf30	0.63	3.66	Hs.372914	NDRG1	0.89	5.68	Hs.501023	MXI1	0.87	2.77
Hs.180903	NCAPH2	0.63	1.81	Hs.80 04	KALRN	0.89	2.67	Hs.5 36075		0.87	1.41
Hs.76364	AIF1	0.62	3.06	Hs.523789	TncRNA	0.89	2.14	Hs.1 02267	LOX	0.86	1.79
Hs.43627	SOX12	0.62	6.59	Hs.634882	ARL6IP1	0.88	2.39	Hs.135213	HSD17B12	0.86	3.51
Hs.643599	PAPPA	0.62	1.96	Hs.585433	KIRREL	0.88	2.60	Hs.1 32342	LPIN2	0.86	1.87
Hs.125038	FAM92A1	0.61	1.82								
Hs.131342	CCL26	0.61	2.59								
Hs.512767	DKFZP761H1	0.61	2.67								
Hs.631974	LOC728488	0.60	1.31								
Hs.112873	IGSF11	0.60	1.65								
Hs.6217		0.60	1.93								

Lists are sorted by descending correlation coefficient with the curve of interest. Maximum induction (in log scale) is also provided.

Table 3

Validation of the early hypoxia gene signatures: statistical differences between Kaplan–Meier survival curves were assessed by computed mean *p*-values obtained from log-rank tests at 5 and 10 years

Oxygen concentration (%)	Hypoxia type	Corr. coeff.	No. of reporters	No. of unique UniGenes	<i>p</i> -value
0	Early	0.6	80	79	0.004
0	Late	0.6	241	210	0.118
0	Late	0.88	74	70	0.110
2	Early	0.6	36	36	0.034
2	Late	0.6	169	147	0.919
2	Late	0.86	34	32	0.842
0 & 2	Chi et al.	n.a.	253	171	0.300
0 & 2	Early	0.6	15	15	0.005
0 & 2	Late	0.6	93	93	0.240

For each signature, its type, correlation coefficient (not applicable for the Chi study), number of reporters and number of unique UniGene cluster IDs are also given.

It is also important to note that when all possible variable combinations were considered for training a multivariate model, four out of the top five performing combinations included the signature score.

Discussion

Early hypoxia

In this paper, we focus on the impact of the early response to hypoxia which with the exception of a few studies (e.g. [12]) is rarely investigated by means of microarrays. Due to potential differences arising for different oxygen concentrations, the two time series were treated independently. Both early hypoxia signatures were significant ($p < 0.01$ at 0%; $p < 0.05$ at 2%) whereas none of the late signatures were. A signature of the 15 common UniGenes (13 symbols) found in the two early signatures was also tested and found to be significant ($p = 0.005$).

Ingenuity Pathways Analysis identified genes within this list including ACACA, AMH, BACH1, CCT2 (related to growth and apoptosis) and IFI6 in Cancer, Tumor Morphology, Amino Acid Metabolism (Table 4). Focusing on the early signature at 0% ($p = 0.004$) with Ingenuity Pathways Analysis revealed a significant contribution of IGF1R (from the canonical insulin pathway), TERT (Telomerase reverse transcriptase), AMH, COL6A3, and ACACA in apoptosis, growth and proliferation. CCNH (from the estrogen receptor pathway) as well as transcription factors ATF3 and BACH1 were also identified. The early signature may reflect activation of the Unfolded Protein Response (ATF3 for instance), which is known to be activated rapidly in response to hypoxia. Cross-talk between oxygen and glucose metabolism via HIF1 have been reported [11,15] as well as the protective role of IGF1 against cell degeneration in animal models after stroke such as hypoxia–ischemia [17].

Early hypoxia gene signature

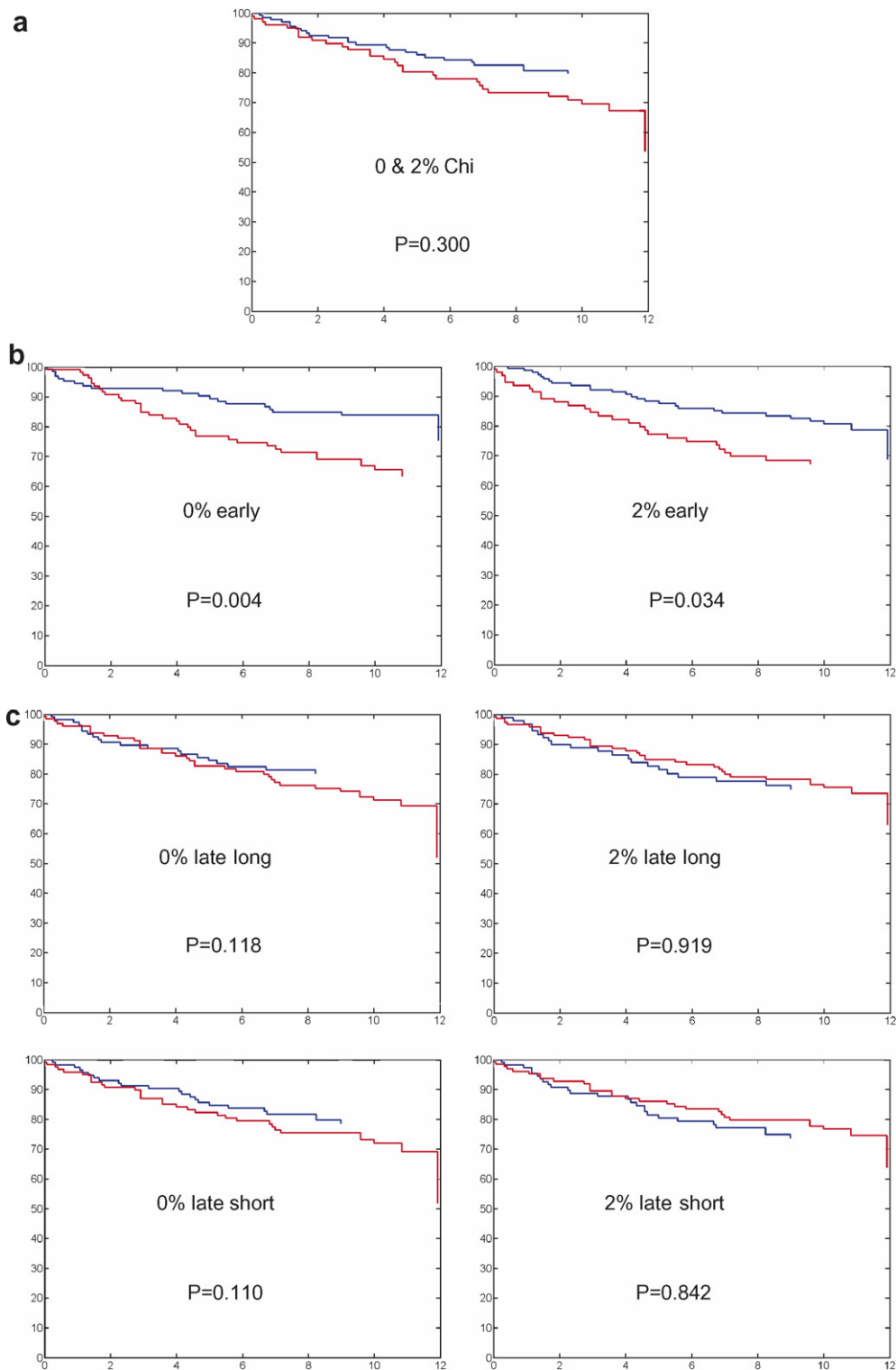


Fig. 2. Kaplan–Meier survival curves for disease specific survival (follow-up to 12 years) and p -values for the different gene signatures. The group with a low (high) expression of the signature is plotted in blue (red). (a) Chi signature, (b) early hypoxia (under 0% or 2%), and (c) late signatures.

Late hypoxia

The late signatures (0% and 2%) were found to be very similar to the Chi signature, with an overlap for each one greater than 40% (and up to 70% with their short versions). The biological theme of the late signatures is also typical

of a hypoxic response [6], with genes encoding proteins involved in the hypoxia pathway and angiogenesis (ANGPTL4, CA9, DDIT4, EGLN3, EGFR, HK2, HIG2, LOX), cell proliferation and apoptosis (BNIP3L, NDRG1, MXI1), glucose transport (many members of the solute carrier family including

Table 4
Gene Ontology provided by Ingenuity for the early signatures

Oxygen concentration	Top functions	Score	Focus genes	Gene symbols
0%	Cellular growth and proliferation, cellular development, connective tissue development and function	27	14	AIF1, ATF3, BACH1, CCL26, COL6A3, GATAD2B, GRK6, IL22RA2, PARG, PCSK1, PEX14, PIK3R4, TERT, TP53I11
0%	Cellular function and maintenance, cancer, cell cycle	25	13	ACACA, ACOX1, ACS1, AMH, CCNH, CCT2, GRM3, HMMR, IFI6, IGF1R, LY86, PTPLAD1, TLE3
2%	Cellular function and maintenance, cell cycle, cancer	27	12	ACACA, AMH, BACH1, CDR2, HTATIP, IFI6, ING1, ITGB4, PEX14, SCD, SGK, SPAG5
2%	Cellular movement, connective tissue development and function, cancer	9	5	CCT2, GRM3, HMMR, ICMT, MUC1
0% & 2%	Cancer, tumor morphology, amino acid metabolism	11	5	ACACA, AMH, BACH1, CCT2, IFI6

For each one, the top two functions which correspond to the highest scores are shown, as well as the genes contributing to the score. The top function of early genes overlapping between 0% and 2% are also given.

GLUT1), metabolism (ENO1, PGK1) and adrenomedullin (ADM). Similar to the four derived late hypoxia signatures, the Chi signature was found to be not significant in this data set. This suggests that late signatures are very similar to the Chi signature not only in gene content but also in terms of prognostic power.

Unsupervised and supervised hypoxia signatures

Cluster analysis [7] is a commonly used technique [3,10] to assess shared functions and common regulation in an unbiased way since the algorithm does not require user-defined (i.e. supervised) assumptions on the numbers or type of clusters to be found. The underlying idea is that genes with a similar expression pattern are likely to also be involved in the same regulatory process (referred to as 'Guilt by association'). This approach was used by Chi et al., to group temporal gene expression under hypoxia. From HMECs and RPTECs, one cluster of global up-regulated genes was then selected after visual inspection to form the 'epithelial cell hypoxia genes' or so-called Chi signature. The gene content of the early hypoxia signatures had virtually no overlap with the Chi signature: 2 UniGenes at 0% (Hs.149032, PIK3R4 and Hs.283749, RNASE4), and only 1 at 2% (Hs.94542, ALKBH1). This suggested that the early signatures are very different from the Chi signature.

Although external changes in milieu (e.g. hypoxia) are sensed very rapidly by the cell, large changes at the transcriptional level occur mostly after several hours. Our supervised method was developed to extract a handful of genes involved in the early hypoxic response (i.e. from 1 to 6 h of hypoxia). Overlap was seldom found between the early and the late signatures: only one UniGene at 0% (Hs.106861, NSD1), and none under 2%. It shows that early signatures are different from late signatures (including the Chi signature) both in gene content and in patients they select.

All early hypoxia signatures whether extracted under 0%, 2% or their overlap were found significant on the Miller data

set. Focusing on the former for multivariate analyses confirmed its prognostic power. At this moment it is unclear why the early hypoxic signatures provide superior prognostic information in this data set. It may reflect differential activation of hypoxia response pathways, but may also select patients with more aggressive tumors (more metastasis), resistance to treatment or both. We are currently investigating additional methods to validate our early gene signatures such as *in vitro* experiments with RNAi to selectively down-regulate gene products (e.g. TERT or IGF1R) or by further testing the gene signature on independent clinical data sets with distinct clinical features. This retrospective study could be useful in the field of radiotherapy and oncology by suggesting the need for a prospective clinical study providing clinical data together with microarray data and an independent measure of hypoxia (by Eppendorf electrodes or immunochemistry staining for instance).

Conclusions

This analysis shows that, consistent with our hypothesis, the early and late hypoxia responses are very different at the transcription level. In a relatively old cohort of patients with primary breast cancer treated by locoregional therapy and systemic adjuvant therapy, it was shown that early hypoxia signatures, but not late hypoxia responses, could correlate with survival differences. It suggests that gene signatures can provide a means to select patients for individualized therapy.

Acknowledgements

We acknowledge financial support from Siemens (to R.S. and M.H.W.S.), the Dutch Science Organization (ZonMW-NWO Top Grant 912-03-047 to B.W.), the Dutch Cancer Society (KWF Grant UM 2003-

2821 to B.W.), and the EU 6th framework program (Euroxy program to B.W.).

* **Corresponding author. Dr. Philippe Lambin**, Department of Radiation Oncology (MAASTRO), GROW Research Institute, University Hospital Maastricht, Dr. Tanslaan 12, 6229 ET Maastricht, The Netherlands. *E-mail address:* philippe.lambin@maastro.nl

Received 27 March 2007; received in revised form 1 May 2007; accepted 1 May 2007; Available online 25 May 2007

References

- [1] Bild AH, Potti A, Nevins JR. Linking oncogenic pathways with therapeutic opportunities. *Nat Rev Cancer* 2006;6:735–41.
- [2] Bussink J, Kaanders JH, van der Kogel AJ. Microenvironmental transformations by VEGF- and EGF-receptor inhibition and potential implications for responsiveness to radiotherapy. *Radiother Oncol* 2007;82:10–7.
- [3] Chang HY, Sneddon JB, Alizadeh AA, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2004;2:E7.
- [4] Chi JT, Wang Z, Nuyten DS, et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med* 2006;3:e47.
- [5] Chung CH, Bernard PS, Perou CM. Molecular portraits and the family tree of cancer. *Nat Genet* 2002;32:533–40.
- [6] Denko NC, Fontana LA, Hudson KM, et al. Investigating hypoxic tumor physiology through gene expression patterns. *Oncogene* 2003;22:5907–14.
- [7] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998;95:14863–8.
- [8] Eriksen JG, Horsman MR. Tumour hypoxia – a characteristic feature with a complex molecular background. *Radiother Oncol* 2006;81:119–21.
- [9] Huang ES, Black EP, Dressman H, West M, Nevins JR. Gene expression phenotypes of oncogenic signaling pathways. *Cell Cycle* 2003;2:415–7.
- [10] Iyer VR, Eisen MB, Ross DT, et al. The transcriptional program in the response of human fibroblasts to serum. *Science* 1999;283:83–7.
- [11] Kietzmann T, Kronen-Herzig A, Jungermann K. Signaling cross-talk between hypoxia and glucose via hypoxia-inducible factor 1 and glucose response elements. *Biochem Pharmacol* 2002;64:903–11.
- [12] Koritzinsky M, Seigneuric R, Magagnin MG, van den Beucken T, Lambin P, Wouters BG. The hypoxic proteome is influenced by gene-specific changes in mRNA translation. *Radiother Oncol* 2005;76:177–86.
- [13] Miller LD, Smeds J, George J, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA* 2005;102:13550–5.
- [14] Ramaswamy S, Tamayo P, Rifkin R, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 2001;98:15149–54.
- [15] Riedemann J, Macaulay VM. IGF1R signalling and its inhibition. *Endocr Relat Cancer* 2006;13:S33–43.
- [16] Sjogren S, Inganas M, Norberg T, et al. The p53 gene in breast cancer: prognostic value of complementary DNA sequencing versus immunohistochemistry. *J Natl Cancer Inst* 1996;88:173–82.
- [17] Smith PF. Neuroprotection against hypoxia-ischemia by insulin-like growth factor-I (IGF-I). *IDrugs* 2003;6:1173–7.
- [18] Sorensen BS, Hao J, Overgaard J, et al. Influence of oxygen concentration and pH on expression of hypoxia induced genes. *Radiother Oncol* 2005;76:187–93.
- [19] Troost EG, Laverman P, Kaanders JH, et al. Imaging hypoxia after oxygenation-modification: comparing [¹⁸F]FMISO autoradiography with pimonidazole immunohistochemistry in human xenograft tumors. *Radiother Oncol* 2006;80:157–64.
- [20] van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.
- [21] 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.
- [22] Vaupel P. The role of hypoxia-induced factors in tumor progression. *Oncologist* 2004;9:10–7.
- [23] Wouters BG, van den Beucken T, Magagnin MG, Lambin P, Koumenis C. Targeting hypoxia tolerance in cancer. *Drug Resist Updat* 2004;7:25–40.