Supplementary Material

Hypoxic gene expression is a prognostic factor for disease free-survival in a cohort of locally advanced squamous cell cancer of the uterine cervix

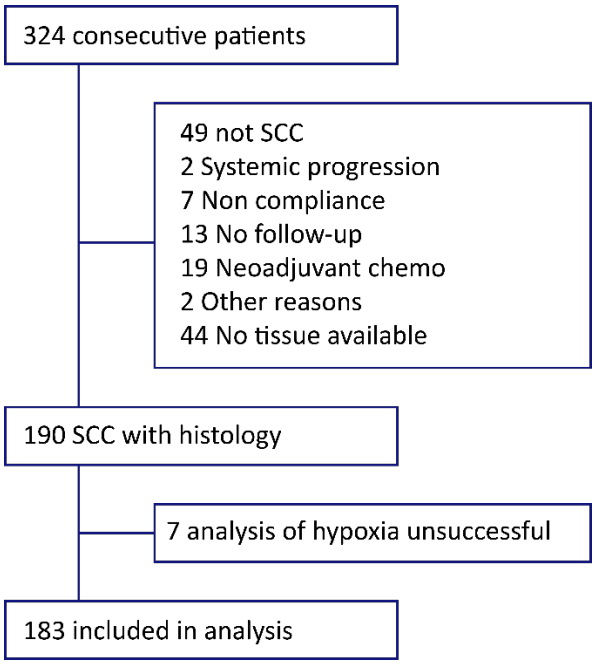
Jan Alsner, Jens Overgaard, Trine Tramm, Jacob Chr. Lindegaard

Supplementary Figure 1

Supplementary Document

Supplementary Figure 1

Patient flow chart.



Supplementary Document

Hypoxia classification

Study design

The study is designed to test the hypothesis that a 15-hypoxia gene expression classifier developed for head and neck cancer [1] can also be used as a prognostic marker for disease-free survival in locally advanced cervical cancer patients with squamous cell carcinoma.

To avoid issues with selecting an optimal cut-off for cervical cancer, the classification into 'more' or 'less' hypoxic samples is performed by the same validated method [2] currently used in a prospective clinical trial on patients with head and neck cancer (NCT02661152). Thus, the same classification of whether a sample is 'more' or 'less' hypoxic is being used, and the cut-off is therefore pre-defined.

Like in the ongoing head and neck cancer trial, future classifications can be performed prospectively on single samples.

Background

There are several challenges when developing a gene expression classifier for hypoxia in routine formalin-fixed paraffin-embedded (FFPE) tumour biopsy material. The following sections very briefly describes how these have been addressed leading up to the final classifier.

Induction by hypoxia, *in vitro* (pH independent)

Initial studies demonstrated that pH had a major impact on some hypoxia induced genes, but not all [3,4]. A later study used gene expression arrays to identity 27 genes that were induced by hypoxia, independently of pH, in human cell lines [5].

Expression at normoxic conditions

Care was taken to include only hypoxia induced genes that were expressed at a relatively high level at normoxic conditions [5]. This is one of the important steps to minimize issues with missing values when analysing old FFPE material (another important step is to develop a classifier that can handle missing values for some of the genes, see details below). In a

previous study, we have analysed the loss of amplifiable mRNA per year since tumour biopsies are stored in FFPE, and estimated a half-life of 4.6 years [6].

Identifying reference genes

From the gene expression data set [5], and based on potential reference genes [4,7], three optimal reference genes (Table 1) were selected based on the similarities in expression patterns across FFPE material from head and cancer patients [1].

Induction by hypoxia, *in vivo* (xenograft models)

The hypoxia induction of the genes identified in [5] was then validated in xenograft tumour models from human squamous cell carcinoma cell lines by correlating high expression levels of the genes with tumour regions of high uptake of a hypoxia tracer [1].

Identifying genes which expression correlates with physical levels of pO_2 in human tumours and development of the classifier

Using FFPE biopsies from head and neck cancer patients (same samples as used for selection of reference genes) where pO_2 levels had been measured by polarographic O_2 needle electrodes [8,9], it was possible to identify 15 of the genes from [5] (Table 1) which could separate a group of 'more' hypoxic tumours from a group of 'less' hypoxic tumours (based on the pO_2 levels). The mean expressions of the 15 genes for the groups of 'more' hypoxic and 'less' hypoxic patients, respectively, is shown in Table 2. Together with the estimated common variance, these mean values form the basis of the classification method. Briefly, the expression pattern in a sample is compared to the two sets of mean values, and the sample is then classified based on which group it resembles the most (see details below). The classifier was developed so that it could classify samples individually and prospectively (e.g. as patients are entering a trial) and handle missing values [1].

Validating the prognostic and predictive (benefit of hypoxic modification) potential

Using samples from a randomized trial on hypoxic modification in head and neck cancer [10], the prognostic potential (significant association of classifier status with loco-regional control in patients treated without hypoxic modification) and predictive potential (significant interaction between classifier status and hypoxic modification on loco-regional control) was identified [1,11].

Repeatability

The repeatability and the technical variability was established from all steps in qPCR procedures by independent replicate analysis of the same samples [2].

Reproducibility

Tumour xenografts were used to test for the influence of variation in fixation procedure on gene expression levels and reproducibility of the procedure [2]. This was performed by sending halves of frozen xenograft tumours to seven different pathology departments across Europe, where the samples were fixed, embedded in paraffin, cut, and a section was returned for classification. Comparisons with classifications on the remaining halves of the tumours kept locally demonstrated that variations in fixation procedures between these institutions did not affect classification.

Classification of samples where gene expression is not measured on the same platform and/or with different primer sets (normalization)

The mean gene expression values for 'more' and 'less' hypoxic samples can only be used if the sample to be classified is analysed on the same platform using the same primer sets. Based on a previously reported method of normalizing gene expression values from one platform to another [12], a representative set of head and neck cancer patients was used to obtain values that allow the expression levels of current samples to be compared with the overall expression levels in the original study [2] (Table 3). Thus, current head and neck cancer samples can be compared directly to characteristic expression levels of 'more' and 'less' hypoxic samples identified from the original pO_2 study (Table 2).

If samples are to be classified by a different primer set and/or on a different platform, a group of representative samples needs to be measured on the new platform to obtain the mean and SD values that are used in the normalization (se details below).

Figure 1 of the main paper shows the mean expression values for the head and neck cancer samples currently used for normalization of head and neck cancer samples and the mean values for the cervical cancer patients in the present study. The samples are measured on the same platform using the same primer sets, and the expression levels are highly

correlated. In the classification of the cervical cancer samples, the mean values of the cervical cancer samples are used for normalization.

Heterogeneity within classifier genes

As part of the validation of the classifier, classifications were performed individually on each of the genes included in the classifier and the results were matched with the overall classification. On average, single gene classifications matched the overall classification in 72% of the cases (range 60–80%) [2].

Intra-tumour heterogeneity

If more than one paraffin block is available for a given patient, then classification is performed on the section with the highest tumour content. Intra-tumour heterogeneity was addressed in 20 tumours where 2–4 different biopsies from each tumour were accessible [2]. In 14 of the 20 tumours there was full concordance between classification of all samples from the same tumour, and in 4 tumours, the biopsy containing the relatively highest proportion of tumour tissue was classified as more hypoxic. In the remaining 2 tumours, the biopsy with the highest tumour tissue content was classified as 'less' hypoxic, but another biopsy with slightly less relative tumour tissue content was classified as 'more' hypoxic.

Minimum tumour content

To examine a possible limit of detection linked to the relative tumour content of a biopsy, 129 samples from 96 patients were compared [2]. There was no lower level where samples could not be classified as' more hypoxic', and in biopsies with 0-20%, 21-40% and 41-60% tumour tissue content, the frequency of 'more hypoxic' tumours was stable. In biopsies with more than 60% tumour tissue content, there was a tendency for a higher frequency of 'more' hypoxic' tumours.

Gene expression

Gene expression is measured as previously described [1,2]. RNA from FFPE tumour biopsies are extracted by using a fully automated, bead-based RNA isolation procedure [13]. cDNA is generated by using the High Capacity cDNA kit (Thermo Fischer Scientific) and gene

expression is quantified by using qPCR. cDNA based on FFPE samples is pre-amplified according to the manufacturer's details (TaqMan PreAmp, Thermo Fischer Scientific) before real time qPCR. To detect transcripts of interest, TaqMan Gene Expression assay (Thermo Fischer Scientific) is used for all potential classifier and reference genes. ΔCt values are generated by normalizing to the geometric mean of the 3 reference genes. Ct values more than 35 or with an SD more than 0.3 are dismissed and interpreted as empty wells. Gene expression levels are quantified as $2^{-\Delta Ct}$ and log2-transformed.

Classification

Classification is performed as described previously [1,2]. For each of the 15 genes in a new sample, distances are calculated between the expression values for the new sample and the mean values in the two pre-defined groups ('more' and 'less' hypoxic) from the original study with pO_2 measurements, and the distances are weighted by an estimate of the common variance in the original study. Finally, the values for each gene in each group is summed up. Thus, two distances, D_{more} and D_{less}, are calculated for each new sample as

$$D_{in} = \sum_{m \in C} \frac{(y_{mn} - \overline{z}_{im})^2}{W_m} \qquad \text{(a)}$$

where the gene $m \in C$, where C refers to the 15 genes in the classifier (Table 1), i is group ('more' or 'less' hypoxic), $\overline{z}$ is the mean expression for the group (Table 2), W is the estimated common variance (Table 2), and y is the gene expression of the classified sample, n.

The tumour is classified to the group it resembles the most, i.e. a sample is classified as 'more' hypoxic if $D_{more} < D_{less}$, and 'less' hypoxic if $D_{less} < D_{more}$.

Table 1. Gene expression assays. All TaqMan Assays are obtained from Thermo Fisher Scientific. The assay marked with * is obtained from DNA Technology (Aarhus, Denmark).

Gene Name	TaqMan Assay no.	Gene Name	TaqMan Assay no.	Gene Name	TaqMan Assay no.
Classifier genes					
ADM	Hs02562698_s1	*EGLN3*	Hs00222966_m1	*P4HA1*	Hs00914594_m1
ALDOA	Hs00605108_g1	*FAM162A*	Hs01055823_m1	*P4HA2*	Hs00990001_m1
ANKRD37	Hs00699181_g1	*KCTD11*	Hs00922550_s1	*PDK1*	Hs01561840_m1
BNIP3	Hs00969293_mH	*LOX*	Hs00184700_m1	*PFKFB3*	Hs00998700_m1
BNIP3L	Hs00188949_m1	*NDRG1*	Hs00608389_m1	*SLC2A1*	Hs00892681_m1
Reference genes					
ACTR3	Hs01029161_m1	*NDFIP1*	Hs00228968_m1		
*RPL37A**	Forward primer:	TGTGGTTCCTGCATGAAGACA			
	Reverse primer:	GTGACAGCGGAAGTGGTATTGTAC			
	Probe:	TGGCTGGCGGTGCCTGGA			

Table 2. Gene expression values for 'more' and 'less' hypoxic groups as defined in [1]. Gene expression is given as log2-transformed fold expression values relative to the three reference genes.

Gene Name	Mean 'More hypoxic'	Mean 'Less hypoxic'	Estimated common variance	Gene Name	Mean 'More hypoxic'	Mean 'Less hypoxic'	Estimated common variance
ADM	-0.749	-2.345	1.400	*LOX*	-1.086	-2.431	1.924
ALDOA	-0.671	-1.552	1.463	*NDRG1*	2.362	0.830	1.978
ANKRD37	-4.157	-5.651	0.700	*P4HA1*	-4.877	-6.355	1.370
BNIP3	-0.629	-1.524	1.244	*P4HA2*	-2.712	-4.210	0.903
BNIP3L	-0.462	-1.086	0.492	*PDK1*	-1.706	-2.306	0.484
EGLN3	-0.546	-1.563	1.559	*PFKFB3*	0.461	-0.243	1.143
FAM162A	-0.639	-1.285	0.508	*SLC2A1*	1.956	0.529	1.766
KCTD11	-2.084	-3.129	1.851				

Normalization

During the development of the final classifier for head and neck cancer patients, some of the TaqMan assays used in the original study on pO_2 measurements were replaced, and the qPCR was performed on a newer instrument. Thus, to allow comparisons with the original mean and variance numbers in Table 2, the log2-transformed fold expression levels of recent head and neck cancer samples are normalized to the scale at which the original mean and variance numbers were developed. The qPCR procedures (including the specific TaqMan assays) used for cervical cancer samples in the present study are identical to the qPCR procedures currently used to classify head and neck cancer patients.

The normalization procedure has been described and validated previously [2,12], and transforms a cohort of recent patients so that the mean and standard deviation (SD) of each

gene in a recent cohort is identical to the mean and standard deviations in the original cohort using

$$y_{trf,mn} = \left(\frac{y_{new,mn} - \overline{z}_{new,m}}{\sigma_{new,m}} \times \sigma_{orig,m}\right) + \overline{z}_{orig,m} \qquad \text{(b)}$$

where $y_{new,gn}$ and $y_{trf,gn}$ are the expression values for a recent patient, n, before and after the transformation for the gene $m \in C$, where C refers to the 15 genes in the classifier (Table 1), $\overline{z}_{new,m}$ and $\sigma_{new,m}$ are the mean and SD for the recent cohort, and $\overline{z}_{orig,m}$ and $\sigma_{orig,m}$ are the mean and SD for the original cohort (Table 3).

As mentioned above, Figure 1 of the main manuscript demonstrates that the expression pattern of the 15 hypoxia induced genes in the 183 patients with locally advanced SCC cervical cancer is similar to the expression pattern in the cohort of head and neck patients used to derive the values for the normalization procedure (b) of prospective head and neck cancer patients. For the classification of the cervical samples in the present study, the mean and SD values for the 183 cervical patients are used for normalization.

Table 3. Mean and standard deviation (SD) for the original cohort used to develop the classifier. Mean gene expression is given as log2-transformed fold expression values relative to the three reference genes.

Gene Name	Mean	SD	Gene Name	Mean	SD
ADM	-2.513	1.239	*LOX*	-2.044	1.419
ALDOA	-1.283	1.074	*NDRG1*	0.553	1.410
ANKRD37	-5.138	1.079	*P4HA1*	-6.079	0.993
BNIP3	-1.494	1.364	*P4HA2*	-3.521	1.087
BNIP3L	-0.601	0.712	*PDK1*	-1.799	1.084
EGLN3	-1.468	1.451	*PFKFB3*	-0.045	0.931
FAM162A	-1.207	0.884	*SLC2A1*	0.927	1.485
KCTD11	-2.520	1.242			

Missing values

In order to minimize the problem with missing values due to low amounts of suitable mRNA in old paraffin sample (as mentioned above), part of the selection criteria during the initial development of the 15-gene classifier was that the genes should be expressed at relatively high levels [5]. Furthermore, the classification procedure allows for missing values of some of the hypoxia induced (classifier) genes (but not any of the reference genes). If a gene cannot be detected in a sample, the D_{more} and D_{less} values for that sample is simply based

on 14 rather than 15 genes. Here, an arbitrary number of 5 classifier genes are allowed to have missing values before classification of a sample is unsuccessful. Table 4 includes the number genes measured in the 190 squamous cell carcinoma samples available for analysis, with 183 samples having 10 or more genes successfully measured.

Table 4. Number of hypoxia induced genes measured successfully in 190 cervical SSC cancer samples.

Number of genes successfully measured	Number of samples
0*	4
6	1
7	1
9	1
10	2
11	5
12	7
13	7
14	22
15	140

* If one or more of the reference genes are not measured successfully, none of the 15 hypoxia genes can be measured.

6-gene classifier, Fjeldbo and Lyng, Norwegian Radium Hospital, Oslo

A 6-gene hypoxia gene classifier for cervical cancer has been developed by Christina Fjeldbo, Heidi Lyng, and co-workers at the Norwegian Radium Hospital, Oslo [14]. Two of the genes are also included in the present analysis (Table 5).

Table 5. Gene expression assays for 6-gene classifier, Norwegian Radium Hospital, Oslo. All TaqMan Assays are obtained from Thermo Fisher Scientific.

Gene Name	TaqMan Assay no.	Gene Name	TaqMan Assay no.	Gene Name	TaqMan Assay no.
Classifier genes					
DDIT3	Hs01090850_m1	*ERO1A*	Hs00205880_m1	*STC2*	Hs01063215_m1
UPK1A	Hs01086736_m1	*KCTD11**	Hs00922550_s1	*P4HA2**	Hs00990001_m1
Reference genes					
CHCHD1	Hs00415053_g1	*SRSF9*	Hs01596548_g1	*TMBIM6*	Hs00162661_m1

* Included in the 15-gene classifier used in the present study.

The reference and hypoxia induced genes were tested in a subset of 34 samples in the present study. Table 6 shows the percentage of samples with missing values for each of the genes in the two signatures.

Half of the samples were from patients diagnosed in 2005-2006 and the other half were diagnosed between 2007-2014. There was a clear correlation between age of diagnosis and missing values with the 'Aarhus' signature, thus for the 18 genes measured in the 17 patients diagnose in the early period, there was 12 missing values (out of 306 measured), whereas there were only 2 missing values (out of 306 measured) in the 17 patients from the more recent period. No clear correlation was observed for the 'Oslo' signature with 16 and 12 missing values in the early vs more recently diagnosed patients (153 measured in each group). This could indicate that the main problem with some of the genes in the 'Oslo' signature is a general low level of expression which is problematic even in relatively recent samples.

Table 6. Percentage of samples with missing values for the 15-gene 'Aarhus' signature and the 6-gene 'Oslo' signature.

Gene Name	Percent missing values	Gene Name	Percent missing values
Reference genes, 'Aarhus'		Reference genes, 'Oslo'	
ACTR3	0 %	*CHCHD1*	14 %
NDFIP1	0 %	*SRSF9*	6 %
RPL37A	0 %	*TMBIM6*	0 %
Classifier genes, 'Aarhus'		Classifier genes, 'Oslo'	
ADM	3 %	*DDIT3*	9 %
ALDOA	3 %	*ERO1A*	0 %
ANKRD37	6 %	*STC2*	23 %
BNIP3	0 %	*UPK1A*	14 %
BNIP3L	0 %		
EGLN3	3 %		
FAM162A	0 %		
KCTD11	9 %	*KCTD11*	9 %
LOX	0 %		
NDRG1	0 %		
P4HA1	9 %		
P4HA2	3 %	*P4HA2*	3 %
PDK1	6 %		
PFKFB3	0 %		
SLC2A1	0 %		

In conclusion, the percentage of missing values in both the reference and classifier genes in the 6-gene classifier was unfortunately relatively high and it was not explored further.

References

[1] Toustrup K, Sørensen BS, Nordsmark M, et al. Development of a Hypoxia Gene Expression Classifier with Predictive Impact for Hypoxic Modification of Radiotherapy in Head and Neck Cancer. Cancer Res. 2011;71(17):5923–5931. https://doi.org/10.1158/0008-5472.CAN-11-1182.

[2] Toustrup K, Sørensen BS, Metwally MAH, et al. Validation of a 15-gene hypoxia classifier in head and neck cancer for prospective use in clinical trials. Acta Oncol. 2016;55(9–10):1091–1098. https://doi.org/10.3109/0284186X.2016.1167959.

[3] Sørensen BS, Hao J, Overgaard J, et al. Influence of oxygen concentration and pH on expression of hypoxia induced genes. Radiother Oncol. 2005;76(2):187–193. https://doi.org/10.1016/j.radonc.2005.06.037.

[4] Sørensen BS, Alsner J, Overgaard J, et al. Hypoxia induced expression of endogenous markers in vitro is highly influenced by pH. Radiother Oncol. 2007;83(3)https://doi.org/10.1016/j.radonc.2007.04.028.

[5] Sørensen BS, Toustrup K, Horsman MR, et al. Identifying pH independent hypoxia induced genes in human squamous cell carcinomas in vitro. Acta Oncol. 2010;49(7):895–905. https://doi.org/10.3109/02841861003614343.

[6] Tramm T, Sørensen BS, Overgaard J, et al. Optimal Reference Genes for Normalization of qRT-PCR Data from Archival Formalin-fixed, Paraffin-embedded Breast Tumors Controlling for Tumor Cell Content and Decay of mRNA. Diagnostic Mol Pathol. 2013;22(3):181–187. https://doi.org/10.1097/PDM.0b013e318285651e.

[7] Hennig G, Gehrmann M, Stropp U, et al. Automated Extraction of DNA and RNA from a Single Formalin-Fixed Paraffin-Embedded Tissue Section for Analysis of Both Single-Nucleotide Polymorphisms and mRNA Expression. Clin Chem. 2010;56(12):1845–1853. https://doi.org/10.1373/clinchem.2010.151233.

[8] Nordsmark M, Overgaard M, Overgaard J. Pretreatment oxygenation predicts radiation response in advanced squamous cell carcinoma of the head and neck. Radiother Oncol. 1996;41(1):31–39. https://doi.org/10.1016/S0167-8140(96)91811-3.

[9] Nordsmark M, Overgaard J. A confirmatory prognostic study on oxygenation status and loco-regional control in advanced head and neck squamous cell carcinoma treated by radiation therapy. Radiother Oncol. 2000;57(1):39–43. https://doi.org/10.1016/S0167-8140(00)00223-1.

[10] Overgaard J, Sand Hansen H, Overgaard M, et al. A randomized double-blind phase III

study of nimorazole as a hypoxic radiosensitizer of primary radiotherapy in supraglottic larynx and pharynx carcinoma. Results of the Danish Head and Neck Cancer Study (DAHANCA) Protocol 5-85. Radiother Oncol. 1998;46(2):135–146. https://doi.org/10.1016/S0167-8140(97)00220-X.

[11] Toustrup K, Sørensen BS, Lassen P, et al. Gene expression classifier predicts for hypoxic modification of radiotherapy with nimorazole in squamous cell carcinomas of the head and neck. Radiother Oncol. 2012;102(1):122–129. https://doi.org/10.1016/j.radonc.2011.09.010.

[12] Tramm T, Mohammed H, Myhre S, et al. Development and Validation of a Gene Profile Predicting Benefit of Postmastectomy Radiotherapy in Patients with High-Risk Breast Cancer: A Study of Gene Expression in the DBCG82bc Cohort. Clin Cancer Res. 2014;20(20):5272–5280. https://doi.org/10.1158/1078-0432.CCR-14-0458.

[13] Bohmann K, Hennig G, Rogel U, et al. RNA Extraction from Archival Formalin-Fixed Paraffin-Embedded Tissue: A Comparison of Manual, Semiautomated, and Fully Automated Purification Methods. Clin Chem. 2009;55(9):1719–1727. https://doi.org/10.1373/clinchem.2008.122572.

[14] Fjeldbo CS, Julin CH, Lando M, et al. Integrative Analysis of DCE-MRI and Gene Expression Profiles in Construction of a Gene Classifier for Assessment of Hypoxia-Related Risk of Chemoradiotherapy Failure in Cervical Cancer. Clin Cancer Res. 2016;22(16):4067–4076. https://doi.org/10.1158/1078-0432.CCR-15-2322.