Abstract
Oral cancer is the eighth most common cancer worldwide and represents a significant disease burden. If detected at an early stage, survival from oral cancer is better than 90% at 5 years, whereas late stage disease survival is only 30%. Therefore, there is an obvious clinical utility for novel metabolic markers that help to diagnose oral cancer at an early stage and to monitor treatment response. In the current study, blood samples of oral cancer patients were analyzed using nuclear magnetic resonance spectroscopy to derive a metabolic signature for oral cancer. Using multivariate chemometric analysis, we obtained an excellent discrimination between serum samples from cancer patients and from a control group and could also discriminate between different stages of disease. The metabolic profile obtained for oral cancer is significant, even for early stage disease and relatively small tumors. This suggests a systemic metabolic response to cancer, which bears great potential for early diagnosis.

Introduction
Oral squamous cell carcinoma (OSCC) is the eighth most common cancer worldwide and represents a well-defined subgroup of head and neck cancer [1,2]. It has a 5-year mortality of 45% to 60% dependent on the setting and patient group, and the disease is frequently associated with tobacco smoking and excessive alcohol intake [3,4]. The main treatment of OSCC is radical surgery. High-stage disease frequently requires postoperative adjuvant radiotherapy to "mop-up" residual disease and is therefore more distressing for patients and, unfortunately, less successful in survival when compared with treatment of low-stage disease. Whereas patients with high-stage tumors of larger size and higher metastatic potential have only a 30% chance of survival at 5 years, patients with low-stage cancer have an 80% to 90% chance of survival.

A high degree of clinical skill is required to detect OSCC, and this is not widely available. The development of detection methods (such as blood tests) that can act as a surrogate for a highly skilled clinical examination would allow detection of OSCC at an early stage as well as monitoring response to treatment and helping to identify residual disease after treatment. This would significantly improve survival as treatment of early stage disease has been shown to be more successful.

Nuclear magnetic resonance (NMR) spectroscopy is a commonly used analytical method to analyze the small molecule composition, that is, the metabolome, of body fluids such as urine and blood serum. Variations in metabolite concentrations have been associated with the biochemical status of organisms and reflect changes in metabolism arising from biologic conditions, including disease and response to chemical treatment. Recent studies demonstrate the applicability of NMR-based metabolomics using serum samples for the diagnosis and prognosis of disease [5-13].

Both $^1$H and $^{31}$P magnetic resonance spectroscopy have previously been used to identify metabolic signatures of squamous cell carcinoma compared with benign tumors and normal tissues [14-21]. Both in vivo and in vitro clinical studies have shown that the choline-creatinine ratio is significantly higher in OSCC than in normal tissue.
Materials and Methods

Patient Characteristics and Sample Collection

After ethical committee approval and written patient consent, samples of venous blood were collected from sequential patients with histologically confirmed OSCC at the University Hospital Birmingham NHS Foundation Trust, Birmingham, United Kingdom. Control samples were taken from normal donors who had no personal history of cancer. All samples were taken at around the same time during the day for each patient.

Samples were harvested by standard venepuncture, and 10 ml of blood was acquired and immediately transferred to a sterile universal container without anticoagulant. This was centrifuged at 3500 rpm for 5 minutes. The supernatant was removed in 1-ml aliquots, and samples were immediately quenched in liquid nitrogen until serum was frozen and later transferred to a −80°C freezer for storage.

Samples were harvested from 15 cancer patients: 5 males and 10 females of average age of 63.8 years. Patient characteristics, staging of disease, and other parameters are shown in Table 1. The UK average age for patients with oral cancer is 63.5 years, which is well represented by this study. There is an unexpectedly high proportion of females in this cohort; the national average is a ratio of 2:1 in favor of males. No significance is attached to the high proportion of females in this cohort, as it is within the range of the national average. However, it is worth noting that women with cancer have a longer life expectancy than men, which could explain the higher proportion of females in this study.

Ten blood samples were harvested from normal healthy donors: six males and four females of average age of 57 years (Table 1). Tumors were staged according to the American Joint Committee on Cancer classification for head and neck and oral cancer [25].

Sample Preparation

Approximately 0.5 ml of human serum was filtered (Nanosep 3K OMEGA; Pall Corporation, Ann Arbor, MI) at 4°C using a centrifuge at 10,000 rpm [26]. Three hundred twenty microliters of filtered serum was diluted adding 90 μl of distilled water. One hundred forty microliters of phosphate buffer 0.4 M containing 0.75% w/v sodium azide and 50 μl of 3.25 mM 3-(trimethylsilyl)propionate-2,2,3,3-d4 (TMSP; Cambridge Isotope Laboratories) in D2O (99.9% pure; Goss Scientific Instruments Ltd, Essex, UK) were added to the filtered serum. The final pH of the solution was adjusted to 7.0 ± 0.1. The sample transferred to the NMR tube contained 600 μl of filtered serum.

Nuclear Magnetic Resonance Data Acquisition

One-dimensional (1D) 1H and 2D 1H J-resolved (JRES) NMR spectra [27] were acquired using a 500-MHz Bruker spectrometer (Bruker Biospin, Rheinstetten, Germany) equipped with a cryogenically cooled probe at 303.15 K. Proton spectra were referenced to the TMSP signal (δ 0.00 ppm). In both pulse sequences, the water resonance was suppressed using excitation sculpting [28]. One-dimensional spectra were acquired using a 60° flip angle, a 5-kHz spectral width, and a relaxation delay of 3 seconds; 128 transients were collected using 16,384 data points and eight dummy scans. Two-dimensional JRES spectra were collected using a double spin-echo sequence with 16 transients per increment at 32 increments. Strong coupling artifacts were suppressed by phase cycling [29]. Two-dimensional COSY experiments were carried out using an 800-MHz Varian spectrometer (Varian, Palo Alto, CA) equipped with a cryogenically cooled probe using a gradient-selected coherence transfer pathway (gCOSY45 [30]) with 16 transients of 8192 complex data points, 256 increments, and a spectral width of 8 kHz in both dimensions.

Nuclear Magnetic Resonance Data Processing and Postprocessing

Nuclear magnetic resonance data were processed using NMRLab [31] and MetaboLab (U. Günther, unpublished software) in the
MATLAB (The MathWorks, Inc., Natick, MA) programming environment. Postprocessing included scaling, alignment, exclusion of water and TMSP signals, binning at 0.005 ppm, and application of a generalized log transformation [32] to enhance small signals in the spectrum.

Before Fourier transformation, 2D JRES were multiplied by a combined sine-bell/exponential window function in the direct dimension and by a sine bell function in the incremented dimension [33]. The spectra were processed and tilted in frequency space by 45°; projections were calculated after tilting and symmetrizing using NMRLab [31] to simplify 1D spectra and to reduce peak congestion. Alternatively, we also recovered coupled 1D spectra by omitting the 45° rotation for metabolite identification using the Chenomx NMR Suite software (version 5; Chenomx Inc., Edmonton, Canada).

**Statistical Analysis**

Unsupervised (principal component analysis, PCA) and supervised (PLS-DA) multivariate analyses in PLS-Toolbox (Version 4.1; Eigenvector Research, Manson, WA) were used for statistical data analysis. Cross-validation applying the “Venetian blind” algorithm has been used for all PLS-DA analyses. Metabolites were identified from loadings plots using the Chenomx NMR Suite software (version 5; Chenomx Inc.).

**Results**

Figure 1 depicts the superimposition of two representative 1H NMR spectra (projections of JRES spectra), which show the average signals from healthy (blue line) and diseased (red line) subjects. The spectra are dominated by the resonances of carbohydrates, in particular, both anomeric forms of glucose (≈3–5.5 ppm, Gluc) and some of the intermediate metabolites of the glycolytic pathway, such as lactate (δ 1.33 and 4.11 ppm, Lac). Other compounds such as amino acids, in particular, valine (Val, 0.9–1.1 ppm) and alanines (Ala, 1.46 ppm), show large methyl signals in the spectrum. Smaller contributions to the spectrum are often more relevant to the analysis and are enhanced by applying a generalized log transformation (Figure 1) [32].

One-dimensional and 2D JRES NMR spectra were recorded for filtered blood sera from patients with head and neck cancers and from a control group of healthy volunteers. Two-dimensional JRES spectra were used to visualize both chemical shifts and scalar couplings along different spectral dimensions and increase peak dispersion and therefore metabolite specificity in 1D projections [34]. Principal component analysis applied to 1D projections is depicted in Figure 2 and shows a clear separation between samples from healthy subjects and OSCC patients. Samples from patients with the disease group geared toward smaller or negative values of PC2; samples from higher stage disease also spread out toward higher positive values of PC1. Furthermore, the PCA result shows clustering according to the stage of disease; samples B06, B07, B08, B09, and B12 are from patients with stages III and IV disease, with associated metastasis, and are observed at higher absolute values of PC1 and PC2. Samples B01 and B15 represent an exception to good separation of low- and high-stage disease, which, in the case of B01, can be attributed to the small size of the tumor that was classified as high-stage IV because it had invaded the locally adjacent bony structures although there was no indication of metastatic disease. Therefore, PCA not only differentiates disease and control samples but also shows significant potential to distinguish early and late stage disease with high specificity.

![Figure 1. Superimposition of two representative 1H NMR spectra (projections of JRES spectra), which show the average signals from the serum samples of healthy subjects (blue line) and patients with disease (red line) after application of a generalized log transformation. Predominant metabolite resonances are labeled on the spectrum. Ala indicates alanine; Gluc, glucose; Lac, lactate; Val, valine.](image-url)
To obtain a more objective statistical estimation and specific loadings, we used PLS-DA for a model discriminating samples from cancer patients and controls (Figure 3). In this case, the sensitivity and specificity for oral cancer detection are both >95% (see receiver operating characteristic curves in Figure W1) after applying the Venetian blind algorithm for cross-validation. Despite limited statistical significance owing to a small number of samples, this result is remarkable considering that some early stage tumors were below 0.005% of the overall body mass.

To probe for a possible bias arising from other factors, we probed the ensemble of data for the influence of age and sex. Both factors do not cluster in the unsupervised PCA. Therefore, PLS-DA analyses were used to build models for disease, age, and sex. In this analysis, a substantially less-pronounced clustering and a lower degree of specificity and sensitivity were observed for age and sex compared with disease versus control (see receiver operating characteristic curves in Figures W1–W3).

Loadings plots were calculated from PLS-DA models to identify discriminatory metabolites for different models. Figure 4, A and B, shows different areas of loadings in the first latent variable of PLS-DA, with positive values representing healthy controls and negative values representing patients with disease. In the samples from cancer patients, the levels of valine, ethanol, lactate, alanine, acetate, citrate, phenylalanine, tyrosine, methanol, formaldehyde, and formic acid were reduced compared with those of healthy controls, whereas signals arising from glucose, pyruvate, acetone, acetoacetate, 3-hydroxybutyrate and 2-hydroxybutyrate, choline, betaine, and, to a lesser degree, dimethylglycine, betaine, and choline showed enhanced loadings. In the aromatic region, additional contributions arise from yet unidentified metabolites. These contributions were summarized in metabolic pathways shown in Figure 5. To avoid possible errors arising from PLS-DA for a small sample size, we compared these results with differences of average late and early stage spectra, which gave the same results.

To identify the metabolites that are responsible for the distinction between different stages of disease, PLS-DA models using only samples from patients with disease were built (not shown). Loadings plots from latent variable 1 are displayed in Figure 6, A and B, showing similar patterns as for the distinction between patients with disease and healthy controls (Figure 4, A and B). In samples from patients with late stage disease, the levels of 2-hydroxybutyrate, 3-hydroxybutyrate, acetone, acetate, acetoacetate, creatinine, asparagine, glucose, dimethylglycine, betaine, and choline were remarkably increased compared with patients with early stage disease, whereas valine, lactate, alanine, pyruvate, lysine, creatine, acetyl-L-carnitine, and carnitine showed reduced concentrations (Figure 6, A and B). In addition, a model using three classes (healthy controls and patients with early and late stage diseases) is shown in Figure W4. The application of this multivariate chemometric model shows a clear separation between the samples from healthy controls and patients with disease. Very narrow groupings are observed for the control samples and the patients with early stage disease. On the contrary, the sera of patients with late stage disease are more widely spread in the scores plot highlighting the higher inhomogeneity of their metabolic profile.

Discussion

This study shows an unexpectedly strong metabolic profile in blood sera of patients with OSCC with a significant signature even for low-stage tumors. Subsequent analysis of metabolite profiles of serum samples from oral cancer patients can distinguish patients with disease from healthy normal controls and can also discriminate patients with high- and low-stage diseases and provide a fingerprint of metabolic changes. Although the statistical relevance of this study
is limited by the small number of samples, a strong signature was obtained using an unbiased and unsupervised PCA. This is an astonishing result considering the relatively small size of the tumors (less than 0.005% of total body mass for early stages) and the fact that some patients (B02, B03, B04, B05, B10, B11, B13, and B14) had localized tumors without metastasis or only small locoregional metastasis. Sensitivity and specificity were both remarkably high for the detection of disease. Discrimination between disease stages in PCA was also remarkable with only two samples from patients with stages IV and III disease (B01 and B15) clustering with low-stage

![Figure 3](image3.png)

**Figure 3.** PLS-DA scores plot of $^1$H NMR spectra of human blood sera using two classes: OSCC patients (● early and □ late stage diseases) and healthy controls (▼).

![Figure 4](image4.png)

**Figure 4.** Loadings plots for the first principal component for different regions of the spectrum. A representative NMR spectrum of human serum from an OSCC patient is shown above the loadings plots. Signals representing the most relevant discriminatory metabolites of healthy individuals (positive loadings) and OSCC patients (negative loadings) are labeled (A and B). 2HB indicates 2-hydroxybutyrate; 3HB, 3-hydroxybutyrate; Ac, acetone; Ace, acetate; AceAce, acetooxacetate; ACL, acetyl- l-carnitine; Ala, alanine; Asn, asparagine; Bet, betaine; Car, carnitine; Cho, choline; Cit, citrate; Cr, creatine; Cre, creatinine; DMG, dimethylglycine; EtOH, ethanol; Gln, glutamine; Gluc, glucose; Gly, glycine; Lac, lactate; Lys, lysine; MeOH, methanol; Om, ornithine; Pro, proline; Pyr, pyruvate; Sar, sarcosine; Ser, serine; Val, valine.
samples. For B01, this has been attributed to the fact that the tumor was small but had invaded the locally adjacent bony structures, although there was no indication of metastatic disease. Based solely on size and in absence of the bony invasion, this tumor would have been classified as stage I or II. Many believe such upstaging to be incorrect, and certainly, this tumor was biologically more similar to low-stage disease and may explain why it clustered with the other low-stage tumors.

Because blood serum represents the effects of metabolism in different organs, it is difficult to assign a metabolic fingerprint to specific metabolic processes. Nevertheless, changes in blood serum metabolite

Figure 5. Schematic representation of the most relevant metabolic differences between oral cancer patients and healthy controls.

Figure 6. Loadings plots (latent variables) from a PLS-DA analysis of OSCC patients showing metabolites that discriminate between early (positive loadings) and late stage diseases (negative loadings). A representative NMR spectrum of human serum from an OSCC patient is shown above the loadings plots.
concentrations of patients with OSCC clearly point to an altered energy metabolism. The relatively high level of glucose and low level of lactate is strikingly different to observations in other types of cancer where, commonly, lactate levels are high [10,35]. It is unlikely that the high level of glucose can be ascribed to any other disorder because none of the patients was diagnosed with any other medical condition leading to elevated glucose. It is possible that the levels of glucose may be linked to the unique behavior of oral cancer, which interferes with the ability of insulin to modulate the uptake of glucose and thus regulates consequent energy metabolism favoring the accumulation of carbohydrates and the process of ketogenesis [36–39]. Several potential precursors of glucose in gluconeogenesis, such as lactate, alanine, and other gluconeogenic amino acids [40–42], were found at lower concentrations in patients with head and neck cancer. This clearly points to an altered energy metabolism, although it is not possible to assign changes to specific pathways considering that blood samples represent a cross-section of metabolic events in multiple organs.

The increased level of ketone bodies can be linked to lipolysis as a backup mechanism for energy production. This signature cannot be a consequence of fasting because all patients with OSCC received optimal nutrition according to a standard protocol of calorific requirement. Enteral tube feeding was used if patients were unable to swallow. Moreover, of the 15 patients, only 2 had lost significant amounts of weight (B03 and B06) and 2 others showed moderate weight loss (B01 and B08). Importantly, samples from these patients were grouped according to the disease stage with the already discussed exception of B01. This shows that the cancer signature dominates in these samples.

Furthermore, several Krebs cycle intermediates such as citrate, succinate, and fumarate were less concentrated in patients with disease suggesting a suppressed Krebs cycle. This represents a typical signature in cancer patients, and it has been previously associated with the “Warburg effect” [43], which assumes that tumors rely on glycolysis as a main source of energy, even in the presence of oxygen.

Elevated levels of 2-hydroxybutyrate may be associated with protein and amino acid catabolism. This is also supported by vastly altered amino acid profiles of head and neck cancer patients. For example, ornithine and asparagine levels were higher in oral cancer patients than in control samples. This result is in agreement with previous observations where blood plasma free amino acid pattern profiles were diagnostically correlated with the organ site for three different types of malignant tumors [44]. Ornithine represents a metabolite found in the urea cycle, which can use amino acid breakdown products to feed into the Krebs cycle. Besides ornithine, the blood levels of most essential and non-essential amino acids (such as threonine, valine, isoleucine, leucine, glycine, tyrosine, and phenylalanine) were found to be lower in the cancer patients. A significant decrease in blood plasma threonine and glycine has previously been reported at the early stage of head and neck cancer [44,45] but has been attributed to dysphagia and consequent insufficient food intake rather than to the disease itself. However, this is an unlikely cause in the present study because optimal nutrition was ensured for all patients.

In agreement with previous reports analyzing OSCC tumor samples [16,18,19], serum samples showed creatine to be low, creatinine to be high, and choline-creatine ratios to be high. Choline and its derivatives represent important constituents in phospholipid metabolism of cell membranes [46] and have been previously identified as markers of cellular proliferation. In the context of the metabolite profile associated with OSCC in this study, choline, together with betaine, dimethylglycine, carnitine, and acetyl carnitine, allows a distinct differentiation between early and late stage diseases.

In summary, patients with OSCC show a distinct signature of altered energy metabolism in blood serum, which includes altered lipolysis (an accumulation of ketone bodies), a distorted Krebs cycle, and amino acid catabolism. Most of the observed effects are more pronounced in late stage OSCC, and some metabolites show specific changes in the later stages of disease. Such measurements might revolutionize cancer treatment if early stage disease can be detected equally well in other cancers and if the obtained signatures show specificity for the individual type of disease.

References


Figure W1. Receiver operating characteristic curves of sensitivity versus specificity (the plots on the left) and threshold versus specificity/sensitivity (the plots on the right) calculated for cross-validated PLS-DA applied to NMR spectra of serum samples from healthy controls (class 1) and patients affected by oral cancer (class 2). A high degree of specificity and sensitivity is observed.
Figure W2. Receiver operating characteristic curves of sensitivity versus specificity (the plots on the left) and threshold versus specificity/sensitivity (the plots on the right) calculated for cross-validated PLS-DA applied to NMR spectra from healthy controls below (class 1) and above (class 2) the mean age. Compared with Figure W1, a lower degree of specificity and sensitivity is observed.
Figure W3. Receiver operating characteristic curves of sensitivity versus specificity (the plots on the left) and threshold versus specificity/sensitivity (the plots on the right) calculated for PLS-DA applied to NMR spectra with groups for sex (male: class 1 and female: class 2). Compared with Figure W1, a lower degree of specificity and sensitivity is observed.
Figure W4. PLS-DA scores plot of $^1$H NMR spectra of human blood sera using three different classes for patients with ● early and ■ late stage diseases and healthy controls (▼). The application of this multivariate chemometric model shows a clear separation between the samples from healthy controls and the patients with disease. Very narrow groupings are observed for the control samples and the patients with early stage disease. On the contrary, the sera of patients with late stage disease are more widely spread in the scores plot, highlighting the higher inhomogeneity of their metabolic profile.