Tongue cancer patients can be distinguished from healthy controls by specific N-glycopeptides found in serum.

Mayank Saraswat[1, 2, *], Antti Mäkitie[3, 4], Tiialotta Tohmola[1, 5], Amy Dickinson[3], Shruti Saraswat[1], Sakari Joenväärä[1,2] and Suvi Renkonen[3,6]

[1]Transplantation Laboratory, University of Helsinki, Haartmaninkatu 3, PO Box 21, Helsinki 00014, Finland;

[2]HUSLAB, Helsinki University Hospital, Helsinki 00290, Finland;

[3]Department of Otorhinolaryngology – Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki 00130, Finland;

[4]Division of Ear, Nose and Throat Diseases, Department of Clinical Sciences, Intervention and Technology, Karolinska Institutet and Karolinska Hospital, Stockholm, Sweden

[5]Department of Biosciences, University of Helsinki, PO Box 65, Helsinki 00014, Finland

[6]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm 11382, Sweden

* Corresponding author.

Transplantation Laboratory, University of Helsinki & HUSLAB, Helsinki University

Hospital, Helsinki, Finland.

Mail to: sarawat.mk@gmail.com

Phone number: +358-449572310

Fax Number: +358-294126700

Running title: N-glycoproteomics of tongue cancer serum

Abbreviations:FDR: False discovery rate, HNSCCs: Head and neck squamous cell carcinomas, OSCC: Oral Squamous Cell Carcinoma.

Keywords: OSCC, HNSCC, N-glycoproteomics, IgG, Transferrin

Total words: 5,581

Statement of Clinical Relevance

Diagnosis of Oral squamous cell carcinoma (OSCC) is based on histology of a visible tumour. By that time, it has become late and the tumour has become advanced grade and treatment outcomes become poor. There are no biomarkers for early detection of OSCC tumours. We have identified a set of glycoforms of serum proteins (IgG1, IgG4, Haptoglobin and Transferrin) which can separate the OSCC patients from healthy controls. The peptide sequence, N-glycosylation site, glycan composition and proposed glycans structures of these glycoforms were established. Serum samples belonged to either stage I or IVA OSCC which were compared to healthy controls and it was possible to separate these patients from healthy controls based of the expression levels of these glycoforms. The study identifies N-glycoforms of several proteins which can help design sensitive future modalities for the diagnosis of OSCC patients as early as Stage I tumour. These glycoforms can potentially serve as sensitive biomarkers of the OSCC.

Abstract

Purpose

There are no blood biomarkers to detect early-stage oral cavity squamous cell carcinoma (OSCC) prior to clinical signs. Most OSCC incidence is associated with significant morbidity and poor survival. We aimed to use mass-spectrometry (MS) technology to find specific N-glycopeptides potentially serving as serum biomarkers for preclinical OSCC screening.

Experimental Design

Serum samples from 14 patients treated for OSCC (stage I or stage IV) with 12 age and sex-matched controls were collected. Quantitative label-free N-glycoproteomics was performed, with MS/MS analysis of the statistically significantly different N-glycopeptides.

Results

Combined with a database search using Web-based software (GlycopeptideID), MS/MS provided detailed N-glycopeptide information, including glycosylation site, glycan composition and proposed structures. 38 tryptic N-glycopeptides were identified, having 19 unique N-glycosylation sites representing 14 glycoproteins. OSCC patients, including stage I tumors, could be differentiated from healthy controls based on the expression levels of these glycoforms. N-glycopeptides of IgG1, IgG4, Haptoglobin and Transferrin had statistically significantly different abundances between cases and controls.

Conclusions and clinical relevance

We are the first to suggest specific N-glycopeptides to serve as potential serum biomarkers to detect preclinical OSCC in patients. These N-glycopeptides are lead candidates for validation as future diagnostic modalities of OSCC as early as Stage I.

1. Introduction

Head and neck squamous cell carcinomas (HNSCCs) encompass a heterogeneous group of tumours. Smoking and alcohol consumption are the most common aetiological risk factors [1]. The outcome of HNSCC patients remains poor and this is largely due to the fact that most of the tumours are diagnosed at an advanced stage [2]. Although the incidence of many HNSCCs has decreased during the last decades, oral cavity cancers (OSCCs), and especially tongue cancers, have shown an increasing incidence [3]. At the moment, the only and ultimate diagnostic tool for these cancers is a tissue biopsy from a clinically visible tumour. For early diagnosis and consequent better prognosis, screening of high-risk individuals or patients with ambiguous symptoms without clinical findings would be of great interest. Further, finding a specific diagnostic biomarker would make this possible.

Up to 75% of serum proteins are thought to be glycosylated [4]. Glycosylation is based on non-template biosynthesis, unlike for example RNA and protein coding [5]. Under physiological conditions this non-linear, branching synthesis is regulated by several factors such as availability of monosaccharides, depending on the nutritional situation, and activity of the glycan-attaching and -processing enzymes [6]. Despite the synthesis of great variety of glycoprotein types and magnitudes taking place under normal conditions, their distributions are thought to be stable [6]. Under pathological conditions, aberrant glycosylation can be caused by different occupancies of glycosylation sites on a given protein, or by the variability of attached glycan structures [6, 7], leading to a specific expression of different glycovariants. In tumours, activity or expression levels of different glycosyltransferases or glycosidases change, leading to molecular changes such as modified glycan branching, sialylation or fucosylation [8]. Another mechanism leading to altered glycosylation of proteins in malignancies is caused by the immune response of the host system to the ongoing malignant process, and thus not surprisingly, altered glycosylation patterns. Nonspecific IgGs and acute phase proteins have been shown to be present in the serum of patients with malignancies [6, 9]. Thus, the problem of low diagnostic accuracy of protein biomarkers might be overcome by choosing a specific glycoprotein instead. Due to the challenges in both measuring and interpreting the presence and magnitude of glycoproteins and their altered forms, most of the glycoprotein biomarkers in use are based on the protein backbone - instead of the glycan moiety or glycopeptides themselves [6, 9]. PSA and CEA are examples of the oldest and most often used glycoprotein biomarkers where only the protein backbone is measured in the patient serum [10].

Only a few clinical studies have previously addressed the topic of glycoproteins in HNSCCs before (all refs). In all these studies, either the amino acid sequence of the glycoprotein, or the sole glycan structure have been studied [11]. Recent advances in quantitative high-throughput mass spectrometry (MS) have enabled the analyses of glycoproteins, providing simultaneous information of amino acids and glycan compositions, in an unknown mixture. This kind of analysis compiles information about multiple glycosylation sites, as well as different glycoforms of N-glycopeptides, and allows the estimation of putative structures of the original glycoprotein.

Our aim was to analyse the serum samples of OSCC patients along with their age matched healthy controls. The aims were to separate the two groups based on their glycoprotein profiles, and to find specific combinations of glycopeptide attributes to describe this classification.

2. Materials and Methods

2.1 Sample collection, handling and study design

Fourteen serum samples were collected from patients with OSCC treated at the Department of Otorhinolaryngology – Head and Neck Surgery, Helsinki University Hospital (Helsinki, Finland). An approval for the study was received from the Ethics

Committee of Medical Sciences (Dnro: 64/13/03/02/2014) and informed consent was obtained from all subjects at the time of serum sample collection. Serum samples were stored at -70 °C until all were tested at the same time. All experiments were performed in accordance with relevant guidelines and regulations. Twelve age and sex-matched serum samples from healthy individuals serves as controls (Blood Bank, Helsinki). The study is case-control comparison type and retrospective in nature. The number of samples were chosen keeping in mind not to keep the sample size too small but practically manageable at the same time. Practicality involved the total number of patients coming to the specified clinic and meeting the requirements to be included in the study in a reasonable time frame. The workflow, from sample preparation to analysis, is represented in Figure 1. The immobilized AffiSep-SNA adsorbent was purchased from Galab Technologies (Geesthacht, Germany). All other lectins were purchased from VectorLabs (Burlingame, CA). Pierce SwellGel Blue Albumin Removal Discs, Pierce Centrifuge Columns (800 µL), and Pierce C18 Spin Columns were purchased from Thermo Scientific (Rockford, IL, USA). The high-purity HPLC reagents, and solvents were purchased from Waters (Milford, MA, USA). All other reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.2 Serum depletion and protein digestion

Patient and control serum samples were thawed on ice and centrifuged at +4 C, 400 g, 5 min to pellet precipitated proteins. The supernatant was collected to be further processed. Excess albumin was depleted from 50 µL of each sample by manufacturer's protocol using Pierce SwellGel Blue Albumin removal discs. The remaining protein concentration was determined by the Pierce BCA procedure. Samples were dried in Savant SPD121P Speed Vac centrifugal evaporator. The pellets, corresponding to 350 µg total protein, were dissolved in 35 µL of 6 M urea, 100 mM Tris-HCl pH 7.4, and reduced by adding DTT to a concentration of 10 mM, following incubation at room temperature (RT) for 60 min. Alkylation was performed by adding iodoacetamide to a concentration of 40 mM, and incubated in the dark at RT for 60 min. Excess iodoacetamide was consumed by adding additional DTT to a concentration of 30 mM and incubating at RT for 60 min. The samples were diluted 1:10 with MQ water and trypsin was added in a mass ratio of 1:50 (trypsin to protein). The digestion was carried out at +37 C overnight.

2.3 Lectin affinity chromatography

Sixty µL of tryptic peptides were diluted by 540 µL of 10mM HEPES buffer pH 7.4 containing 1mM CaCl2 and 1mM MnCl2. The mixture was applied to lectin-agarose mix column slurry containing Con-A: SNA: LCA: AAL in ratio of 5:3:3:1 for a final volume of lectin resin slurry of 150 µL. Micro-columns with lectin mix resins were incubated at 4°C on rotation overnight. Next day, 3 washes with HEPES buffer were performed and N-glycopeptides were eluted with sugar-mix solution containing fucose (100mM), α-methyl mannoside (200mM), α-methyl glucoside (200mM) and lactose (400mM) followed by second elution with 1% formic acid. N-glycopeptides were cleaned by C18 micro-columns according to manufacturer's instructions. Resulting N-glycopeptides were dissolved in 0.1% formic acid before being analyzed by UPLC-MS.

2.4 Ultra-performance liquid chromatography-mass spectrometry

In glycopeptide analysis, MS processing is done in two phases: first the peptides are quantitated by MS^E run and then second MS/MS run is performed on those proteins with altered expression levels between the compared groups. This MS/MS run is done to identify the peptide sequence and potential glycan compositions. In our study, a web-based tool called GlycopeptideID was used to identify N-glycopeptides.

The ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) system used was a Waters SYNAPT G2 High Definition MS connected to a Waters nanoACQUITY UPLC. The MS was run in positive mode with nano flow and a 1 second scan time. Calibration was performed using sodium formate (over 50-2500 m/z). The trapping column was a nanoACQUITY UPLC Trap, 180 µm × 20 mm (5 µm), SymmetryR C18, and the analytical column was a nanoACQUITY UPLC, 75 µm × 100 mm (1.8 µm), HSS T3. One microliter of each sample was injected. The LC-MS was operated in sensitivity mode with mass range 100–2000 m/z. The LC-MS/MS was operated in sensitivity mode with mass range 50–2500 m/z The CID collision energy ramp was 20–60 V. The UPLC was operated under the following conditions: Solvent A: H2O + 0.1% formic acid, Solvent B: ACN + 0.1% formic acid. Flow rate 300 nL/min. Gradient: 0–1 min 1% B, 1–2 min 5% B, 2–45 min 30% B, 45–48 min 50% B, 48– 50 min 85% B, 50–53 min 85% B, 53–54 min 1% B, 54–60 min 1% B (Curve: 6). Total run-time was 60 min.

2.5 Glycopeptide identification

The MS/MS spectra were deconvoluted in Waters MassLynx 4.1 software using the MaxEnt3 module and saved as peak lists (file format .pkl). Identification of detected glycopeptides was performed on the web-based software GlycopeptideID, developed for automated CID MS/MS spectrum analysis (http://glycopeptideid.appliednumerics.fi). Although the software is revised from earlier versions, the principle of this method is explained in detail in two publications by S. Joenväärä a et al. and H. Peltoniemi et al. [12]. Search functions of this software were also utilized in another study and described in details [13]. Briefly, a database of tryptic peptides from known human serum proteins is generated, and the combined and deconvoluted MS/MS spectra are imported as peak lists. First, each spectrum is searched against the tryptic peptide database (Uniprot, release 04_2013, 191,896 sequences, 3 misscleavages allowed, modifications were C alkylation fixed and M oxidation variable, sequence must contain NXS/T/C, X≠P) for amino acid sequence determination, and different possibilities are scored (peptide score). Next, possible glycan compositions are searched against a glycan database and resulting glycopeptides are fitted into the spectrum (glycan score). The total score for a glycopeptide hit is the sum of the peptide and glycan scores. The results are ranked, and for each possible result, an annotation spectrum is drawn for visual assessment of the matching y and b ions from peptide, glycan fragments and glycopeptide fragments. False-discovery rate was determined by searching the spectra against a reversed peptide database (target-decoy search). The false-discovery rate estimation is based on matches with the reverse database generated by reversed tryptic peptide sequences. In the resulting data, glycan compositions are given as one-letter abbreviations H: Hexose, N: Hexosamine, S: Sialic acid, and F: Fucose, with the number following showing the amount of the monosaccharides. For example, S1H5N4 stands for a glycan containing one sialic acid (Nacetylneuraminic acid), five hexoses (mannose and galactose), and four hexosamines (N-acetylglucosamine). Proteins are given as the UniProt entry name, e.g. HPT_HUMAN stands for human haptoglobin. For simplicity, we refer to identify proteins using the first part of the entry name in the results. These abbreviations are used throughout the text. Our final dataset (Supplementary Table 2B) consisted of 1556 .pkl-files (1 MS/MS spectra per file) collected from several separate MS/MS runs. All the annotated spectra for identified N-glycopeptides are provided in supplementary information. Settings used for searching deconvoluted peak list with GlycopeptideId were dm/z 0.05 Da for precursor and fragments and 50 ppm for precursors and 20 ppm for fragments.

2.6 Data accessibility: All the relevant data is available in the supplementary information section.

3. Results

Fourteen serum samples of patients with OSCC and 12 age and sex-matched samples of healthy controls were included in this study. All the patients had tongue cancers and they were treated with curative intent except one patient. Samples were collected in the operation theatre, immediately before the start of the operation for resecting the cancer. A Head & Neck pathologist confirmed the diagnosis for each patient preoperatively using a tissue biopsy. None of the patients received any treatment before surgery. Patient details are given in Supplementary Table 1.

In the first MS run 5,212 potential N-glycopeptide spectra were observed and quantified out of which 4,094 had altered expression levels among cases and controls. The quantification was based on 78 runs (14 OSCC cases and 12 controls, each run in triplicate). After lectin affinity chromatography and the first MS run recognizing the potential N-glycopeptides (4,094 ions, Supplementary Table 2A) with altered expression levels, the second MS/MS run led to identification of 38 tryptic N-glycopeptides (FDR 1.98%) with significantly altered serum expression levels between OSCC patients and controls. Figure 1 shows the principal component analysis (PCA) of OSCC patients and controls.

3.1 Identified proteins

When the deconvoluted spectra of the MS2 fragmented glycopeptides were searched against the SwissProt human tryptic peptide database (described in methods) using Glycopeptide ID search engine, containing the information about N-glycosylation consensus sites, 38 different glycoforms from 19 N-glycosylation sites and 14 glycoproteins were identified (Supplementary Table 2B). Two conditions had to be met in order to the result to be considered as reliable: at least 25% of total y and b peptide fragment ions were found and at least three of the peptide fragment ions were sequential.

3.2 Glycan compositions

MS/MS results in fragment peaks, from which the monosaccharide compositions can be identified. An assessment of the structure for a glycan can also usually be drawn. A representative annotated spectra form the study is shown in Figure 2. Annotated spectra, from the output of GlycopeptideId software, of all identified N-glycopeptides are given in the supplemental information file named “Supplemental Annotated Spectra”.

Most of the acquired glycan compositions were suggested to be bi-antennary complex-type structures, no high-mannose or hybrid structures were seen. Also, tri- and tetra-antennary structures, with possible numerous sialylation sites, were absent. Sixty-seven percent of all identified glycopeptides were fucosylated, 53% core-fucosylated, and 30% sialylated. Core-fucosylated glycoforms were mostly over-expressed (81%, fold changes ranged from 1.13-44.25), and sialylated forms mostly under-expressed (75%, fold changes ranged from -1.22- -5.90), in OSCC patients’ sera compared with that of healthy controls.

3.3 Results of relative quantification

Thirty-eight quantitated and identified glycoforms were among the ones which were found to differ between the serum samples of OSCC patients and healthy individuals (Supplementary Table 2).

3.3.1 IgGs

Immunoglobulin subspecies IgG1 was the major source of the identified peaks with 10 ions representing seven unique glycoforms (Supplementary Table 2). When comparing the expression levels of these in cancer and control samples, the fold changes were found to vary from 9.5 to 15 being higher in cancer samples. All glycoforms were core-fucosylated, and core-fucosylation was found in at least five separate glycan fragments in the CID spectra. Among the 10 identified N-glycopeptides of IgG1, Asp180 of IgG1 was found to be decorated with complex N-glycans, out of which all were fucosylated except one glycopeptide which was sialylated (m/z 1106.518, Supplementary Table 2B). Figure 3 presents proposed structures associates with different m/z values of the N-glycopeptides ions having same peptide sequence (IgG1). Microheterogeneity can be seen at specified glycosylation site in the figure with at least four unique glycoforms being present. IgG4 (Asp177) had a core-fucosylated complex glycoform and N-glycopeptide was present in 2 charge states. A core-fucosylated glycoform of IgG4 (z=4) was found to be expressed in 2.6 fold higher levels in cancer samples when compared with controls. We did not observe any significantly (FC >2) altered expression levels of glycosylated IgG2. Figure 4 shows the box plots of IgG1 glycopeptides expression levels in all samples studied. Box plot of IgG4 glycoforms expression is shown in Supplementary Figure 1.

3.3.2 Transferrin and Haptoglobin

The acute phase proteins transferrin and haptoglobin were among the glycoproteins with significantly altered glycoforms, when comparing OSCC serum to that of healthy controls. Figure 5 shows the box plot of expression levels of all glycoforms of haptoglobin. Supplementary Figure 2 presents the box plot of expression levels of 2 glycoforms of transferrin. Of the haptoglobin glycoforms, all of the glycosylation site

241, biantennal complex type with sialylation (S2H5N4) was found to have 10.9 times higher level in cancer samples than in controls. Significant fold change (>2) was not observed in the non-sialylated glycoform. In the case of Transferrin, there were three glycoforms of two different N-sites (432 and 630), all being sialylated and two also fucosylated (S2H5N4, S1H5N4F2 and S2H11N4F2), and with significantly lower expression (fold chages -2,4, -3,1 and -5,9) in cancer samples.

4. Discussion

Specific glycoforms of certain N-glycoproteins seem to be able to distinguish OSCC patients from healthy controls. In the current study, among the glycoproteins with different expression levels of the particular N-glycopeptids, sialylated structures were interestingly found to be mostly decreased in the OSCC samples, when comparing with healthy controls' sera. Core-fucosylation was clearly more frequently present and fucosylated glycoforms were significantly overexpressed in OSCC compared with the controls.

Examples of typical changes taking place in cancer glycoproteins are truncation and branching, sialylation and fucosylation. In the previous literature, a general increase in sialylation due to changes in specific glycosyltransferase expression has been linked with malignancies [14]. Increased sialylation has been reported to be associated with poor prognosis in gastric, and colorectal carcinomas [15]. Our finding of sialylated glycoforms being expressed in lower levels in OSCC is a novel finding of interest and shows that changes in glycosylation really are disease specific.

Core-fucosylation is a known regulating factor of glycoproteins' function and has been shown to take place in different cancers, including hepatocellular, breast and lung cancers [16]. Hepatocellular carcinoma can be distinguished from hepatic infection and inflammation by a core-fucosylated alfa fetoprotein expression in serum

[17]. Fucosylation of serum glycoproteins has not been studied in OSCC before. With this series, we were able to show that 67% of the N-glycopeptides significantly differentially expressed in OSCC versus healthy controls were fucosylated, and most of them were specifically core-fucosylated. In contrast with sialylated glycoforms, the fucosylated glycoforms were expressed in significantly higher levels in OSCC patients' sera, when compared with that of healthy controls.

A distinctive clinical problem with HNSCCs is the lack of biomarkers. During the last two decades, several studies have aimed to overcome this shortage [18]. Guerra et al conducted a review of the diagnostic accuracy of different proposed serum biomarkers in 2015 [18]. In the 65 studies included, enzyme-linked immunosorbent assay (ELISA) was the most used method, others being microarray, quantitative polymerase chain reaction and liquid chromatography mass spectrometry [18]. Mostly proteins but also DNAs, mRNAs, microRNAs and metabolites were suggested as potential biomarkers. From these only 13% of the single markers and 34 % of the panels (two or more biomarkers combined together) showed some real capacity to identify HNSCCs from healthy controls [18]. A main conclusion of Guerra et al was that none of the previously studied biomarkers seemed to be specific enough to reliably identify HNSCC patients. One advantage of studying serum glycopeptides as potential biomarkers is the significant responsiveness of their synthesis to biological conditions - due to their non-template biosynthesis [5]. This might offer the possibility to detect more disease-specific changes.

In this study, we were able to quantitatively analyze the N-glycopeptide profile of the serum samples of OSCC patients and healthy controls. The observed serum N-glycopeptide profiles of specifically IgG1 but also IgG4, haptoglobin and transferrin differed significantly between OSCC patients and healthy controls. These glycoproteins are typically expressed in acute phase response, and thus the interesting part is not their overall presence in OSCC patients' sera but the specific glycoforms observed in our comparison. Analyzing the information about N-glycopeptides as well as their attached glycan structures and their glycosylation sites is challenging, and there are not many publications we could compare our results with. For example, no studies describing the specific glycopeptide profiles of head and neck cancers exist. Only one study with released serum N-glycans (as opposed to N-glycopeptides in our study) in OSCC has been published [19]. This study had only 3 glycan compositions common with our study:

Composition 1: H3N4F1 (not significantly different between OSCC cases and controls in previously published study). In our study, composition 1 (H3N4F1) was found on multiple IgG1, IgG2 and IgG4 N-glycopeptides and in IgG1 (Fold change 14.1 and 10.5, Higher in OSCC) and IgG4 (Fold change 2.6 and 2.1, Higher in OSCC) it was significantly different between cases and controls.

Composition 2: S1H5N4 (not significantly different between OSCC cases and controls in their study). In our study, this composition was found on 1 N-glycopeptide each of Hemopexin (significantly different, 1.7 times higher in OSCC cases) and IgA1 (Not significantly different).

Composition 3: S2H5N4 (Significantly different between cases and controls in previously published study, 1.171 times lower in OSCC cases than controls). In our study this composition was found on 4 different N-glycopeptides (2 from Haptoglobin, 1 from Haptoglobin related protein and 1 form serotransferrin). 3 out of these 4 N-glycopeptides having this composition were singificantly different between OSCC cases and controls and fold changes ranged from 1.5 to 3.4.

This comparison also highlights that although these 2 tehcniques (Glycomics and Glycoproteomics) are complementary to each other, they still present a different picture of the biological differences. One of the reasons is that in Glycomics, all glycans having identical compositions (masses) will be detected as one species after glycan release even though they might have been attached to the very different proteins in very different amounts. However, in Glycoproteomics, the glycans still remain attached to the peptides and therefore will be detected as different peaks. This creates the abundance difference seen in the studies comparing Glycomics and Glycoproteomics.

Among other studies, two previous cancer studies analyzing a large amount of serum glycopeptides with label free MS in an unknown mixture have been published; one studying pancreas cancer vs acute pancreatitis vs healthy controls and the other oesophageal squamous cell carcinoma vs healthy controls [20, 21]. When the glycoforms of IgG1, IgG4, Haptoglobin and Transferrin in OSCC patients' sera were compared with those of patients with pancreatic cancer, they were found to be totally different [20]. The oesophagus is anatomically adjacent to oral cavity and also the histologies of esophageal and oral cavity squamous cell carcinomas share significant resemblance. In the case of oesophageal cancer one of the top four glycoproteins most significantly separating cancer patients' sera from those of healthy controls was haptoglobin [21]. The glycan structure S2H5N4, with glycosylation site 241, was found to be significantly overexpressed in both oesophageal and oral cavity cancer [21]. Other glycoforms of haptoglobin found in these two studies were different from each other as well as other glycopeptides separating these two cancers' sera from controls.

It seems that the specific glycoforms of glycoproteins distinguishing OSCC from healthy controls are specific to the disease and not just general alterations representing cancer.

One of the most interesting findings of this study was that of the four glycoforms of IgG1 that separate OSCC from controls, all were core-fucosylated, and their expression levels differed starkly between groups - with fold changes ranging from 9.5 to 15.1. We have previously studied the protein expression level changes in the sera of the same OSCC patients versus controls and none of the proteins found had a fold change over ten [22]. Although in that study, we were able to find a set of serum proteins whose expression levels reliably distinguished OSCC patients from healthy controls, it thus seems that studying the expression of glycopeptides is even more sensitive in differentiating OSCC patients from healthy individuals. Due to the depletion of the 12 most abundant proteins from the serum protein analysis, aiming to gain a larger (dynamic) range of proteins identified, which would not otherwise be detected, IgGs, haptoglobin and transferrin were not included in that study. This is a limitation of the current study, that we are unable to make a clear comparison between abundance in the two studies. However, in our opinion, it is quite unlikely that the high fold changes found in N-glycopeptides of IgG1, IgG4, haptoglobin and transferrin would solely be due to protein level changes. In addition, would that be the case, if a certain glycoform of e.g. IgG would provide the required specificity to serve as a biomarker for OSCC, the underlying reason for this phenomenon would be secondary.

We realize that sample size was modest in our study and it would count as potential limitation of the current pilot study. This was mainly due to the relatively low number of cases encountered at the clinic where the study based and it took approximately 2 years to collect these samples. However, the results are significant and interesting and we believe it would initiate validation of results at larger centers globally. We have also started to collaborate with other centers in Finland and we aim to carry out future studies of similar nature on larger sample set.

To conclude, we found IgG1s core-fucosylated glycoforms from serum to be able to distinguish OSCCs from controls. The fold changes were found to be several times higher than the differences found in the protein expression levels. It would be of the utmost importance to screen this set of glycoforms as biomarkers in a larger cohort of patients to establish their routine clinical use. This could be done using Multiple Reaction Monitoring (MRM) technology on triple quadrupole instrument [23]. Compared with MS techniques aiming to scan through a large scale of different peptides/glycopeptides, this technique allows higher sensitivity and also lower abundant compounds can be measured from complicated mixtures [23].

Funding

This study was supported by Korvatautien tutkimussäätiö and Jane and Aatos Erkko foundation. However, the funders had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Conflict of Interests

Authors declare no conflict of interests.

5. References

[1] J. Decker, J. C. Goldstein, N Engl J Med 1982, 306, 1151; E. M. Smith, J Natl Cancer Inst 1979, 63, 1189; E. L. Wynder, M. H. Mushinski, J. C. Spivak, Cancer 1977, 40, 1872.

[2] K. Dahiya, R. Dhankhar, World J Methodol 2016, 6, 77.

[3] B. A. van Dijk, M. T. Brands, S. M. Geurts, M. A. Merkx, J. L. Roodenburg, Int J Cancer 2016, 139, 574.

[4] E. Song, Y. Mechref, Biomark Med 2015, 9, 835.

[5] F. M. Tuccillo, A. de Laurentiis, C. Palmieri, G. Fiume, P. Bonelli, A. Borrelli, P. Tassone, I. Scala, F. M. Buonaguro, I. Quinto, G. Scala, Biomed Res Int 2014, 2014, 742831.

[6] U. Kuzmanov, N. Musrap, H. Kosanam, C. R. Smith, I. Batruch, A. Dimitromanolakis, E. P. Diamandis, Clin Chem Lab Med 2013, 51, 1467.

[7] S. A. Brooks, Mol Biotechnol 2009, 43, 76.

[8] R. Saldova, M. R. Wormald, R. A. Dwek, P. M. Rudd, Dis Markers 2008, 25, 219.

[9] J. N. Arnold, R. Saldova, U. M. Hamid, P. M. Rudd, Proteomics 2008, 8, 3284.

[10] S. R. Stowell, T. Ju, R. D. Cummings, Annu Rev Pathol 2015, 10, 473.

[11] C. Suarez Nieto, A. Cuesta Garcia, E. Fernandez Bustillo, J. C. Mendez Colunga, C. Alvarez Marcos, Clinical otolaryngology and allied sciences 1986, 11, 41; R. M. Rawal, P. S. Patel, B. P. Patel, G. N. Raval, M. M. Patel, J. M. Bhatavdekar, S. A. Dixit, D. D. Patel, Head & neck 1999, 21, 192; R. K. Shetty, S. K. Bhandary, A. Kali, Journal of clinical and diagnostic research : JCDR 2013, 7, 2818; S. Manoharan, M. Padmanabhan, K. Kolanjiappan, C. R. Ramachandran, K. Suresh, Clinica Chimica Acta 2004, 339, 91.

[12] S. Joenvaara, I. Ritamo, H. Peltoniemi, R. Renkonen, Glycobiology 2008, 18, 339; H. Peltoniemi, S. Joenvaara, R. Renkonen, Glycobiology 2009, 19, 707.

[13] M. Saraswat, S. Joenväära, L. Musante, H. Peltoniemi, H. Holthofer, R. Renkonen, Molecular & Cellular Proteomics 2014.

[14] O. M. Pearce, H. Laubli, Glycobiology 2016, 26, 111.

[15] F. L. Wang, S. X. Cui, L. P. Sun, X. J. Qu, Y. Y. Xie, L. Zhou, Y. L. Mu, W. Tang, Y. S. Wang, Cancer Detect Prev 2009, 32, 437; J. J. Park, M. Lee, Gut Liver 2013, 7, 629.

[16] F. Geng, B. Z. Shi, Y. F. Yuan, X. Z. Wu, Cell Res 2004, 14, 423; C. F. Tu, M. Y. Wu, Y. C. Lin, R. Kannagi, R. B. Yang, Breast Cancer Res 2017, 19, 111; H. Yin, Z. Tan, J. Wu, J. Zhu, K. A. Shedden, J. Marrero, D. M. Lubman, J Proteome Res 2015, 14, 4876.

[17] A. Mehta, T. M. Block, Dis Markers 2008, 25, 259.

[18] E. N. Guerra, D. F. Rego, S. T. Elias, R. D. Coletta, L. A. Mezzomo, D. Gozal, G. De Luca Canto, Crit Rev Oncol Hematol 2016, 101, 93.

[19] S.-Y. Guu, T.-H. Lin, S.-C. Chang, R.-J. Wang, L.-Y. Hung, P.-J. Fang, W.-C. Tang, P. Yu, C.-F. Chang, PLoS ONE 2017, 12, e0178927.

[20] H. Kontro, S. Joenvaara, C. Haglund, R. Renkonen, Proteomics 2014, 14, 1713.

[21] A. Mayampurath, E. Song, A. Mathur, C. Y. Yu, Z. Hammoud, Y. Mechref, H. Tang, J Proteome Res 2014, 13, 4821.

[22] M. Saraswat, A. Makitie, R. Agarwal, S. Joenvaara, S. Renkonen, British journal of cancer 2017, 117, 376.

[23] Q. Hong, L. R. Ruhaak, S. M. Totten, J. T. Smilowitz, J. B. German, C. B. Lebrilla, Anal Chem 2014, 86, 2640.

Figure legends

Figure 1: Principal component analysis (PCA).

A. PCA of all quantified potential N-glycopeptides in Oral Squamous Cell Carcinoma cases and controls. PCA was performed with software Progenesis QI for Proteomics. In this panel, all N-glycopeptides are considered for PCA. Orange circles are cases and purple circles are controls.

B. PCA of only those potential N-glycopeptides which pass the cutoff of ANOVA p value 0.05 in Oral Squamous Cell Carcinoma cases and controls comparison. PCA was performed with software Progenesis QI for Proteomics and ANOVA was calculated with the same. In this panel, ANOVA p value cutoff (0.05) passing N-glycopeptides are considered for PCA. Orange circles are cases and purple circles are controls.

Figure 2. Representative annotated MS/MS spectra of an N-glycopeptide.

This figure contains the annotated MS/MS spectra of an N-glycopeptide (m/z 1107.1462, IgG1 glycoform) with peptide sequence of TKPREEQYNSTYR and glycan composition of H3N5F1. Here H is a hexose, N is N-acetylglucosamine and F is fucose. Total Glycopeptide score (explained in methods) is 69.72 and 58% of the fragment intensity was explained by annotation. This spectrum is the output of software GlycopeptideId, which is publicly available web-based software. This software takes .pkl or .mgf deconvoluted spectrum files as its input.

Figure 3: Proposed structures of the IgG1 N-glycopeptide ions.

An N-glycopeptide of the IgG1 is shown here which was found associated with seven ions (m/z values of intact N-glycopeptides are indicated below the proposed structure). Four ions shown from left to right in the figure had unique proposed structures. The other 3 ions at the right side in the figure (m/z of 1039.456 (z=3), 1093.47 (z=3) and 1107.146 (z=3)) had same proposed structures as the ones on the left (m/z 779.847 (z=4), 820.36 (z=4) and 830.609 (z=4) respectively). These ions had different m/z values because of differential amount of charges on them. The m/z values given are of the intact N-glycopeptide ion including the peptide part. All of these glycoforms were core-fucosylated and had higher expression values in OSCC compared to controls. The fold changes (FC) of every N-glycopeptide ion is shown at the top of the proposed structure. The spectrum of these N-glycopeptide ions was matched to database entries (GlyycomeDB) to infer these proposed structures. It is not possible with current technology (CID-MS/MS) used in our experiment to infer linkage information (such as α- and β-glycosidic bonds), only the order of the attachment of monosaccharides is known. Therefore the linkage information is not shown n the diagram. Blue squares represent N-acetylhexosamines, green circles mannoses and yellow circles galactoses.

Figure 4: Box plot of glycoforms of IgG1.

Box plots depicting the expression levels of IgG1 glycoforms in all samples studied. Standard error with one sigma is shown in the graph.

Figure 5: Box plot of glycoforms of Haptoglobin.

Box plots depicting the expression levels of Haptoglobin glycoforms in all samples studied. Standard error with one sigma is shown in the graph.