

UCSF

UC San Francisco Previously Published Works

Title

Mass spectrometry-based analyses showing the effects of secretor and blood group status on salivary N-glycosylation

Permalink

<https://escholarship.org/uc/item/0cv5x5j6>

Journal

Clinical Proteomics, 12(1)

ISSN

1542-6416

Authors

Albertolle, Matthew E

Hassis, Maria E

Ng, Connie Jen

et al.

Publication Date

2015-12-01

DOI

10.1186/s12014-015-9100-y

Peer reviewed

RESEARCH

Open Access



# Mass spectrometry-based analyses showing the effects of secretor and blood group status on salivary *N*-glycosylation

Matthew E. Albertolle<sup>1,2†</sup>, Maria E. Hassis<sup>1,2†</sup>, Connie Jen Ng<sup>1,2</sup>, Severino Cuison<sup>1,2</sup>, Katherine Williams<sup>1,2</sup>, Akraporn Prakobphol<sup>1,2</sup>, Andrew B. Dykstra<sup>1,2</sup>, Steven C. Hall<sup>1,2</sup>, Richard K. Niles<sup>1,2</sup>, H. Ewa Witkowska<sup>1,2</sup> and Susan J. Fisher<sup>1,2\*</sup>

## Abstract

**Background:** The carbohydrate portions of salivary glycoproteins play important roles, including mediating bacterial and leukocyte adhesion. Salivary glycosylation is complex. Many of its glycoproteins present ABO and Lewis blood group determinants. An individual's genetic complement and secretor status govern the expression of blood group antigens. We queried the extent to which salivary glycosylation varies according to blood group and secretor status. First, we screened submandibular/sublingual and parotid salivas collected as ductal secretions for reactivity with a panel of 16 lectins. We selected three lectins that reacted with the largest number of glycoproteins and one that recognized uncommon lactosamine-containing structures. Ductal salivas representing a secretor with complex blood group expression and a nonsecretor with a simple pattern were separated by SDS-PAGE. Gel slices were trypsin digested and the glycopeptides were individually separated on each of the four lectins. The bound fractions were de-*N*-glycosylated. LC-MS/MS identified the original glycosylation sites, the peptide sequences, and the parent proteins.

**Results:** The results revealed novel salivary *N*-glycosites and glycoproteins not previously reported. As compared to the secretor, nonsecretor saliva had higher levels of *N*-glycosylation albeit with simpler structures.

**Conclusions:** Together, the results suggested a molecular basis for inter-individual variations in salivary protein glycosylation with functional implications for oral health.

**Keywords:** Saliva, Secretor, Glycosylation, *N*-glycosite, Lectin, Mass spectrometry

## Background

Saliva—the product of the parotid, submandibular, sublingual and minor salivary glands—is a complex biological fluid that exists in two phases. The fluid phase flows over oral surfaces and is swallowed. Salivary components can also be immobilized within the oral cavity, e.g., the pellicle that coats the tooth surface. In both phases, saliva plays many critical roles in general aspects of health (reviewed in [1, 2]). Examples include regulation of the oral microbiome by modulating the adherence, growth

and/or viability of a diverse array of organisms. Other important functions include promoting tissue homeostasis, taste and lubrication. The critical roles of this body fluid are illustrated by the myriad pathologies that accompany xerostomia (abnormally low salivary flow), which include dental caries and an increased susceptibility to oral infections, e.g., candidiasis.

Saliva has inorganic as well as protein, glycoprotein, peptide and carbohydrate constituents. During the past decade, several compilations of the salivary proteome have been published. In total, more than 3000 proteins/peptides have been identified (reviewed in Amado et al. [3]) with ductal saliva having about 400 major constituents [4]. Specific salivary functions are parsed among these components. In addition to its well-known

\*Correspondence: sfisher@cgl.ucsf.edu

<sup>†</sup>Matthew E. Albertolle and Maria E. Hassis contributed equally

<sup>2</sup> Sandler-Moore Mass Spectrometry Core Facility, University of California San Francisco, San Francisco, CA 94143, USA

Full list of author information is available at the end of the article

enzymatic activity, fluid phase salivary amylase binds to streptococci mediating their clearance from the oral cavity [5]. Saliva also contains a rich repertoire of peptides—members of the histatin, statherin, proline rich protein, and cystatin families—that regulate bacterial adhesion, have microbicidal activity and modulate calcium phosphate chemistry [6, 7]. In general, the role of the carbohydrate motifs in body fluids is much less well understood.

Salivary glycoproteins play a particularly important role in oral health because their carbohydrate constituents interact with a wide array of bacteria. Interestingly, their glycan repertoire is, in part, genetically determined as these structures include the carbohydrate epitopes that comprise the ABO, Lewis (Le) and other blood group determinants. Secretor status further complicates glycosylation patterns. Non-secretors have an inactive form of fucosyltransferase 2 (FUT2), which provides the glycan scaffold for Le<sup>b/y</sup> and blood-type motifs [8]. Thus, individuals who lack this transferase should have less fucosylated glycoforms than secretors who have an active FUT2. What are the biological consequences? Many of these glycans serve as receptors that mediate adhesion of bacteria that colonize the oral cavity—e.g., the T-Ag for *Actinomyces naeslundii*; the sT-Ag for several streptococcal strains [9]; and the Le determinants, sLe<sup>a</sup>, Le<sup>b</sup>, and Le<sup>x</sup>, for *H. pylori* strains [10, 11]. The major glycan carried by the glycosylated proline rich protein (gPRP) is a biantennary oligosaccharide with a difucosylated lactosamine (Le<sup>y</sup>) sequence on one antenna and an unsubstituted lactosamine sequence on the other. *Fusobacterium nucleatum*, a microorganism associated with periodontal disease, binds to the gPRP via terminal unmodified lactosamine sequences; fucosylation of this disaccharide blocks bacterial binding [12].

Here we asked, at a global level, whether secretor status affects the arrangement of carbohydrate chains along the amino acid backbone of salivary proteins. To answer this question we used a modification of a workflow that our group devised [13]. First, we identified the lectins (carbohydrate binding proteins) that best discriminated between salivary samples from secretors vs. nonsecretors. To do so, we screened parotid and submandibular/sublingual (SMSL) salivas collected as the ductal secretions from the two groups with a large panel of lectins that recognized various aspects of glycan structure. To compare the *N*-glycosylation sites of secretors and nonsecretors, we used columns formed from the lectins with the highest discriminating power or unusual specificities to fractionate electrophoretically separated and trypsin digested saliva samples from the two donor groups. After elution of the bound glycopeptides, the samples were treated with peptide-*N*-glycosidase F (PNGase F), an amidase, and the original *N*-glycosylation sites (NXS/T;

X ≠ proline) were identified by ESI HPLC–MS/MS via a +1 Da mass shift at the modified Asn. Previously reported and novel *N*-glycosites were identified. Incidentally, we also detected genetic polymorphisms along the peptide backbones some of which were novel. The impact of secretor status on *N*-oligosaccharide composition was demonstrated by MALDI MS analyses of the released glycans, which showed higher degrees of fucosylation among the secretor oligosaccharides. Interestingly, more sites were identified when the saliva sample was from a nonsecretor. However, studies on larger number of individuals are needed to further investigate the potential impact of a secretor status on *N*-glycosite occupancy.

## Results and discussion

### Lectin selection

As a first step, we screened SMSL and parotid saliva samples collected as the ductal secretions from twenty individuals for reactivity with a 16-lectin panel (Additional file 1: Figure S1). The donors were chosen as representing the spectrum of glycosylation commonly observed in the general US population with regard to the addition of carbohydrate blood group determinants to the core oligosaccharide structures of glycoproteins. We also evaluated relative expression of L-selectin carbohydrate ligands, e.g., MECA-79 reactivity, which are added to salivary components [14]. From the original twenty individuals, we selected four with the following characteristics: donor (1) secretor with blood type O, Le<sup>a+</sup>, Le<sup>b+</sup>, Le<sup>y+</sup>, low MECA-79; donor (2) secretor with blood type B, Le<sup>a+</sup>, Le<sup>b+</sup>, Le<sup>y+</sup>, low MECA-79; donor (3) secretor with blood type O, Le<sup>a+</sup>, Le<sup>b+</sup>, Le<sup>y+</sup>, low MECA-79; and donor (4) nonsecretor with blood type A, Le<sup>a-</sup>, Le<sup>b-</sup>, Le<sup>y-</sup>, high MECA-79. The 16 lectins were chosen based on their carbohydrate specificity, which spanned a wide spectrum from elements that are commonly found in the majority of *N*-linked structures to unusual sugar sequences and/or linkages (Table 1). The results are shown in Fig. 1. Panel A shows the electrophoretic banding patterns of SMSL and parotid, proteins (Coomassie blue staining) and glycoproteins (Alcian blue silver staining), respectively. Of the lectins that were screened, AAL, jacalin (JAC) and wheat germ agglutinin (WGA) reacted with the largest number of bands spanning the greatest molecular weight range (compare Fig. 1b with Additional file 1: Figure S1). Thus, using the rationale that they would capture the highest number of *N*-linked glycopeptides, they were chosen for the separation experiments. In this regard, we also included the *Lycopersicon esculentum* agglutinin (LEA) lectin, which has a much narrower specificity consistent with the banding pattern shown in Fig. 1c, but includes lactosamine units that could be substrates for the addition of the Le antigens.

**Table 1 Lectins and their carbohydrate binding specificities used to screen SMSL and parotid saliva**

Name	Carbohydrate sequence(s)	References
<i>Artocarpus integrifolia</i> (Jacalin)	Gal $\beta$ 1-3GalNAc, Gal $\alpha$ 1-6Gal	[62]
<i>Erythrina christagalli</i> (ECA)	Gal $\beta$ 1-4GlcNAc	[63]
<i>Arachis hypogaea</i> (peanut, PNA)	Gal $\beta$ 1-3GalNAc	[64]
<i>Glycine max</i> (soybean, SBA)	GalNAc $\alpha$ 1-3GalNAc GalNAc $\alpha$ / $\beta$ 1-3/4Gal	[65]
<i>Vicia villosa</i> (VVA)	GalNAc-Ser	[66]
<i>Datura stramonium</i> (DSA)	(GlcNAc $\beta$ 1-4)2-4, Gal $\beta$ 1-4GlcNAc	[67]
<i>Lycopersicon esculentum</i> (tomato, LEA)	(GlcNAc $\beta$ 1-4)1-4	[68]
<i>Triticum vulgare</i> (wheat germ, WGA)	(GlcNAc $\beta$ 1-4)2-5, Neu5Ac	[69, 70]
	Man $\beta$ 1-4GlcNAc $\beta$ 1-4GlcNAc	[71]
<i>Canavalia ensiformis</i> (Con A)	Branched <i>N</i> -linked hexa-saccharide	[72]
<i>Galanthus nivalis</i> (GNA)	Man $\alpha$ 1-3Man	[73]
<i>Lens culinaris</i> (LCA)	Fuca1-6GlcNAc- <i>N</i> -Asn containing <i>N</i> -linked oligosaccharides	[72]
<i>Aleuria aurantia</i> (AAL)	Fuca1-6/3GlcNAc	[74]
<i>Lotus tetragonolobus</i> (LTA)	Fuca1-2Gal $\beta$ 1-4(Fuca1-3)GlcNAc	[75]
<i>Sambucus nigra</i> (SNA)	Neu5Ac $\alpha$ 2-6Gal/GalNAc	[76]
<i>Evonymus europaeus</i> (EEA)	Gala1-3(Fuca1-2)Gal $\beta$ 1-3/4GlcNAc	[77]
<i>Phaseolus vulgaris erythroagglutinating</i> (PHA-E)	<i>N</i> -linked bi-antennary	[78]

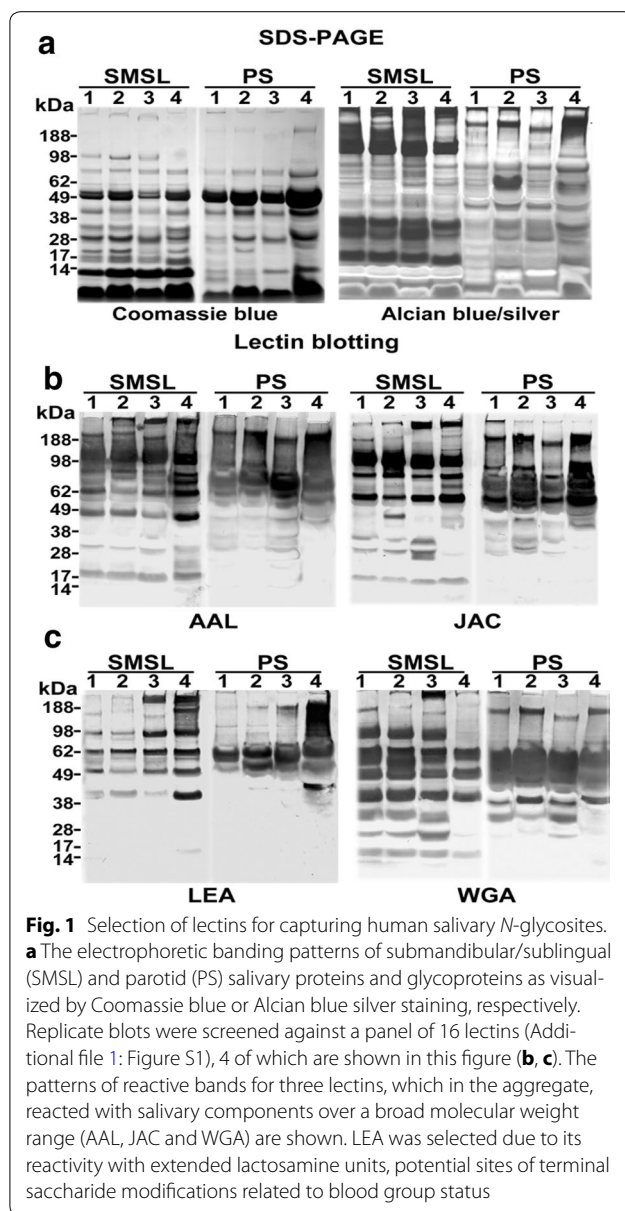
### Piloting the workflow

An overview of the general method that we devised is shown in Fig. 2. SMSL or parotid salivas were separated by preparative SDS-PAGE. The entire gel was horizontally rastered, macerated and subjected to trypsin digestion. The resulting mixture of peptides and glycopeptides was separated on an immobilized lectin column. The bound glycopeptides were eluted and treated with PNGase F, removing the *N*-linked oligosaccharides and converting the asparagines to aspartic acids. LC-MS/MS of the digest enabled sequencing of the peptides and identification of the original sites of carbohydrate attachment.

Prior to scaling up the analysis, we performed a pilot experiment to test the proposed workflow by analyzing 100  $\mu$ L of parotid and the same amount of SMSL saliva from one donor (secretor, blood type O, Le<sup>a</sup>, Le<sup>b</sup>-, Le<sup>y</sup>-, and MECA-79 low). The electrophoretically separated samples were rastered into 12 gel slices and chromatographed on immobilized AAL. LC-MS/MS analyses, a total of 24 runs, identified 31 *N*-glycosites from 21 glycoproteins (Fig. 3). To our knowledge, 4 were not previously reported in saliva: IGHG2\_HUMAN@176 (Ig gamma-2 chain C region), PRB3\_HUMAN@66 (Basic salivary proline-rich protein 3), PSG1\_HUMAN@104 and @111 (Pregnancy-specific beta-1-glycoprotein 1), the latter two within the same tryptic peptide. The PSG1 *N*-glycosites were recently reported in the *N*-glycoproteome of human metastatic hepatocellular carcinoma cell lines [15]. At the glycoprotein level, IGHG2\_HUMAN and

PRB3\_HUMAN were known salivary components [16, 17]; the detection of PSG1\_Human in this body fluid was novel. This was despite the fact that the reproductive age, female donor was not pregnant at the time of sample collection. Nevertheless, this finding raises the interesting possibility that this family of glycoproteins, which are produced in large amounts by the placenta and circulate at high levels in maternal blood, can be detected in saliva during pregnancy. Whether their levels could be indicative of the placental dysfunction that is associated with complications such as preeclampsia is an interesting question that could be addressed in the future. Finally, we note that future studies are necessary to validate these new glycosites since chemical deamidation of asparagine residues during sample preparation can potentially lead to false positive identification of glycosites [18, 19].

The majority of identified *N*-glycosites and associated glycoproteins were identified in SMSL secretions, with parotid saliva contributing only two unique *N*-glycosites, N83 and N90, within the same tryptic peptide of PIGR\_HUMAN (Polymeric Ig receptor). The undersampling limitations of the data-dependent 'shotgun' mode of data acquisition that was utilized did not allow for unequivocal conclusions about the absence of specific *N*-glycosites in a sample [20], a question that would best be answered with targeted analyses. Of note, a recently published library containing selected reaction monitoring assays that were developed to enable antibody-independent MS-based analyses of potential *N*-glycosylation sites in body fluids included a number of peptide targets



identified in this study (see Additional file 2: Table S1) [21]. In conclusion, the results of the pilot experiment demonstrated the feasibility of applying, to SMSL and parotid salivas, a workflow combining protein fractionation (via SDS-PAGE), in-gel trypsin digestion, lectin capture of glycopeptides carrying *N*-linked oligosaccharides, and LC-MS/MS identification for profiling glycosylation patterns in saliva.

#### The effects of secretor status on *N*-glycosylation patterns

Salivary glycoproteins display a diverse carbohydrate repertoire due, in part, to the addition of blood group determinants to many oligosaccharides. Therefore, we were interested, at a global level, in the effects of this

#### Submandibular/sublingual or parotid saliva

Preparative SDS PAGE

Raster protein bands

In-gel trypsin digestion

*N*-glycopeptide capture via lectin chromatography

PNGase F removal of *N*-linked glycans

LC-MS/MS and MALDI TOF analyses of peptides

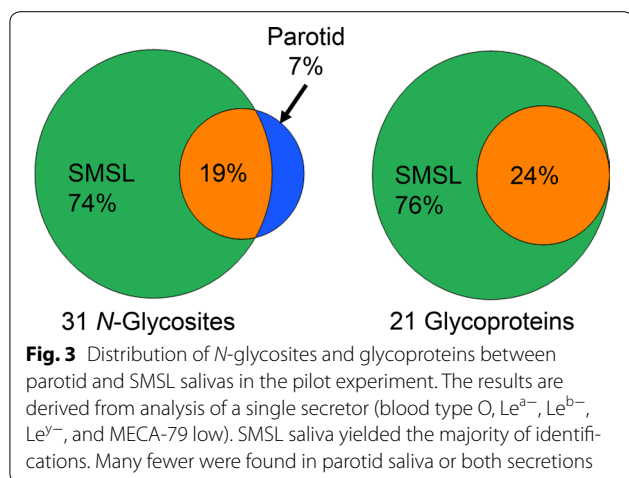
Data analysis to identify *N*-glycosites, their parent glycoproteins, and postulate *N*-glycan structures

**Fig. 2** Experimental workflow used to identify salivary *N*-glycosites. Parotid and SMSL salivas collected as the ductal secretions were separated by preparative SDS-PAGE followed by rastering of the gel bands and in-gel trypsin digestion. *N*-glycopeptides were captured from the peptide extracts via their ability to bind to at least one of four lectins (AAL, JAC, WGA and/or LEA). After elution and PNGase F removal of *N*-linked glycans, LC-MS/MS was performed to identify the original sites of oligosaccharide attachments and the protein scaffolds

phenomenon on *N*-linked glycosylation patterns. To address this point, we applied the method outlined in Fig. 2 to 500  $\mu$ L of SMSL or parotid saliva, expanding the rastering of the SDS-PAGE gels to 18 slices and the LC-MS/MS runs to 300, for the analysis of SMSL and parotid samples at both ends of the glycosylation spectrum. All results presented in this section were generated using saliva samples from two donors, a secretor and a nonsecretor.

The first donor was a secretor with blood type AB and complex patterns of terminal fucosylation as determined by broad Le immunoreactivity (anti-Le<sup>a</sup>, anti-Le<sup>b</sup> and anti-Le<sup>y</sup>) and relatively low levels of the MECA-79 epitope (Fig. 4a). The second donor with the simplest pattern was a nonsecretor with blood type O. Immunoblotting revealed that the salivary secretions displayed Le<sup>a+</sup>, but were Le<sup>b-</sup> and Le<sup>y-</sup> with high levels of MECA-79 immunoreactivity (Fig. 4a). An overview of the protein composition of the samples, as determined by SDS-PAGE showed similar patterns to those observed in saliva

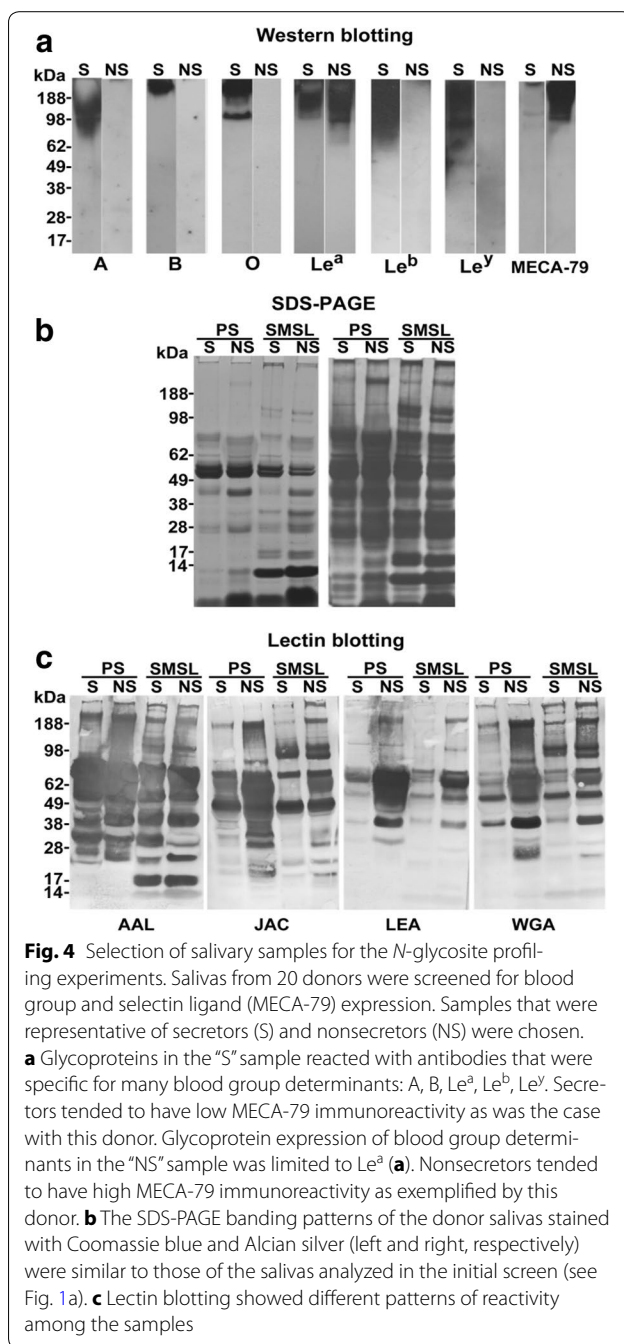




samples from the other donors (compare Figs. 4b and 1a), which was also true for the lectin blotting results (AAL, jacalin, LEA, and WGA; compare Figs. 4c and 1b). Thus, we glycotyped the donors and showed that, in general terms, their SMSL and parotid samples had the expected repertoire of proteins and glycoproteins.

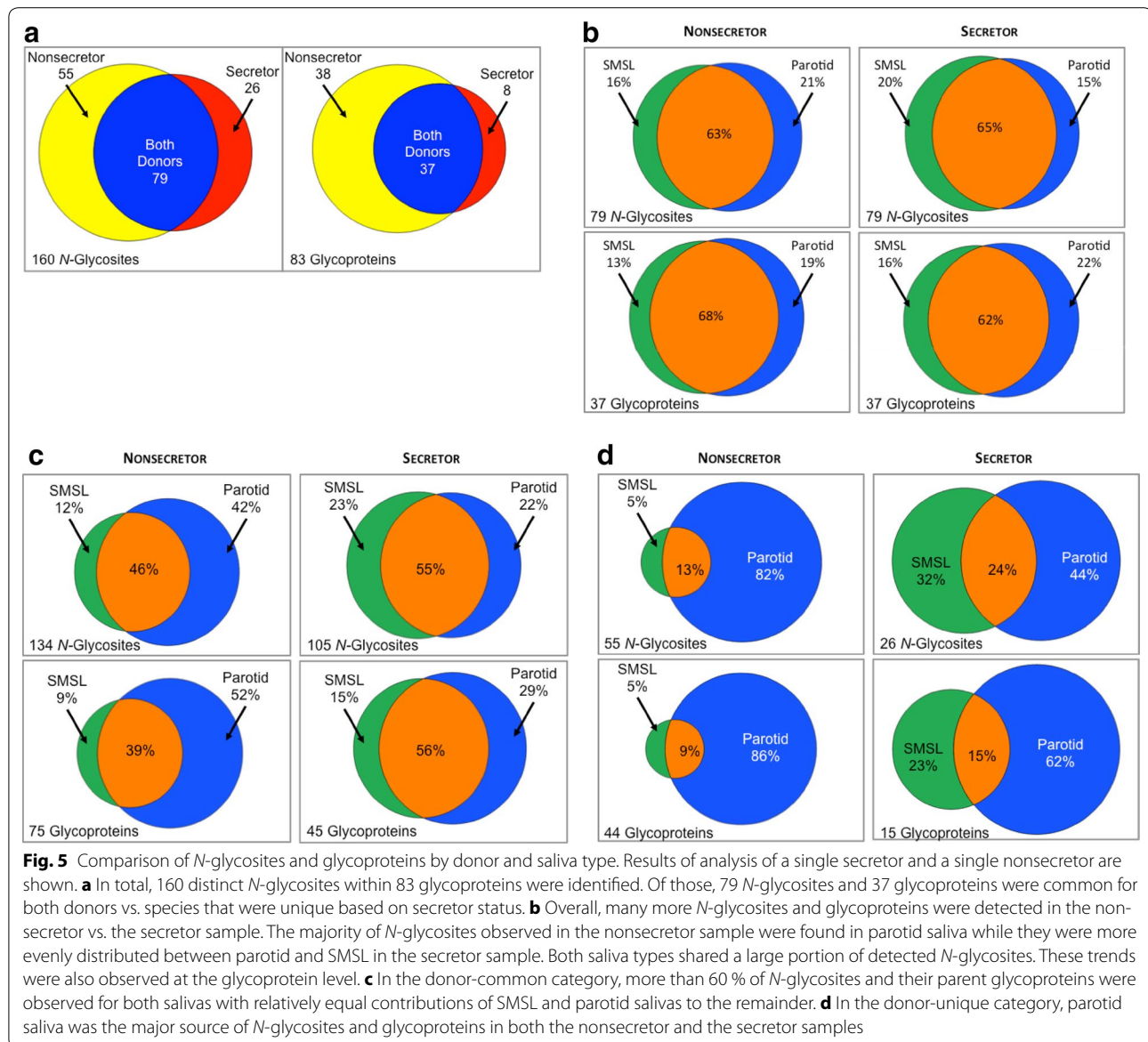
With regard to an overview of the LC-MS/MS results, Additional file 2: Table S1 lists all peptide identifications and their associated *N*-glycosites, including the sites identified exclusively in the pilot experiment. The spectra are shown in Additional file 3: Figure S2. Overall, 2296 distinct peptides (54,298 spectral counts) were identified in the SMSL and parotid salivas of the nonsecretor vs. 1928 distinct peptides (43,197 spectral counts) in salivas from the secretor. Using our established criteria (see “Methods”), a total of 160 distinct *N*-glycosites within 83 glycoproteins were identified. First, we assessed the results in terms of the *N*-glycosites and their parent glycoproteins that were common for the donors vs. species that were unique based on secretor status. In the common category, 79 *N*-glycosites (Fig. 5a, left) and 37 glycoproteins (Fig. 5a, right) were detected in both donors. More donor-unique *N*-glycosites were identified in the nonsecretor samples (55 from 38 proteins) as compared to those of the secretor (26 from 8 proteins).

We then examined, in detail, the distribution of *N*-glycosites across both donors and saliva types. Out of 160 distinct *N*-glycosites, 134 (4249 spectral counts) were confirmed in the nonsecretor salivas as compared to 105 distinct *N*-glycosites (4562 spectral counts) in the secretor samples (Fig. 5b). In 8 cases, 7 involving both nonsecretor and secretor samples and 1 instance in the analysis of the nonsecretor sample, the two *N*-glycosites resided within a single tryptic peptide. Some *N*-glycosites were observed within peptides of different lengths—due to truncation, proteolytic cleavage or missed cleavage



sites—thus increasing the confidence of their assignments (examples shown in Additional file 2: Table S1, Additional file 3: Figure S2).

As shown in Fig. 5b, more *N*-glycosite species (upper panels) were found in the nonsecretor samples (134 total, upper left) than in those of the secretor (105 total, right). This was also true at the level of the glycoproteins from which they were derived (75 vs. 45; bottom panels). However, spectral counting suggested that higher



copy numbers of *N*-glycosites were detected in the secretor vs. the nonsecretor sample (Additional file 4: Figure S3A). This was despite the fact that the spectral count distribution of non-glycosylated peptides was virtually identical for both sample types (Additional file 4: Figure S3B). Parotid saliva was a major source of *N*-glycosites in the nonsecretor samples (Fig. 5b, upper left) while a more balanced contribution of SMSL and parotid salivas was observed for the secretor (Fig. 5b, upper right). It is interesting to note that this result was consistent with the lectin binding patterns of the electrophoretically separated samples (Fig. 4c). For both donors, a large fraction of detected *N*-glycosites was common to the two saliva types.

In the donor-common category, more than 60 % of *N*-glycosites and their parent glycoproteins were observed for both salivas with relatively equal contributions of SMSL and parotid salivas to the remainder (Fig. 5c, upper and lower panels). In contrast, parotid saliva was the major source of donor-unique *N*-glycosites and glycoproteins in both the nonsecretor and the secretor samples (Fig. 5d, upper and lower panels).

In a cumulative analysis of the entire data set, the relative abundance of each *N*-glycosite was estimated by the number of spectral counts observed for all deglycosylated peptides encompassing that site. The counts varied significantly from 1 to more than 400 (433 for ZA2G\_HUMAN@109 [Zinc-alpha-2-glycoprotein] in

the nonsecretor and 445 for AMY1\_HUMAN@476 [Amylase] the secretor. The *N*-glycosites were binned into 7 groups based on the number of associated spectral counts (Additional file 5: Figure S4). *N*-glycosites that were common among the donors had similar distributions in nonsecretor (median/average = 15/49) and secretor samples (median/average = 16/59, left Panel). *N*-glycosites that were unique to a particular donor were present in lower abundances (Additional file 5: Figure S4, right Panel).

In addition, by virtue of our experimental design, we also observed protein isoforms and mutations. For example, an *N*-glycosite unique for isoform 3 of C1q tumor necrosis factor-related protein 3 (C1QT3\_HUMAN) was observed in both donors. Interestingly, we failed to detect the canonical structure for the carbonic anhydrase 6 (CAH6\_Human) tryptic peptide encompassing the *N*-glycosite at position 67. Rather, a common mutation, 68 M→L (dbSNP ID: rs2274328), was found in both donor samples (Additional file 3: Figure S2, spectrum 26a). In addition, the nonsecretor carried another mutation at 90S→G (dbSNP ID: rs2274329) that resided with 68 M→L in the same tryptic deglyco-peptide (Additional file 3: Figure S2, spectrum 26b), suggesting compound heterozygosity for these variants. To the best of our knowledge, this double mutant was not previously reported. A number of recent studies addressed the impact of polymorphic CAH6 structures on salivary parameters with the goal of discerning associated phenotypes of potential clinical significance [22–25].

Together, these data pointed to the potential impact of secretor status on utilization of *N*-glycosites. Namely, these simpler glycosylation patterns of the nonsecretor could render the peptide backbone of salivary glycoproteins more accessible to glycosyltransferases, which would have the net effect of increasing the number of unique *N*-glycosylation sites [26]. However, the secretor sample had more oligosaccharides, in absolute terms, which could interact with the lectins we used, e.g., fucose residues that bind to AAL. This finding is consistent with the ability of secretors to add additional fucose residues to the simpler glycan termini typical of nonsecretors. Given that one glycosylation site can carry oligosaccharides with many different structures, the net effect would likely be an increased probability of any given secretor glycopeptide binding to AAL.

### Lectin performance

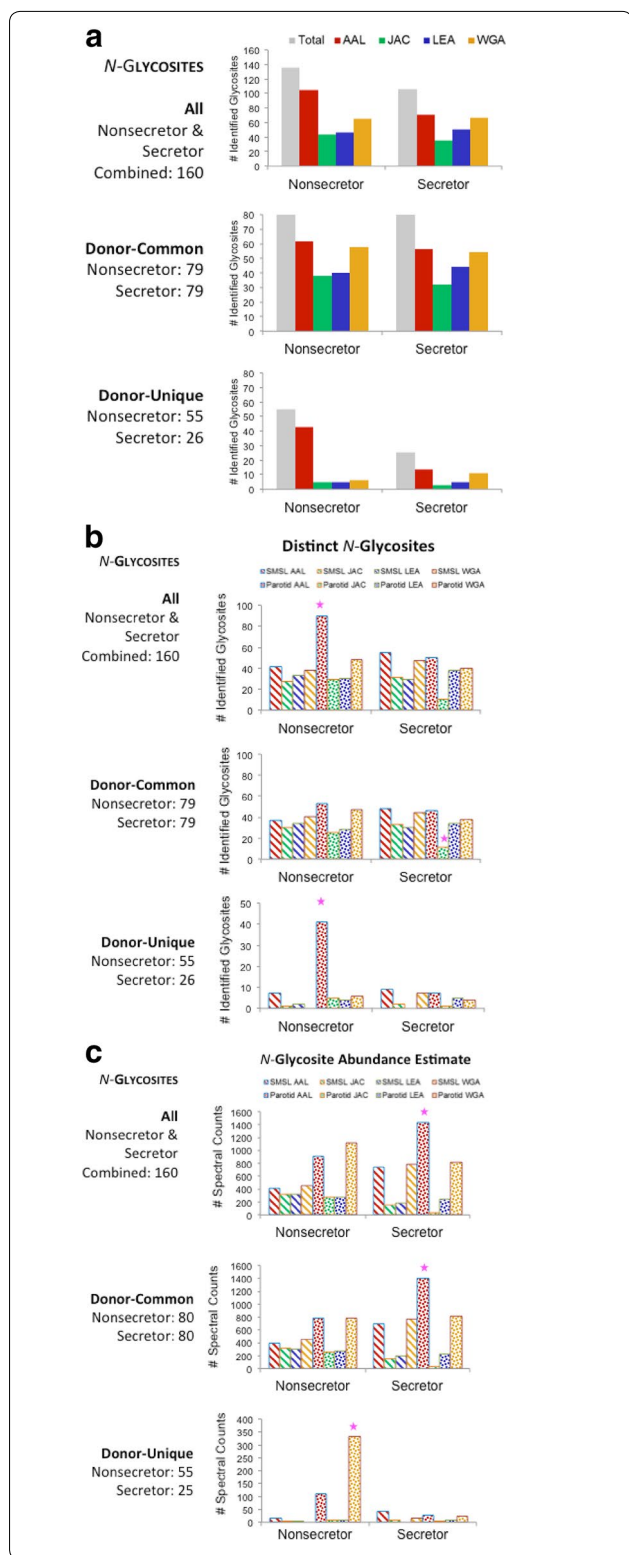
Parotid and SMSL salivas from a single secretor and a single nonsecretor of the glycotypes described above were studied in this series of experiments. Figure 6a illustrates the efficiency of the lectins employed in terms of the number of *N*-glycosites identified summed for

SMSL and parotid salivas. None of the results achieved statistical significance in terms of being greater than 2 SD above or below the mean values. However, trends were observed. With regard to the total numbers, AAL enrichment tended to lead to the highest number of identifications (top panels). There was no difference in lectin performance in terms of *N*-glycosites that were common among donors (middle panel). A trend was observed in which a higher number of donor-unique *N*-glycosites tended to be captured by AAL from the nonsecretor sample (lower panel). Next, we explored these results in terms of each saliva type (Fig. 6b). Overall, AAL was the only lectin that outperformed the others, capturing significantly more *N*-glycosites from the parotid saliva sample of the nonsecretor (upper Panel). With regard to *N*-glycosites that were common among donors, the same trend was observed (middle Panel). However, we also noted the low recovery of *N*-glycosites from the secretor parotid sample following jacalin capture. With regard to donor-unique sites, AAL capture from the parotid sample of the nonsecretor once again led to the highest number of unique identifications (lower panel). AAL separation also led to the greatest number of unique *N*-glycosite identifications per sample-lectin combination (Additional file 6: Figure S5; left Panel). However, it was the secretor-SMSL-jacalin combination that resulted in the highest fraction of recovered unique glycopeptides (Additional file 6: Figure S5; right Panel).

Finally, we used relative abundance (e.g., spectral counts) to compare lectin performance (Fig. 6c). Overall, AAL tended to capture the highest number of *N*-glycosites from the parotid sample of the donor who was a secretor; WGA followed by AAL had the highest efficiency in terms of enrichment from parotid saliva of the nonsecretor (upper panel). In terms of donor-common species (middle panel), the highest number of *N*-glycosites was observed in the AAL bound fraction of the secretor parotid saliva sample. Finally, in terms of donor-unique sites, WGA captured the highest numbers from the nonsecretor parotid saliva sample (lower panel). Of note, the latter result was largely driven by the preferential isolation of FIBB\_HUMAN@394.

The overall performance of AAL was in keeping with the fact that this lectin binds fucose-containing oligosaccharides such as the Le determinants that differed among our donors. In contrast, *N*-glycosites that were common among donors were more likely to have broader lectin affinity perhaps due to their less specialized underlying carbohydrate structures. The poor performance of jacalin was in accord with its *N*-acetylgalactosamine specificity, a sugar that is typically a component of *O*-linked oligosaccharides, which this study did not interrogate.





**Fig. 6** Lectin capture efficiency, in terms of number of *N*-glycosites identified and relative abundance. Results shown are of analysis of a single secretor and a single nonsecretor. **a** (upper) Overall, AAL enrichment tended to yield the greatest number of *N*-glycosites. **a** (middle) With regard to *N*-glycosites that were common among donors, lectin performance did not depend on secretor status (**a**, lower) whereas more *N*-glycosites tended to be captured by AAL from the nonsecretor sample. **b** (upper) Overall, AAL demonstrated the greatest capture efficiency, enriching more *N*-glycosites from the parotid saliva sample of the nonsecretor (starred). **b** (middle) With regard to *N*-glycosites that were common among donors, the low level of jacalin capture of parotid sites from the secretor sample was evident (starred). **b** (lower) With regard to donor-unique *N*-glycosites, AAL capture from parotid saliva of the nonsecretor was once again most productive in terms of a number of identified *N*-glycosites (starred). **c** (upper) As to relative abundances in terms of spectral counts and overall performance, WGA and AAL capture tended to have the highest efficiency. **c** (middle) In terms of donor-common species, the highest number of *N*-glycosites tended to be found in the AAL bound fraction of the secretor parotid saliva sample. **c** (lower) In terms of donor-unique sites, WGA captured the highest numbers from the nonsecretor parotid saliva sample (starred)

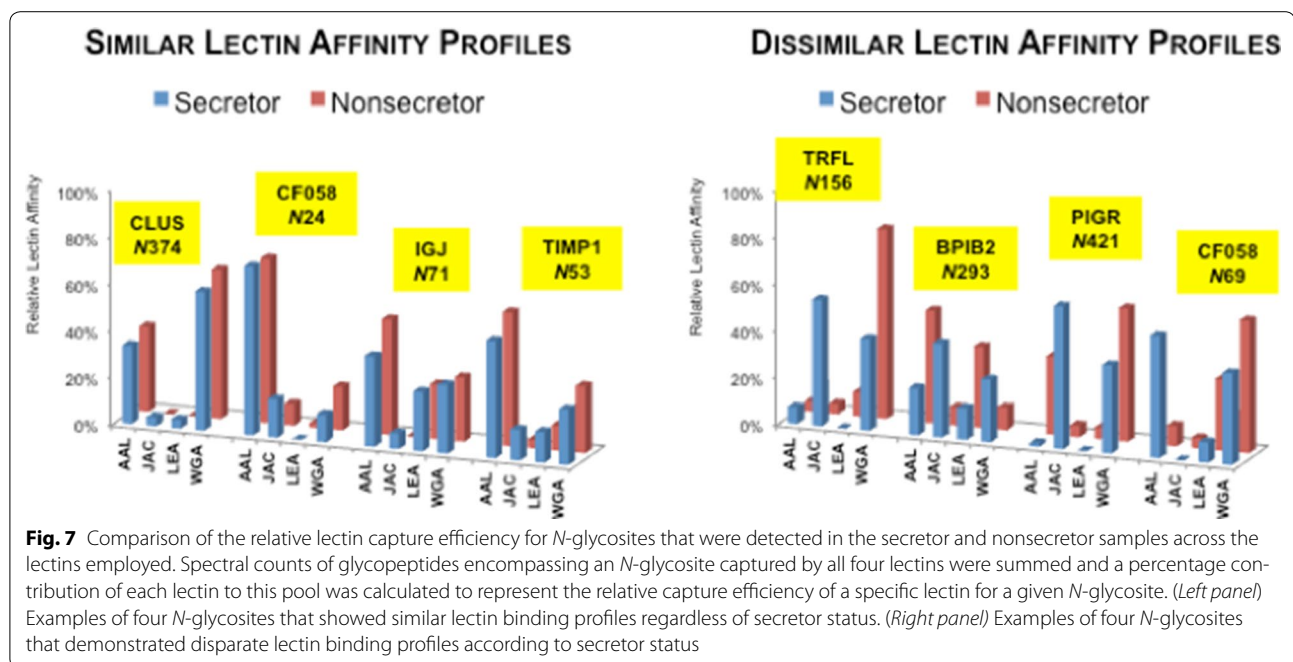
the number of unique identifications is the goal, then AAL capture is the best discriminator of *N*-glycosites as applied to parotid saliva samples from nonsecretors vs. secretors. If measures of relative abundance are of interest, then WGA ± AAL should be used for enrichment.

**Differences in *N*-glycosite occupancy between the donors**

For this analysis, we compared one secretor and one nonsecretor and considered only common sites with ≥30 counts per donor. Of the 45 *N*-glycosites that met these criteria, most were highly correlated between the nonsecretor and the secretor samples ( $r > 0.69$  for 60 % and 0.97 for 26.7 %). Figure 7 (left Panel) shows four examples according to their lectin binding profiles. To the right are four examples of *N*-glycosites that varied in their lectin enrichment levels according to secretor status. In addition to differences in individual *N*-glycosite lectin interactions according to secretor status, this analysis demonstrated interesting differences in glycosylation between the sites along the same peptide backbone, e.g., CF058\_HUMAN (N24 vs. N69;  $r = 0.99$  and 0.24, respectively).

Other examples included sites that were unique among donors (not shown). These included FIBB\_HUMAN@394 (Fibrinogen beta chain), which was observed with 304 counts only in parotid saliva of the non-secretor. Likewise, FIBG\_HUMAN@78 (Fibrinogen gamma chain) was unique to parotid saliva of the nonsecretor, albeit at a much lower relative abundance (18 spectral counts). Both peptides primarily interacted with WGA. The other relatively abundant *N*-glycosites that were specific to the nonsecretor samples included

Together these results suggested that the choice of lectins for separating saliva samples depends on the type of information that is being sought. If maximizing



DMBT1\_HUMAN@1889 (Deleted in malignant brain tumors 1 protein), RNT2\_HUMAN@212 (Ribonuclease T2) and LG3BP\_HUMAN@125 (Galectin-3-binding protein). Of the 25 secretor-unique *N*-glycosites, 8 resided in FCGBP\_HUMAN (IgGfC-binding protein) detected in the AAL and/or WGA-retained fractions of SMSL.

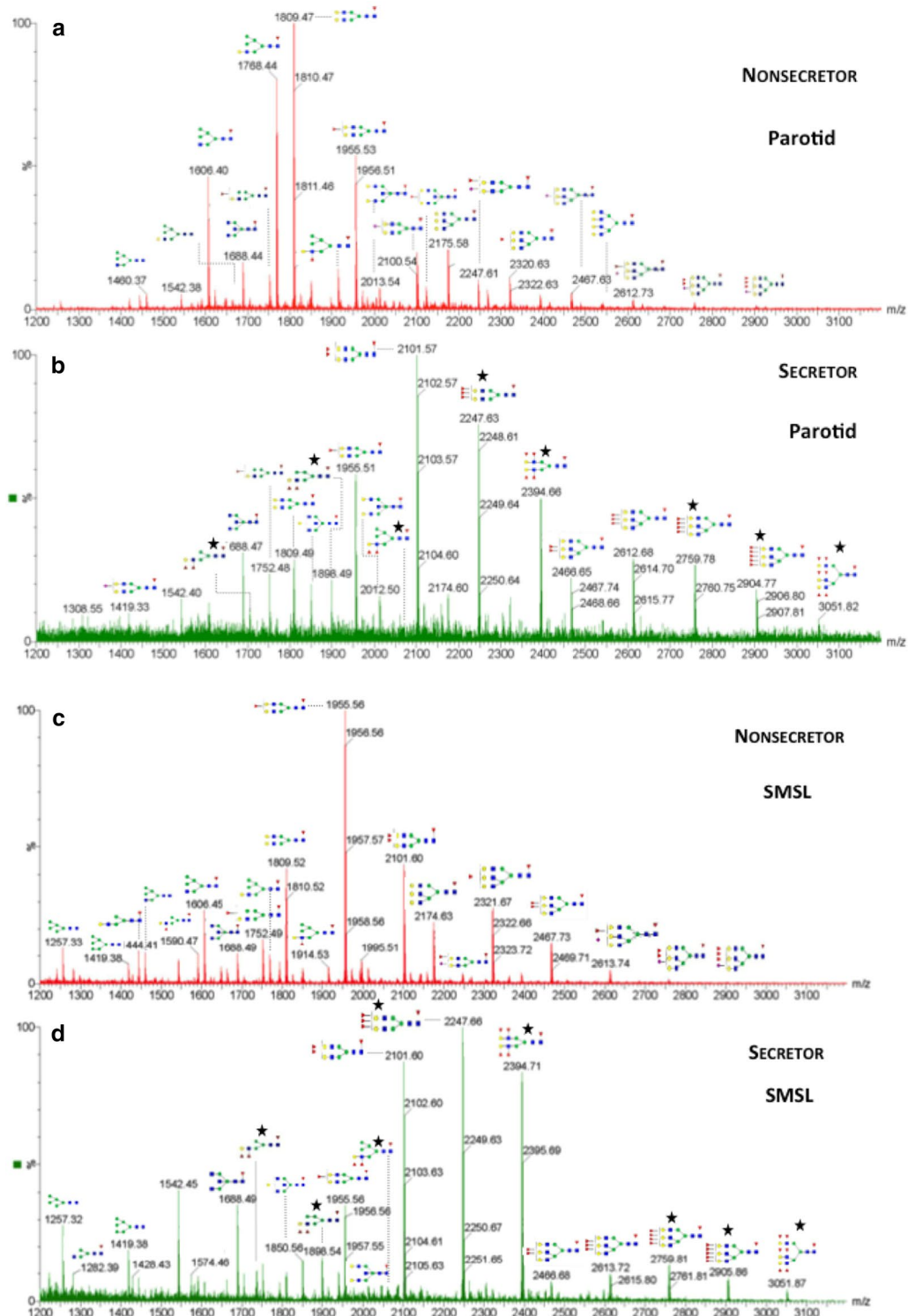
Thus, the results of these experiments show that secretor status might play an important role in determining glycosylation patterns along the peptide backbone. The numerous examples of this phenomenon that we observed suggested that there could be interesting and important biological consequences. Examples include the display of carbohydrate receptors that are recognized by bacterial adhesins, which could play a role in specifying the oral ecology. In this context, it is interesting to note that non-secretors are predisposed to chronic periodontitis [27] and Sjögren's syndrome [28]. Perhaps these "hot spots" of variable glycosylation could be used as diagnostics to predict oral health. Additionally, secretor status correlates with susceptibility to or protection from numerous systemic conditions, including those with an autoimmune etiology such as Crohn's disease (e.g., [29]). Thus, the information that can be gained from analyzing salivary expression of blood group antigens could have translational potential in terms of the clinical practice of medicine and/or dentistry.

#### Oligosaccharide profiling, nonsecretor vs. secretor salivas

Samples from a single secretor and a single nonsecretor were analyzed. Based on donor blood group status, we

theorized that the glycan profiles of the nonsecretor samples were characterized by saccharides with fewer fucosylated species. MS-based analyses confirmed this result. The parotid profiles are shown in Fig. 8a, b (nonsecretor and secretor, respectively); the SMSL profiles are shown in Fig. 8c, d (nonsecretor and secretor, respectively). Parotid saliva contained high mannose as well as hybrid, bi- and tri-antennary structures with higher levels of fucosylation in the secretor as compared to the nonsecretor sample (marked with black stars). In general, the structures we observed (Additional file 7: Table S2) were highly correlated with the core fucosylated oligosaccharides that were described by Guile et al. in their analysis of the human parotid gland glycome [30]. The same general pattern was observed in the SMSL sample with even higher levels of fucosylation. Thus, our structural analyses confirmed that, in saliva, the blood group expression patterns are borne out at the oligosaccharide structural level. We note that other groups have reported differential glycan complexity of secretors compared to nonsecretors [31, 32]. Thomsson et al. detected ABO(H) blood group specific structures that decorated *O*-glycans of salivary MUC5B obtained from nonsecretors demonstrating a higher degree of sialylation compared to secretors [31]. Similarly, an absence of glycan masses corresponding to H-antigen structures in saliva collected from a nonsecretor blood group A individual was observed by Everest-Dass et al. [32].

Glycosylation is a post-translational modification that affects up to 50 % of secreted and cellular



**Fig. 8** Glycan profiles of nonsecretors vs. secretors for parotid (**a** and **b**) and SMSL (**c** and **d**) salivas. N-linked oligosaccharides were released by PNGase F digestion and analyzed by MALDI MS. Putative structures of several glycans that were detected are shown. Overall, both salivas contained high mannose and hybrid bi- and tri-antennary structures. Higher levels of fucosylation, consistent with the presence of blood group H-antigen structure (marked with *black stars*), were observed in the secretor samples. SMSL glycans had a higher level of fucosylation as compared to those observed in parotid saliva. Monoisotopic masses of detected glycans and their structure assignments are listed in Additional file 7: Table S2. Mono-saccharides are indicated using symbols defined by the Consortium for Functional Glycomics [61]

proteins [33]. Oligosaccharide addition adds structural diversity to proteins, and a broad range of glycan functions have been reported including roles in innate immunity [34, 35] and cancer [36]. There have been many studies of the salivary proteome [4, 37], the salivary glycome [32, 38, 39] and glycoproteome [16, 35, 40–45]. However, the majority have not taken secretor status and blood group antigen expression into account. Given the general lack of information about the impact of these variables on inter-individual variations in protein glycosylation, this is a very fruitful future direction for further work focused on saliva and other body fluids.

### Conclusions

We employed a workflow that combined SDS-PAGE protein fractionation, in-gel tryptic digestion, lectin affinity capture of *N*-linked glycopeptides and LC-MS/MS to interrogate, at a global level, human ductal salivary *N*-glycosylation. We considered two variables—blood group (ABO and Lewis) and secretor status. As expected, saliva from a nonsecretor (as compared to a secretor) carried lower level of fucosylation, as confirmed by mass spectrometry analyses of the released *N*-glycans. Intriguingly, a significantly higher number of occupied *N*-glycosites was found in a nonsecretor. Thus, the results of this analysis suggested that blood group and secretor status are important sources of inter-individual variations in salivary glycosylation. The consequences likely include significant effects on the oral ecology.

### Methods

Unless otherwise noted, all reagents were purchased from Thermo-Fisher.

#### Sample collection

The protocols for collecting human saliva were approved by the University of California, San Francisco Committee on Human Research and written informed consent was obtained from all participants. Briefly, the donors thoroughly rinsed their mouths with water prior to sample collection, which was done at the same time in all cases (early afternoon) to mitigate the possible confounding effects of diurnal variation on salivary composition [46]. Salivary flow was stimulated by the application of citric acid to the tongue. Parotid secretions were obtained by using a Lashley cup [12, 47]. SMSL saliva was obtained by using a Block and Brotman collector that fit around the gland openings [4, 48]. The secretions were collected on ice. Immediately thereafter, a proteinase inhibitor cocktail (Pierce) was added. Then the samples were briefly vortexed, divided into 1 mL aliquots and frozen at  $-80^{\circ}\text{C}$ .

#### Analysis of blood group type and secretor status

SMSL and parotid saliva samples from 20 individuals were electrophoretically separated (4–12 % polyacrylamide gels) and transferred to nitrocellulose. The blots were probed with antibodies that recognize the A, B, Le<sup>y</sup> (Abcam), Le<sup>a</sup>, Le<sup>b</sup> (Neomarkers) blood group antigens and sulfated sialyl Le<sup>x</sup> (the L-selectin ligand, MECA-79, BD Biosciences). The method that we used has been published [49].

#### Lectin screening

We used our published method [50]. Each blot consisted of electrophoretically separated samples from 4 individuals. In all, 16 lectins were screened. They are listed in Table I along with the carbohydrate determinants that they recognize.

#### Gel fractionation and digestion

Preparative SDS-PAGE was carried out using a Novex 4–12 % mini-gel and an X-Cell electrophoresis apparatus (Life Technologies). Either 80 (pilot experiment) or 500  $\mu\text{L}$ s of saliva was loaded onto the gel and electrophoresis was performed at 100 V for 2.5 h. Protein bands were visualized by staining with SimplyBlue SafeStain Coomassie G-250 (Invitrogen, Life Technologies, Carlsbad, CA, USA). The entire gel was rastered into 12 (pilot experiment) or 18 slices. Each slice was diced into 1–2 mm pieces, transferred to 2.0-mL Eppendorf tubes and in-gel digestion was performed using a modification of the method described by Shevchenko et al. [51]. Specifically, the first supernatant was removed and gel pieces were incubated with 25 mM ammonium bicarbonate (ABC)/50 % acetonitrile (ACN) to extract peptides. The second supernatant was removed and combined with the first followed by concentration to 25–50  $\mu\text{L}$  in a SpeedVac. Then the samples were brought to a volume of 110  $\mu\text{L}$  with Lectin Buffer A (25 mM Tris, pH 7.4, 50 mM sodium chloride, 10 mM calcium chloride, and 10 mM magnesium chloride; Sigma Aldrich, St. Louis, MO, USA) and passed through 0.2  $\mu\text{m}$  cellulose acetate spin filters (Agilent Technologies, Santa Clara, CA, USA) to remove the remaining gel fragments.

#### Lectin column assembly and glycopeptide enrichment

Columns were produced/packed and lectin affinity enrichment was performed as previously described [13]. Briefly, *Artocarpus integrifolia* (jacalin, JAC), wheat germ agglutinin (WGA), *Aleuria aurantia* lectin (AAL), or *Lycopersicon esculentum* agglutinin (LEA) lectins (Vector Labs, Burlingame, CA) were suspended 10–20 mg/mL in PBS. One hundred mg POROS AL 20  $\mu\text{m}$  beads (Applied Biosystems, Foster City, CA, USA) were washed twice with 1 mL PBS. Then the lectin solution was added,



followed by NaBH<sub>3</sub>CN to a final concentration of 50 mM and the suspension was shaken overnight at room temperature. The beads were washed with 1 mL Tris-HCl pH 7.4, then incubated for 30 min in the same buffer containing 50 mM NaBH<sub>3</sub>CN to block any remaining aldehyde groups. Five washes with 1 M NaCl were performed. Then the beads were packed into 4.6 × 50-mm PEEK HPLC columns by using PBS pumped with a maximum flow rate that was equal to 2000 psi.

Lectin enrichment was accomplished by using a three-step isocratic separation method. One hundred µL of each trypsin digest was applied to a lectin column. First, the column was equilibrated with Lectin Buffer A at a flow rate of 50 µL/min. Second, the samples were loaded and the column was washed for 9 min. Finally, the bound glycopeptides were eluted from the column by using 0.5 M acetic acid for WGA, LEA, and AAL or 100 mM melibiose for jacalin at a flow rate of 500 µL/min for 5.0 min. Then the column was re-equilibrated with Lectin Buffer A for 6 min at a flow rate of 2.5 mL/min. The bound fraction, collected between 9 and 14.2 min, was desalted by using Oasis HLB solid phase extraction (SPE) cartridges. Briefly, an SPE cartridge was wetted with 3 mL of 80 % ACN/0.1 % formic acid (FA) and equilibrated with 3 mL of aqueous 0.1 % FA. Then the sample was loaded and the column washed with 3 mL 0.1 % FA. Next, the peptides were eluted with 1.8 mL 80 % ACN/0.1 % FA into 2-mL eppendorf tubes, neutralized with 200 µL 1 M ABC and vacuum centrifuged to a volume of ~100 µL.

#### **PNGase F deglycosylation**

The eluted *N*-linked glycopeptides were deglycosylated by overnight incubation with 1000 Units of PNGase F (Glycerol-Free, New England Biolabs) as described previously [13]. Then the samples were desalted by using MicroSpin Columns, 5–200 µL loading volume (The Nest Group, Inc.; Southborough, MA, USA). Samples were centrifugally concentrated to dryness and resuspended in 40 µL 0.1 % FA.

#### **Peptide mass spectrometry**

Four microliters from each sample of the original gel bands were separated by using a nanoLC Ultra 2D Plus system (Eksigent/AB Sciex, Foster City, CA, USA) interfaced with a LTQ Orbitrap Velos mass spectrometer (ThermoFisher Scientific). The HPLC was interfaced to the mass spectrometer using a Captive Spray tapered 20 µm I.D. tip (Michrom, Auburn, CA) and applied voltage of 1.5 kV. The peptides were initially loaded onto a guard column (Acclaim PepMap300 C18 300 µm i.d. × 5 mm, 5 µm particle size, 100 Å pore size; ThermoFisher) and washed with the aqueous loading solvent that consisted of 2 % Solvent B (98 % ACN/0.1 % FA)

in Solvent A (2 % ACN/0.1 % FA), flow rate 10 µL/min for 10 min. Then the peptides were separated on a C18 Acclaim PepMap100 column (75 µm i.d. × 150 mm, 3 µm particle size, 100 Å pore size; ThermoFisher Scientific) heated at 48 °C with a column oven. Peptides were eluted at a flow rate of 600 nL/min initially with 2–40 % Solvent B for 60 min then 40–90 % B from 60 to 65 min. The mass spectrometer was calibrated using a solution of caffeine, MRFA, and Ultramark 1621 according to the manufacturer's specifications operating in a data dependent mode. In positive ion mode, MS scans from *m/z* 300–1500 with a full width at half-maximum resolution of 30,000 were acquired in the Orbitrap analyzer. Product ions were generated with a collision cell energy of 35 and an activation *Q* of 0.25. MS/MS scans of the 6 most abundant ions were acquired in the linear ion trap. A mass exclusion time was applied for 30 s with a repeat count of 2 and repeat duration of 20 s and an exclusion list size of 500.

#### **Database searching and *N*-glycosite assignment**

Mass spectrometry data files were processed individually with integrated peak picking using Mascot (Matrix Science, Boston, MA, USA) version 2.2 and Mascot Daemon version 2.2.2 to generate MGF files in an extract\_msn format. Protein identification was accomplished by using the MGF files to search the UniProt Swiss-Prot release 2012\_10 with all human isoforms that included 73,982 protein sequences. ProteinPilot (version 4.5, AB Sciex, Foster City, CA, USA) was used to perform the combined database searches with all bands from a single lectin and secretion. The following ProteinPilot data filters were used: carbamidomethylation of cysteines, Orbi/FT MS (1–3 ppm), LTQ MS/MS as instrument type, and thorough search. Peptides selected for *N*-glycosite analyses had a minimum confidence of 95 %. *N*-glycosites were assigned as previously described [13], which included utilizing the motif NXS/T where X ≠ proline and asparagine was hydrolyzed to aspartic acid. All MS and MS/MS data were manually curated to generate the final *N*-glycosite list. ProteinProspector tools [52] were used to calculate theoretical monoisotopic masses and isotopic envelope distributions for precursors and *m/z* values for product ions [53]. The results of database searches for both donors and secretions were aligned using in-house-generated Java-based software and analyzed in Excel. Estimates of relative abundances employed spectral counts [54] using data combined from all gel slices.

#### **Isolation and enrichment of *N*-linked glycans**

Prior to glycan removal, 100 µL of the SMSL or parotid saliva samples was clarified by centrifugation (3000×*g*) at 4 °C for 20 min. according to the procedure described



by Everest-Dass et al. [55]. The supernatant was mixed with ice-cold acetone (9:1; v/v) and incubated overnight at  $-20^{\circ}\text{C}$ . The next day, the samples were centrifuged to  $13,000\times g$  for 10 min. The supernatant was decanted. Then the precipitate was dried for 10 min at room temperature before dissolving in 200  $\mu\text{L}$  of 3 M urea. *N*-glycan release was performed as described by Papac et al. with slight modifications [56]. Briefly, a Millipore 96-well plate with PVDF membranes (0.45  $\mu\text{m}$ ) was washed ( $3\times$  each solvent) with 100  $\mu\text{L}$  of 70 % ethanol, 100  $\mu\text{L}$  of water and 100  $\mu\text{L}$  of 3 M guanidine hydrochloride (Sigma-Aldrich, Inc.). All washes were performed under mild vacuum. Twenty  $\mu\text{L}$  of each sample was added per well, mixed with 17 mM DTT and incubated at  $37^{\circ}\text{C}$  for 1 h. Then the wells were washed  $3\times$  with 250  $\mu\text{L}$  of milliQ-water and 50  $\mu\text{L}$  of 25 mM iodoacetamide (Sigma-Aldrich, Inc.) was added to each well. After 30 min incubation at room temperature, the wells were washed again with 250  $\mu\text{L}$  of milliQ-water. Membrane blocking was accomplished by room temperature incubation of each well with 100  $\mu\text{L}$  of 1 % PVP-360 (Sigma Aldrich, Inc.) for 30 min. Finally, the membranes were washed  $3\times$  with 250  $\mu\text{L}$  milliQ-water. Then each well was filled with 13  $\mu\text{L}$  milliQ-water, 2  $\mu\text{L}$  of  $10\times$  reaction buffer (0.5 M sodium phosphate, pH 7.5 at  $25^{\circ}\text{C}$ ) and 5  $\mu\text{L}$  of PNGase F enzyme. The plate was incubated overnight at  $37^{\circ}\text{C}$ ; empty wells were filled with water to minimize sample evaporation.

For enrichment, graphite carbon-packed tips (TopTip<sup>TM</sup> Reversed Phase C-18, Glygen) were prepared according to the manufacturer's protocol. Each tip was washed  $3\times$  with 50  $\mu\text{L}$  of 0.05 % TFA in 60 % ACN, followed by  $3\times$  washes with 50  $\mu\text{L}$  of 0.05 % TFA in water. Glycan samples from several wells containing the same saliva type were pooled and applied to a single tip, which was washed  $3\times$  with 50  $\mu\text{L}$  0.05 % TFA. Captured glycans were eluted with 20  $\mu\text{L}$  of 0.05 % TFA in 60 % ACN.

#### MALDI-TOF analysis

The matrix (20 mg/mL) was prepared by dissolving 2',4',6'-trihydroxyacetophenone monohydrate (THAP; Sigma Aldrich, Inc.) in 10 mM ammonium citrate/25 % ACN. Samples were mixed with an equal volume of THAP matrix and 1  $\mu\text{L}$  was deposited onto a 96-well stainless steel MALDI plate (Waters). Orthogonal (o) MALDI ion mobility (IM) TOF MS analyses were performed in positive and in negative ion modes on a Synapt G2 HD mass spectrometer equipped with a TriWave<sup>TM</sup> IM analyzer (Waters) using "sensitivity" settings with a typical resolution of 10,000. Data were analyzed using MassLynx software (Waters). The glycan feature catalog was generated using DriftScope<sup>TM</sup> software to extract glycan ions and the MaxEnt3 algorithm to deconvolute

the data and generate monoisotopic *m/z* values. Glycomod software tools [57, 58] were employed to match the experimental monoisotopic masses to potential oligosaccharide compositions using the following parameters: mass tolerance  $\pm 100$  ppm;  $\text{Na}^+$  (positive mode) or  $[\text{M}-\text{H}]^-$  (negative mode); free/PNGase F released *N*-linked oligosaccharides; and monosaccharide residues defined as hexose, *N*-acetylhexosamine, deoxyhexose and sialic acid. Glycoworkbench tools were employed to generate stick figures of putative glycan structures [59, 60].

#### Additional files

**Additional file 1: Figure S1.** Representative parotid and SMSL salivary protein reactivity with 16 lectins: [*Artocarpus integrifolia* (JAC), Wheat germ agglutinin (WGA), *Aleuria aurantia* lectin (AAL), *Sambucus nigra* (SNA), *Canavalia ensiformis* (Con A), *Lens culinaris* (LCA), *Phaseolus vulgaris* erythroagglutinating (PHA-E), *Galanthus nivalis* (GNA), *Lycopersicon esculentum* agglutinin (Tomato or LEA), *Lotus tetragonolobus* (LTA), *Erythrina christagalli* (ECA), *Datura stramonium* (DSA), *Glycine max* (soybean, SBA), *Arachis hypogaea* (peanut, PNA), *Evonymus europaeus* (EEA), *Vicia villosa* (VVA)]. AAL, JAC, and WGA interacted with the highest number of glycoproteins across a wide molecular weight range. Results are shown for parotid and SMSL saliva from 4 donors.

**Additional file 2: Table S2.** Protein Identifications and Associated *N*-glycosites.

**Additional file 3: Figure S2.** Annotated MS and MS/MS spectra of deglycosylated *N*-linked glycopeptides identified in parotid and SMSL salivas.

**Additional file 4: Figure S3.** Spectral count distribution for *N*-glyco- and non-glycopeptides in the secretor vs. nonsecretor samples. Spectral counting suggested that higher copy numbers of *N*-glycosites were detected in the secretor vs. the nonsecretor sample (Figure S3A). This was despite the fact that the spectral count distribution of non-glycosylated peptides was virtually identical for both samples types (Figure S3B).

**Additional file 5: Figure S4.** *N*-Glycosite distributions binned according to spectral counts. (Left Panel) Similar distributions of donor-common *N*-glycosites were observed across all bins regardless of secretor status. (Right Panel) Lower abundance *N*-glycosites were unique to a particular donor.

**Additional file 6: Figure S5.** Efficiency of *N*-glycopeptide capture by lectin, saliva type and secretor status. Sites that were unique to a particular sample-lectin combination are shown in red and those that were shared among sample-lectin combinations are depicted in blue. (Left Panel) the number of *N*-glycosites and (Right Panel) the relative abundances in terms of spectral counts. Green stars denote  $>2$  SD from the mean. In terms of unique *N*-glycosites detected, AAL captured the largest number in parotid saliva of the nonsecretor. In terms of relative abundances, JAC captured the most copies of unique *N*-glycosites from SMSL saliva.

**Additional file 7: Table S2.** PNGase F-Released *N*-Glycans Detected by MALDI Ion Mobility MS.

#### Abbreviations

Le: Lewis; FUT2: fucosyltransferase 2; gPRP: glycosylated proline rich protein; SMSL: submandibular/sublingual; PNGase F: peptide-*N*-glycosidase F; NXS/T: N = asparagine, X = any amino acid except proline, S = serine, T = threonine; MS: mass spectrometry; MS/MS: tandem mass spectrometry; ESI HPLC-MS/MS: electrospray ionization high-performance liquid chromatography/tandem mass spectrometry; MALDI MS: matrix-assisted laser desorption ionization mass spectrometry; TOF MS: time of flight mass spectrometry; LC-MS/MS: liquid chromatography-tandem mass spectrometry; Orbi/FT MS: Orbitrap/fourier transform mass spectrometry; LTQ: Linear trap Quadrupole; Oasis HLB:

hydrophilic-lipophilic-balanced; AAL: *Aleuria aurantia* Lectin; JAC: *Artocarpus integrifolia*; WGA: wheat germ agglutinin; LEA: *Lycopersicon esculentum* agglutinin; SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis; MUC5B: Mucin 5B; ABC: ammonium bicarbonate; ACN: acetonitrile; SPE: solid phase extraction; FA: formic acid; MRFA: L-methionyl-arginyl-phenylalanyl-alanine acetate H<sub>2</sub>O; MGF: Mascot generic format; TFA: Trifluoroacetic acid; THAP: 2',4',6'-trihydroxyacetophenone monohydrate; S: secretor; NS: nonsecretor; DTT: dithiothreitol; o: Orthogonal; SNA: *Sambucus nigra*; Con A: *Canavalia ensiformis*; LCA: *Lens culinaris*; PHA-E: *Phaseolus vulgaris* erythroagglutinating; GNA: *Galanthus nivalis*; LTA: *Lotus tetragonolobus*; ECA: *Erythrina christagalli*; DSA: *Datura stramonium*; EEA: *Evonymus europaeus*; VVA: *Vicia villosa*.

#### Authors' contributions

MA and MH performed gel fractionation, trypsin digestions, glycopeptide enrichment and deglycosylation. SC isolated the N-linked glycans. AP collected saliva from donors, and carried out the analyses for blood group type, secretor status and lectin screening. CN, AD, KW and SC acquired the mass spectrometry data. MA, MH, KW, SH and HEW interpreted the mass spectrometry data. RN coded the in-house Java-based algorithm used for database search alignments. KW and SH performed a critical review of the manuscript. SF conceived the experimental design and co-wrote the manuscript with HEW. All authors read and approved the manuscript.

#### Author details

<sup>1</sup> Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California San Francisco, San Francisco, CA 94143, USA. <sup>2</sup> Sandler-Moore Mass Spectrometry Core Facility, University of California San Francisco, San Francisco, CA 94143, USA.

#### Acknowledgements

This work was supported by R01DE021041. We gratefully acknowledge the University of California, San Francisco Sandler-Moore Mass Spectrometry Core Facility, which received support from the Sandler Family Foundation, the Gordon and Betty Moore Foundation, National Institutes of Health (NIH)/National Cancer Institute Cancer Center Support Grant P30 CA082103, the Canary Foundation, and the NIH/National Center for Research Resources Shared Instrumentation Grant S10RR029446-01.

#### Competing interests

The authors declare they have no competing interests.

Received: 19 August 2015 Accepted: 25 November 2015

Published online: 30 December 2015

#### References

- Bosch JA. The use of saliva markers in psychobiology: mechanisms and methods. *Monogr Oral Sci*. 2014;24:99–108. doi:10.1159/000358864.
- Helmerhorst EJ, Oppenheim FG. Saliva: a dynamic proteome. *J Dent Res*. 2007;86(8):680–93 (86/8/680 [pii]).
- Amado FM, Ferreira RP, Vitorino R. One decade of salivary proteomics: current approaches and outstanding challenges. *Clin Biochem*. 2013;46(6):506–17. doi:10.1016/j.clinbiochem.2012.10.024.
- Denny P, Hagen FK, Hardt M, Liao L, Yan W, Arellano M, et al. The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions. *J Proteome Res*. 2008;7(5):1994–2006. doi:10.1021/pr700764j.
- Scannapieco FA, Torres G, Levine MJ. Salivary alpha-amylase: role in dental plaque and caries formation. *Critic Rev Oral Biol Med Off Publ Am Assoc Oral Biol*. 1993;4(3–4):301–7.
- de Sousa-Pereira P, Amado F, Abrantes J, Ferreira R, Esteves PJ, Vitorino R. An evolutionary perspective of mammal salivary peptide families: cystatins, histatins, statherin and PRPs. *Arch Oral Biol*. 2013;58(5):451–8. doi:10.1016/j.archoralbio.2012.12.011.
- Melino S, Santone C, Di Nardo P, Sarkar B. Histatins: salivary peptides with copper(II)- and zinc(II)-binding motifs: perspectives for biomedical applications. *FEBS J*. 2014;281(3):657–72. doi:10.1111/febs.12612.
- Henry S, Oriol R, Samuelsson B. Lewis histo-blood group system and associated secretory phenotypes. *Vox Sang*. 1995;69(3):166–82.
- Prakobphol A, Tangemann K, Rosen SD, Hoover CI, Leffler H, Fisher SJ. Separate oligosaccharide determinants mediate interactions of the low-molecular-weight salivary mucin with neutrophils and bacteria. *Biochemistry*. 1999;38(21):6817–25. doi:10.1021/bi990145m.
- Walz A, Odenbreit S, Stuhler K, Wattenberg A, Meyer HE, Mahdavi J, et al. Identification of glycoprotein receptors within the human salivary proteome for the lectin-like BabA and SabA adhesins of *Helicobacter pylori* by fluorescence-based 2-D bacterial overlay. *Proteomics*. 2009;9(6):1582–92. doi:10.1002/pmic.200700808.
- Prakobphol A, Boren T, Ma W, Zhixiang P, Fisher SJ. Highly glycosylated human salivary molecules present oligosaccharides that mediate adhesion of leukocytes and *Helicobacter pylori*. *Biochemistry*. 2005;44(6):2216–24. doi:10.1021/bi0480180.
- Gillece-Castro BL, Prakobphol A, Burlingame AL, Leffler H, Fisher SJ. Structure and bacterial receptor activity of a human salivary proline-rich glycoprotein. *J Biol Chem*. 1991;266(26):17358–68.
- Drake PM, Schilling B, Niles RK, Braten M, Johansen E, Liu H, et al. A lectin affinity workflow targeting glycosite-specific, cancer-related carbohydrate structures in trypsin-digested human plasma. *Anal Biochem*. 2011;408(1):71–85. doi:10.1016/j.ab.2010.08.010.
- Prakobphol A, Thomsson KA, Hansson GC, Rosen SD, Singer MS, Phillips NJ, et al. Human low-molecular-weight salivary mucin expresses the sialyl Lewis x determinant and has L-selectin ligand activity. *Biochemistry*. 1998;37(14):4916–27. doi:10.1021/bi972612a.
- Li X, Jiang J, Zhao X, Wang J, Han H, Zhao Y, et al. N-glycoproteome analysis of the secretome of human metastatic hepatocellular carcinoma cell lines combining hydrazide chemistry, HILIC enrichment and mass spectrometry. *PLoS One*. 2013;8(12):e81921. doi:10.1371/journal.pone.0081921.
- Gonzalez-Begne M, Lu B, Liao L, Xu T, Bedi G, Melvin JE, et al. Characterization of the human submandibular/sublingual saliva glycoproteome using lectin affinity chromatography coupled to multidimensional protein identification technology. *J Proteome Res*. 2011;10(11):5031–46. doi:10.1021/pr200505t.
- Ferreira JA, Daniel-da-Silva AL, Alves RM, Duarte D, Vieira I, Santos LL, et al. Synthesis and optimization of lectin functionalized nanoprobes for the selective recovery of glycoproteins from human body fluids. *Anal Chem*. 2011;83(18):7035–43. doi:10.1021/ac200916j.
- Palmisano G, Melo-Braga MN, Engholm-Keller K, Parker BL, Larsen MR. Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J Proteome Res*. 2012;11(3):1949–57. doi:10.1021/pr2011268.
- Hao P, Ren Y, Datta A, Tam JP, Sze SK. Evaluation of the effect of trypsin digestion buffers on artificial deamidation. *J Proteome Res*. 2015;14(2):1308–14. doi:10.1021/pr500903b.
- Tabb DL, Vega-Montoto L, Rudnick PA, Variyath AM, Ham AJ, Bunk DM, et al. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res*. 2010;9(2):761–76. doi:10.1021/pr9006365.
- Huttenhain R, Surinova S, Ossola R, Sun Z, Campbell D, Cerciello F, et al. N-glycoprotein SRMAtlas: a resource of mass spectrometric assays for N-glycosites enabling consistent and multiplexed protein quantification for clinical applications. *Mol Cell Proteomics*. 2013;12(4):1005–16. doi:10.1074/mcp.O112.026617.
- Frassetto F, Parisotto TM, Peres RC, Marques MR, Line SR, Nobre Dos Santos M. Relationship among salivary carbonic anhydrase VI activity and flow rate, biofilm pH and caries in primary dentition. *Caries Res*. 2012;46(3):194–200. doi:10.1159/000337275.
- Koc Ozturk L, Uluçan K, Akyuz S, Furuncuoğlu H, Bayer H, Yarat A. The investigation of genetic polymorphisms in the carbonic anhydrase VI gene exon 2 and salivary parameters in type 2 diabetic patients and healthy adults. *Mol Biol Report*. 2012;39(5):5677–82. doi:10.1007/s11033-011-1374-1.
- Aidar M, Marques R, Valjakka J, Mononen N, Lehtimäki T, Parkkila S, et al. Effect of genetic polymorphisms in CA6 gene on the expression and catalytic activity of human salivary carbonic anhydrase VI. *Caries Res*. 2013;47(5):414–20. doi:10.1159/000350414.
- Melis M, Atzori E, Cabras S, Zonza A, Calo C, Muroli P, et al. The gustin (CA6) gene polymorphism, rs2274333 (A/G), as a mechanistic link between PROP tasting and fungiform taste papilla density and maintenance. *PLoS One*. 2013;8(9):e74151. doi:10.1371/journal.pone.0074151.

26. Dalziel M, Crispin M, Scanlan CN, Zitzmann N, Dwek RA. Emerging principles for the therapeutic exploitation of glycosylation. *Science*. 2014;343(6166):1235681. doi:10.1126/science.1235681.
27. Tabasum ST, Nayak RP. Salivary blood group antigens and microbial flora. *Int J Dental Hygiene*. 2011;9(2):117–21. doi:10.1111/j.1601-5037.2010.00451.x.
28. Manthorpe R, Staub Nielsen L, Hagen Petersen S, Prause JU, Lewis blood type frequency in patients with primary Sjogren's syndrome. A prospective study including analyses for A1A2BO, Secretor, MNSs, P, Duffy, Kell, Lutheran and rhesus blood groups. *Scand J Rheumatol*. 1985;14(2):159–62.
29. Maroni L, van de Graaf SF, Hohenester SD, OudeElferink RP, Beuers U. Fucosyltransferase 2: a genetic risk factor for primary sclerosing cholangitis and Crohn's disease-A comprehensive review. *Clin Rev Allergy Immunol*. 2014. doi:10.1007/s12016-014-8423-1.
30. Guile GR, Harvey DJ, O'Donnell N, Powell AK, Hunter AP, Zamz S, et al. Identification of highly fucosylated N-linked oligosaccharides from the human parotid gland. *Eur J Biochem/FEBS*. 1998;258(2):623–56.
31. Thomsson KA, Schulz BL, Packer NH, Karlsson NG. MUC5B glycosylation in human saliva reflects blood group and secretor status. *Glycobiology*. 2005;15(8):791–804. doi:10.1093/glycob/cwi059.
32. Everest-Dass AV, Jin D, Thaysen-Andersen M, Nevalainen H, Kolarich D, Packer NH. Comparative structural analysis of the glycosylation of salivary and buccal cell proteins: innate protection against infection by *Candida albicans*. *Glycobiology*. 2012;22(11):1465–79. doi:10.1093/glycob/cws112.
33. Khoury GA, Baliban RC, Floudas CA. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific Reports*. 2011. doi:10.1038/srep00090.
34. Prakobphol A, Xu F, Hoang VM, Larsson T, Bergstrom J, Johansson I, et al. Salivary agglutinin, which binds *Streptococcus mutans* and *Helicobacter pylori*, is the lung scavenger receptor cysteine-rich protein gp-340. *J Biol Chem*. 2000;275(51):39860–6. doi:10.1074/jbc.M006928200.
35. Qin Y, Zhong Y, Zhu M, Dang L, Yu H, Chen Z, et al. Age- and sex-associated differences in the glycoproteins of human salivary glycoproteins and their roles against influenza A virus. *J Proteome Res*. 2013. doi:10.1021/pr400096v.
36. Drake PM, Cho W, Li B, Prakobphol A, Johansen E, Anderson NL, et al. Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin Chem*. 2010;56(2):223–36. doi:10.1373/clinchem.2009.136333.
37. Bandhakavi S, Stone MD, Onsongo G, Van Riper SK, Griffin TJ. A dynamic range compression and three-dimensional peptide fractionation analysis platform expands proteome coverage and the diagnostic potential of whole saliva. *J Proteome Res*. 2009;8(12):5590–600. doi:10.1021/pr900675w.
38. Everest-Dass AV, Abrahams JL, Kolarich D, Packer NH, Campbell MP. Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. *J Am Soc Mass Spectrom*. 2013. doi:10.1007/s13361-013-0610-4.
39. Sondej M, Denny PA, Xie Y, Ramachandran P, Si Y, Takashima J, et al. Glyco-profiling of the human salivary proteome. *Clin Proteomics*. 2009;5(1):52–68. doi:10.1007/s12014-008-9021-0.
40. Bandhakavi S, Van Riper SK, Tawfik PN, Stone MD, Haddad T, Rhodus NL, et al. Hexapeptide libraries for enhanced protein PTM identification and relative abundance profiling in whole human saliva. *J Proteome Res*. 2011;10(3):1052–61. doi:10.1021/pr100857t.
41. Ramachandran P, Boonthueung P, Pang E, Yan W, Wong DT, Loo JA. Comparison of N-linked glycoproteins in human whole saliva, parotid, submandibular, and sublingual glandular secretions identified using hydrazide chemistry and mass spectrometry. *Clin Proteomics*. 2008;4(3–4):80–104. doi:10.1007/s12014-008-9005-0.
42. Larsen MR, Jensen SS, Jakobsen LA, Heegaard NH. Exploring the sialome using titanium dioxide chromatography and mass spectrometry. *Mol Cell Proteomics*. 2007;6(10):1778–87. doi:10.1074/mcp.M700086-MCP200.
43. Sun S, Zhao F, Wang Q, Zhong Y, Cai T, Wu P, et al. Analysis of age and gender associated N-glycoproteome in human whole saliva. *Clin Proteomics*. 2014;11(1):25. doi:10.1186/1559-0275-11-25.
44. Vitorino R, Alves R, Barros A, Caseiro A, Ferreira R, Lobo MC, et al. Finding new posttranslational modifications in salivary proline-rich proteins. *Proteomics*. 2010;10(20):3732–42. doi:10.1002/pmic.201000261.
45. Xu Y, Bailey UM, Punyadeera C, Schulz BL. Identification of salivary N-glycoproteins and measurement of glycosylation site occupancy by boronate glycoprotein enrichment and liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid Commun Mass Spectro RCM*. 2014;28(5):471–82. doi:10.1002/rcm.6806.
46. Hardt M, Witkowska HE, Webb S, Thomas LR, Dixon SE, Hall SC, et al. Assessing the effects of diurnal variation on the composition of human parotid saliva: quantitative analysis of native peptides using iTRAQ reagents. *Anal Chem*. 2005;77(15):4947–54. doi:10.1021/ac050161r.
47. Lashley K. Reflex secretion of the human parotid gland. *J Exp Psychol*. 1916;1:461–93.
48. Block P, Brotman S. A method of submaxillary saliva collection without cannulization. *NY State Dent J*. 1962;28:116–8.
49. Prakobphol A, Leffler H, Fisher SJ. The high-molecular-weight human mucin is the primary salivary carrier of ABH, Le(a), and Le(b) blood group antigens. *Critic Rev Oral Biol Med Off Publ Am Assoc Oral Biol*. 1993;4(3–4):325–33.
50. Drake PM, Schilling B, Niles RK, Prakobphol A, Li B, Jung K, et al. Lectin chromatography/mass spectrometry discovery workflow identifies putative biomarkers of aggressive breast cancers. *J Proteome Res*. 2012;11(4):2508–20. doi:10.1021/pr201206v.
51. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc*. 2006;1(6):2856–60. doi:10.1038/nprot.2006.468.
52. Protein Prospector. <http://prospector.ucsf.edu>. Accessed 20 May 2014.
53. Chalkley RJ, Hansen KC, Baldwin MA. Bioinformatic methods to exploit mass spectrometric data for proteomic applications. *Methods Enzymol*. 2005;402:289–312. doi:10.1016/S0076-6879(05)02009-4.
54. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*. 2004;76(14):4193–201. doi:10.1021/ac0498563.
55. Everest-Dass AV, Abrahams JL, Kolarich D, Packer NH, Campbell MP. Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. *J Am Soc Mass Spectrom*. 2013;24(6):895–906. doi:10.1007/s13361-013-0610-4.
56. Papac DI, Briggs JB, Chin ET, Jones AJ. A high-throughput microscale method to release N-linked oligosaccharides from glycoproteins for matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis. *Glycobiology*. 1998;8(5):445–54.
57. Cooper CA, Gasteiger E, Packer NH. GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics*. 2001;1(2):340–9. doi:10.1002/1615-9861(200102)1:2<340:AID-PROT340>3.0.CO;2-B.
58. GlycoMod. <http://expasy.org/glycomod>. Accessed 29 Nov 2014.
59. Ceroni A, Dell A, Haslam SM. The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. *Source Code Biol Med*. 2007;2:3. doi:10.1186/1751-0473-2-3.
60. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res*. 2008;7(4):1650–9. doi:10.1021/pr7008252.
61. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Marth JD, et al. Symbol nomenclature for glycan representation. *Proteomics*. 2009;9(24):5398–9. doi:10.1002/pmic.200900708.
62. Ahmed H, Chatterjee BP. Further characterization and immunochemical studies on the carbohydrate specificity of jackfruit (*Artocarpus integrifolia*) lectin. *J Biol Chem*. 1989;264(16):9365–72.
63. Kaladas PM, Kabat EA, Iglesias JL, Lis H, Sharon N. Immunochemical studies on the combining site of the D-galactose/N-acetyl-D-galactosamine specific lectin from *Erythrina cristagalli* seeds. *Arch Biochem Biophys*. 1982;217(2):624–37.
64. Swamy MJ, Gupta D, Mahanta SK, Surolia A. Further characterization of the saccharide specificity of peanut (*Arachis hypogaea*) agglutinin. *Carbohydr Res*. 1991;213:59–67.
65. Piller V, Piller F, Cartron JP. Comparison of the carbohydrate-binding specificities of seven N-acetyl-D-galactosamine-recognizing lectins. *Eur J Biochem/FEBS*. 1990;191(2):461–6.
66. Shen ZM, Shi WX, Sun C, Yang JT. Conformation and activity of mannose- and N-acetylgalactosamine-specific lectins from *Vicia villosa* seeds. *Biochimie*. 1993;75(11):949–54.
67. Crowley JF, Goldstein IJ, Arnarp J, Lonngren J. Carbohydrate binding studies on the lectin from *Datura stramonium* seeds. *Arch Biochem Biophys*. 1984;231(2):524–33.

68. Nachbar MS, Oppenheim JD, Thomas JO. Lectins in the US Diet. isolation and characterization of a lectin from the tomato (*Lycopersicon esculentum*). *J Biol Chem*. 1980;255(5):2056–61.
69. Allen AK, Neuberger A, Sharon N. The purification, composition and specificity of wheat-germ agglutinin. *Biochem J*. 1973;131(1):155–62.
70. Peters BP, Ebisu S, Goldstein IJ, Flashner M. Interaction of wheat germ agglutinin with sialic acid. *Biochemistry*. 1979;18(24):5505–11.
71. Yamamoto K, Tsuji T, Matsumoto I, Osawa T. Structural requirements for the binding of oligosaccharides and glycopeptides to immobilized wheat germ agglutinin. *Biochemistry*. 1981;20(20):5894–9.
72. Debray H, Decout D, Strecker G, Spik G, Montreuil J. Specificity of twelve lectins towards oligosaccharides and glycopeptides related to N-glycosylproteins. *Eur J Biochem/FEBS*. 1981;117(1):41–55.
73. Shibuya N, Goldstein IJ, Van Damme EJ, Peumans WJ. Binding properties of a mannose-specific lectin from the snowdrop (*Galanthus nivalis*) bulb. *J Biol Chem*. 1988;263(2):728–34.
74. Debray H, Montreuil J. Aleuria aurantia agglutinin. A new isolation procedure and further study of its specificity towards various glycopeptides and oligosaccharides. *Carbohydr Res*. 1989;185(1):15–26.
75. Pereira ME, Kabat EA. Blood group specificity of the lectin from *Lotus tetragonolobus*. *Ann N Y Acad Sci*. 1974;234:301–5.
76. Shibuya N, Goldstein IJ, Broekaert WF, Nsimba-Lubaki M, Peeters B, Peumans WJ. The elderberry (*Sambucus nigra* L.) bark lectin recognizes the Neu5Ac(alpha 2-6)Gal/GalNAc sequence. *J Biol Chem*. 1987;262(4):1596–601.
77. Petryniak J, Goldstein IJ. Evonymus europaea lectin. *Methods Enzymol*. 1987;138:552–61.
78. Cummings RD, Kornfeld S. Characterization of the structural determinants required for the high affinity interaction of asparagine-linked oligosaccharides with immobilized *Phaseolus vulgaris* leucoagglutinating and erythroagglutinating lectins. *J Biol Chem*. 1982;257(19):11230–4.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

