Functional specialization of human salivary glands and origins of proteins intrinsic to human saliva

Authors: Marie Saitou[1], Eliza Gaylord[2], Erica Xu[1,3], Lubov Neznanova[4], Sara Nathan[2], Anissa Grawe[2], Jolie Chang[5], William Ryan[5], Stefan Ruhl[4], Sarah M. Knox[2], Omer Gokcumen[1]

Affiliations:

1. Department of Biological Sciences, University at Buffalo, The State University of New York
2. Program in Craniofacial Biology, Department of Cell and Tissue Biology, School of Dentistry, University of California, San Francisco
3. Currently at Weill-Cornell Medical College, Physiology and Biophysics Department.
4. Department of Oral Biology, School of Dental Medicine, University at Buffalo, The State University of New York
5. Department of Otolaryngology, School of Medicine, University of California, San Francisco

Correspondence: OG.: omergokc@buffalo.edu, SK: Sarah.Knox@ucsf.edu, SR: shruhl@buffalo.edu

ABSTRACT

Salivary proteins facilitate food perception and digestion, maintain the integrity of the mineralized tooth and oral epithelial surfaces, and shield the oro-digestive tract from environmental hazards and invading pathogens. Saliva, as one of the easiest to collect body fluids, also serves in diagnostic applications, with its proteins providing a window to body health. However, despite the availability of the human saliva proteome, the origins of individual proteins remain unclear. To bridge this gap, we analyzed the transcriptomes of 27 tissue samples derived from the three major types of human adult and fetal salivary glands and integrated these data with the saliva proteome and the proteomes and transcriptomes of 28+ other human organs, with tissue expression confirmed by 3D microscopy. Using these tools, we have linked saliva proteins to their source for the first time, an outcome with significant implications for basic research and diagnostic applications. Furthermore, our study represents the first comparative transcriptomic analysis of human adult and fetal exocrine organs, providing evidence that functional specialization occurs late in salivary gland development, and is driven mainly by the transcription of genes encoding secreted saliva proteins. Moreover, we found that dosage of abundant saliva proteins secreted by the salivary glands is primarily regulated at the transcriptional level, and that secreted proteins can be synthesized by distinct subsets of serous acinar cells, revealing hitherto unrecognized heterogeneity in the acinar cell lineage. Our results reveal the functional underpinnings of these secretory organs, paving the way for future investigations into saliva biology and pathology.

SIGNIFICANCE STATEMENT

Saliva facilitates the perception and digestion of food while regulating the homeostasis of microbes in the mouth and their transduction to the digestive tract. Proteins in saliva fluid are responsible for these diverse and important functions. To understand the glandular origins of proteins in saliva, we performed the hitherto most comprehensive investigation of gene expression in healthy human salivary glands. Our results revealed that each mature salivary gland type contributes a distinct repertoire of proteins to saliva. We provide evidence that these unique, gland-specific secreted protein profiles emerge late in gland development. Our results suggest that the origins of salivary proteins in whole mixed saliva can be parsed out only through an integrative analysis of both protein and RNA levels.

INTRODUCTION

The mouth is the main portal of entry to the gastrointestinal tract with saliva serving as the quintessential gatekeeper (1, 2). Saliva exerts a multitude of important functions in the oral cavity and beyond, that depend upon its repertoire of proteins. These functions include the breakdown of dietary starch by the salivary enzyme amylase (3), the provision of calcium phosphate to maintain mineralization of tooth enamel (4), and host defense against pathogenic microorganisms (5, 6). Furthermore, saliva possesses physicochemical properties to keep the oral cavity moist and well lubricated, which are also provided by saliva proteins, especially mucins (7, 8). Thus, genetic and dosage variation in the saliva proteome will have important biomedical consequences (9–12). At the extreme, malfunctioning of the salivary glands, for instance, due to radiation treatment of head and neck cancer, or by the relatively common autoimmune disease, Sjögren's syndrome, can lead to major disruptions in protein homeostasis within saliva causing severe complications in oral health that debilitate patient quality of life (13, 14). Therefore, understanding how the composition of the saliva proteome is attained and regulated remains an important avenue of inquiry.

A major complication in studying saliva and harnessing its proteome profile for diagnostic applications is the fact that this oral biofluid is a complex mixture derived from multiple sources. Saliva gets predominantly synthesized and secreted by three major pairs of anatomically and histologically distinct craniofacial secretory organs: the parotid, submandibular, and sublingual salivary glands (**Figure 1A**). Each of these glands produces a characteristic spectrum of salivary proteins that are thought to be predominantly based on their unique composition of mucous and serous acinar cells. In addition to these 'intrinsic' (*i.e.*, synthesized by the glands) proteins that sustain most of the major functions of the saliva, saliva also contains 'extrinsic' proteins that originate in other organs and systems such as the bloodstream and cells lining the oral integuments (1, 15, 16). Indeed, proteins can diffuse into saliva from other organs and organ systems via crossing capillary barriers, interstitial spaces, and the membranes of acinar and ductal cells. Whole mouth saliva also contains a plethora of proteins derived from the oral microbiome (more than 700 different species of microorganisms cohabit saliva), adding to the functional diversity of human saliva (17, 18). A multitude of studies has been conducted to catalog salivary proteins and distinguish intrinsic from extrinsic protein components (19, 20). These studies include proteomic analyses comparing whole saliva or saliva collected from the ducts of the parotid or submandibular/sublingual (21, 22) to body fluids such as plasma, urine, cerebrospinal fluid, and amniotic fluid (23). However, discrepancies between published data sets, likely due to variation in collection procedures, sample integrity, storage conditions, sample size, and analytical methods (12, 24, 25), have impaired the establishment of a robust catalog of proteins – an outcome that has severely hampered the use of saliva as a physiological and pathophysiological research tool and as a reliable fluid for disease diagnosis.

To address this gap in knowledge, a major effort has been made in compiling a systematic knowledge database of proteins in human saliva, the recently established Human Salivary Proteome Wiki (https://salivaryproteome.nidcr.nih.gov/), a community-based, collaborative web portal that harbors information on salivary proteins identified by high-throughput proteomic technologies including mass spectrometry. The accumulating data on the protein content of saliva led to several intriguing questions in regard to human saliva: *What are the relative contributions of different salivary glands to the composition of the salivary proteome? To what degree are the protein levels in the saliva regulated at RNA or post-translational levels? How different are the secretomes of the adult human salivary glands and what genes drive these differences? Finally, when do those differences in the secretomes between glands become apparent during fetal development?* To help answer these questions, we have constructed the first comprehensive dataset of the transcriptomes of the major human salivary glands types and integrated these data with the mass spectrometric proteomic analyses from the Human Salivary Proteome Wiki and Human Protein Atlas (26). We sequenced total RNA of 27 salivary gland samples collected from adult and fetal human submandibular, sublingual, and parotid glands and analyzed transcriptomes across both gland type and developmental stage. By integrating these RNA data with the proteomic profiles, we have newly defined the mature salivary secretome, allowing us to not only distinguish between intrinsic and extrinsic saliva proteins but also to reveal common and gland-specific secreted molecules of the major glands. Importantly, we provide here the first evidence that transcriptomic differences between the major mature glands emerge late during fetal development and are driven predominantly by gene transcripts encoding for distinct secreted proteins.

RESULTS AND DISCUSSION

Distinct transcript profiles can be found in adult but not in fetal human salivary glands

To comprehensively identify gene expression differences between the three major salivary glands, we conducted a transcriptome analysis of multiple healthy adult and fetal sublingual (SL, sl), submandibular (SM, sm), and parotid (PAR, par) gland tissues (**Table S1, Figure 1A**). We comparatively analyzed the expression levels of 167,278 transcripts consolidated into 40,882 coding and noncoding genes using a DeSeq2 pipeline (27) and gene-transcript information on Ensembl (http://asia.ensembl.org/index.html) (**Table S2, Supplementary Methods**). We were able to clearly differentiate mature glands from fetal glands without any *a priori* hypothesis based only on the first principal component of transcriptome data (**Figure 1B**). The second principal component of the transcriptome data evidently separated the mature gland types. However, the same analysis could not differentiate among fetal gland types (**Figure 1B**). We verified these results using a hierarchical clustering analysis (**Figure 1C**), where the transcripts of the mature glands significantly clustered into a major branch distinct from that of the fetal glands. Moreover, we found that the transcripts of the mature glands further branched according to their glandular origin, whereas those of the fetal glands did not. We were able to quantify these observations using a Pearson correlation analysis (**Figure S1**). Despite their transcriptional distinctiveness, the three major mature gland types still share a high number of genes that are similarly up or downregulated in transcription compared to their fetal status, suggesting overall functional congruity as mature secretory exocrine organs (**Figure 1D**). Collectively, our results suggest that the three major salivary glands acquire specific gene expression profiles at a relatively late stage of embryogenesis.

Genes encoding secreted proteins account for the majority of transcriptomic differences among mature salivary gland types as well as between salivary glands and other organ systems

To identify genes that are expressed specifically in salivary glands, we compared transcript levels in salivary glands to those of 54 other tissues in the GTEx portal, including other epithelial organs that secrete fluids such as the pancreas, mammary tissue (non-lactating), and intestine (28, 29) **(Figure S2**). This analysis showed that some of the most abundantly expressed genes in mature salivary glands showed little or no expression in other tissues. These transcriptome-wide observations are concordant with our manual inspection of the Human Protein Atlas database (26). The majority of these genes encode proteins that were found in highest concentrations in saliva (1), among them most prominently salivary amylase (*AMY1*), salivary histatins (*HTN1, HTN3*), statherin (*STATH*), low-molecular-weight salivary mucin (*MUC7*), salivary proline-rich proteins (*PRB1-4*, *PRH2*), and salivary cystatins (*CST1, CST4*) (**Figure 2A**). Using gene and protein annotations available through the Human Protein Atlas (30), we confirmed that 20 of the 30 highest expressed genes in adult glands encode secreted proteins that are specifically enriched in salivary glands as compared to other organs (**Figure 2A**, **Figure S2**). Our analysis also found that these 20 genes were expressed approximately 100-fold higher in mature salivary glands than in fetal glands (**Table S2**). Furthermore, the proportion of total gene transcripts that encode secreted proteins was significantly higher in mature glands than in their fetal counterparts ($p < 0.05$, Mann-Whitney test, **Figure S3**). Consistent with a previous study demonstrating that the secretome (including hormones, cytokines, growth factors, *etc.*) holds the largest fraction of the organ-specific proteome (26, 30), our data suggest that the major transcriptomic differences between adult salivary glands and their fetal counterparts, as well as other similar organs, is their specific repertoire and abundance of secreted proteins.

To determine which transcripts account for the differences between the mature gland types, we conducted a comparative analysis between glandular transcripts (**Figure 2**). We found that the differences among mature gland types were mainly due to genes that code for secreted proteins. For example, *AMY1*, salivary histatins (*e.g.*, *HTN1, HTN3*), and salivary proline-rich proteins (*e.g.*, *PRH1, PRH2*) were expressed at highest levels in both parotid and submandibular gland but much less or not at all in sublingual gland tissue. Transcripts for *MUC7* were virtually absent in parotid gland tissue but were expressed at highest levels in the sublingual gland and to a lesser extent in the submandibular gland tissue, whereas salivary cystatins (*e.g.*, *CST1*, *CST2*) were found to be highest expressed in submandibular gland tissue (**Figures 2A and 2B**). A remarkably small number of genes showed significant differences in expression among fetal glands (**Figure 2B, lower panel**). One interesting observation was that *MUC7* was abundantly expressed in all fetal gland tissues (**Figure 2A, middle row**) suggesting an as of yet unknown functional importance of *MUC7* early in gland development that is lost later during the parotid gland maturation. The gene encoding the enzyme lactoperoxidase (*LPO*) was also found to be expressed during gland development. However, in contrast to *MUC7*, it showed a similar expression pattern at both fetal and adult stages, being enriched in the parotid and submandibular glands with low expression in the sublingual (**Figure 2B**). LPO displays anti-microbial activity in saliva and other exocrine body fluids such as milk and tear fluid (31), but a role during gland development has not been described.

Functional enrichment analysis indicates that major salivary glands act as immune hubs and that PAR and SM are highly metabolically active

To investigate how transcriptomic differences between the major salivary glands may impact gland function, we conducted gene ontology (GO) analysis (32, 33) on genes expressed in major salivary glands **(Table S3)**. Specifically, we searched for functional enrichment among genes that are up- or down-regulated in mature glands as compared to fetal glands (**Figure 1D**). Genes that were predominantly expressed in the fetal tissues were found to be significantly enriched for categories linked to growth and development, including cell cycle processes, cell division, and other fundamental cellular processes (**Figure 3)**. Mature salivary gland tissues were enriched for genes involved in innate (*e.g.*, complement cascade related genes such as *CD55* and *TNFSF13*) and adaptive immunity (*e.g.*, genes involved in B cell regulation including *SLC39A10* and *BCL6*) and transport (*e.g.*, calcium-activated chloride channel *ANO1* and sodium/iodide symporter *SLC5A5*). This upregulation of genes involved in responses to foreign antigens is consistent with the adult immune system being highly stimulatory compared to the fetal immune system that was shown to be immunosuppressive (34). Mature salivary glands are also well known to contain multiple immune cell types such as plasma B cells that produce IgA antibody, and dendritic cells that present antigens (35).

In addition to age-based differences in immune gene expression, the sets of upregulated immune-related genes also differed between the adult gland types. For instance, we found distinct complement cascades and immunoglobulin production pathways (*e.g.*, IGHV1-58 and C6) that were shared by PAR and SM but were different from those shared by the SM and SL (**Figure 3**). Furthermore, the SM was unique in expressing a large set of immune-system-related genes (~600, **Figure 3**), suggesting that the SM responds in a distinct manner to pathogenic cues. Although it is not known whether these differences in immune response gene expression between glands result in tissue-specific outcomes in humans, glandular inflammatory conditions (sialadenitis) show a predilection for certain gland types. For example, Heerfordt syndrome causes parotitis (36), whereas chronic sclerosing sialadenitis predominantly affects the submandibular glands (37).

We also found that the SM and PAR are enriched for metabolic and biosynthetic processes compared to the SL, including nucleotide, glucose and fatty acid metabolism (**Figure 3**). These outcomes are consistent with PAR and SM producing and secreting the highest quantities of salivary fluid (~90% of total saliva when at rest) and the highest number of secreted proteins. This high metabolic activity is further supported by F-18 fluorodeoxy-D-glucose (FDG) positron emission tomography (PET) based imaging. This technique utilizes metabolic trapping of radioactive glucose molecules to generate metabolic maps of highly glycolytic tissues and pathologies such as cancer (38) and routinely visualizes healthy PAR or SM tissue (39, 40). Although it remains uncertain whether the functional metabolic output in humans is gland-specific, rodent studies have demonstrated specific differences in metabolic levels among the salivary gland tissues, such as increased Krebs cycle activity for the resting PAR compared to the SM (41)In either case, to gain a mature, functional state, the salivary gland likely tunes metabolic processes for the efficient production and secretion of proteins.

A number of gene sets were also found to be downregulated in at least one of the mature salivary glands compared to the fetal organs (**Figure 3**). These included genes involved in gene transcription and chromatin organization (*e.g.*, *TOP2A* and *SMARCAD1*), extracellular matrix organization, cell division (*e.g.*, *CDK1*, *CDK6, CDC20*) and developmental processes (*e.g.*, epithelial tube formation). Given the SM and PAR and to a lesser extent the SL generate a high abundance of secreted proteins (**Figure 2, Figure S3**), it comes as no surprise that to maintain homeostasis these adult organs reduce the expression of other energy-expensive proteins such as those involved in cell division. Indeed, a recent study demonstrated that salivary glands, along with two other highly secretory tissues, pancreas, and liver, have a greater propensity than other organs to suppress the expression and translation of other expensive host cell proteins (42).

Together, our functional enrichment analyses suggest that the mature salivary glands primarily function as protein secretion and immunity hubs. While all major salivary glands apparently share core functions in immunity, individual gland types differ in specific immune-related pathways and in their ability to secrete specific sets of proteins. In future studies, it will be important to investigate how functional salivary gland tissue differentiation adapted to different environmental conditions, food resources, and pathogen challenges during primate evolution and among geographically and culturally distinct human populations (10, 11, 43–45).

Gland-specific expression of abundant salivary secreted proteins is predominantly regulated at the transcriptome level

By combining advanced quantitative technologies such as RNAseq and mass spectrometry it has become possible to dissect observed protein level variance into contributions from transcriptional, translational, and post-translational processes (46). Thus, to determine if saliva protein abundance was regulated at the transcriptomic, translational or post-translational level, we compared transcript levels in each glandular tissue type with protein abundances in the corresponding glandular ductal secretions (**Figure 4, Table S2**). For that, we first compared our RNAseq dataset to protein expression data available through the Human Salivary Protein Wiki, a saliva-centric, curated mass spectrometry-based proteome database that recently became available (https://salivaryproteome.nidcr.nih.gov/). This database provides quantitative data on abundances of proteins in whole mixed saliva as well as in ductal secretions obtained from parotid and submandibular/sublingual glands. The proteomic profiling by mass spectrometry was able to accurately quantify 2,911 proteins (**Table S2**). We found at least 10 normalized RNAseq reads per gene in at least one of the salivary glands for 85%, 81%, and 80% of the top

100 proteins expressed in whole saliva, parotid, and submandibular/sublingual secretions, respectively.

Overall, we could not find a genome-wide correlation between transcript levels in either of the glands and their corresponding protein levels in the respective glandular secretions ($R^2 < 0.1$). Similarly, there was no direct correlation between glandular transcriptomic abundances and protein abundances observed in the whole saliva ($R^2 < 0.1$). However, we found that highly abundant proteins in ductal saliva are also highly expressed at the RNA level in the corresponding glandular tissues (**Figure 4**). This obvious correlation between abundant protein and transcript suggests that most prominent secreted proteins in saliva are regulated primarily at the transcriptome level. Nevertheless, we noted that certain secreted proteins that were abundantly detected at the mRNA level in glandular tissues and at the protein level in ductal saliva, such as STATH, LYZ, MUC7, and HTN1, among others, were detectable by mass spectrometric analysis at much lower amounts or not at all in whole mouth saliva **(Figure 4 and Figure S4)**.

This reduction or loss can be due to these proteins being proteolytically degraded once exposed to the mouth environment, or through adsorption to oral surfaces after secretion from salivary glands. Indeed, multiple studies have demonstrated STATH, LYZ and HTN1 to be selectively adsorbed from saliva onto the enamel surface (acquired pellicle, (47–49)). It is also possible that mass spectrometric analysis could not detect certain proteins in saliva due to, for example, dense glycosylation that prevents trypsin cleavage, or other molecular features that impede identification of unique peptides in the mass spectrometer (43, 50, 51). In that regard, a recent mass spectrometry-based proteomic analysis of healthy parotid salivary gland has revealed multiple proteins including HTN1 and LYZ to be highly expressed in the glandular tissue (52), thereby supporting our prediction of protein loss after secretion from the gland.

To find out more about the genes that did not correlate in their expression levels with the abundance of proteins they encode, we conducted functional enrichment analysis (**Table S3**). We found that genes that were abundantly expressed in glandular tissues but could not be detected at the protein level in salivary secretions, were enriched in functions related to intracellular housekeeping processes, as well as those typifying exocrine tissues including vesicle-mediated transport (*e.g.*, *MCFD2*), regulated exocytosis (*e.g.*, *TCN1*), and cell secretion (*e.g.*, *FURIN*). In contrast, genes encoding proteins that were identified in saliva, but not at the RNA level in glandular tissues, were enriched in functions characteristic of epithelial cells, including keratinization and cornification (*e.g.*, *KRT1, SPRR1A*). One particular protein that is abundantly found in saliva (top 10%) but is not highly expressed (bottom 10%) at the RNA level in glandular tissues, is albumin (**Table S2**). This finding proves that the majority of albumin in the whole saliva is not derived from salivary glands but rather diffuses into whole mouth fluid from the plasma (it accounts for more than 50% of all plasma proteins) through epithelial leakage and gingival crevicular fluid as had been suggested earlier (53).

In order to further define the origins of proteins in saliva, we next compared the most abundant proteins, ranked according to peptide abundance in the human salivary proteome and according to transcript abundance by our RNAseq analysis, to the publically available mass spectrometry-based proteomes of 29 healthy human organ tissues from the Human Protein Atlas project (52). These 29 organs included the parotid gland as well as other organs that secrete fluids such as the pancreas, prostate, and gallbladder (**Figure S5**) (52). Through this comparative analysis, we delineated 14 out of the 50 secreted saliva proteins to be highly enriched in salivary glands and saliva as well as at least 5 other proteins that besides the salivary gland were found to be highly expressed in one or two of the other organs tested (**Figure S5**). These findings suggesting salivary glands as the organs of origin for these proteins in saliva are also supported by our comparative analysis of salivary gland mRNA to the 54 tissues/organs within the GTEx transcriptomes (**Figure S2**). However, as some of these proteins, including MUC7, LPO and PIP are known to be present in other body fluids (not included in the GTEx transcriptome database) such as tear fluid and respiratory mucus (54, 55), these proteins are not exclusive to the salivary glands and are likely markers of multiple human exocrine organs (the pancreas being the exception). Other proteins, such as CST2, CST5, ZG16B and SMR3B also showed little to no protein or transcript expression in the other tissues/organs, including secretory mammary gland, pituitary, prostate, pancreas, and lung, and have only been detected in tear fluid (55).

In addition to our analysis of abundant secreted proteins in saliva, we also examined secreted proteins that displayed high abundance in salivary glands at the transcript level but low abundance in saliva at the protein level (**Figure 2A**). When compared to the Human Protein Atlas proteome or the GTEx transcriptome database, three proteins, STATH, HTN1 and HTN3, showed little protein or gene expression in the organs analyzed, with only STATH being detected in another body fluid (sweat gland secretion), suggesting these proteins are highly specific to the salivary glands and saliva. However, further studies are needed to determine whether they are exclusive to the salivary glands or are expressed by other glandular systems (e.g., lacrimal glands, sweat glands).

To further consolidate our findings that the most abundant proteins in saliva are directly related to RNA transcript levels and show gland-specific expression, we conducted immunofluorescent imaging using tissue sections from the three adult gland types. We found clear concordance of gland-specific expression on the protein level with RNA transcript levels for a number of genes including *AMY1*, lactoperoxidase (*LPO*), cysteine-rich secretory protein 3 (*CRISP3*), *STATH*, *MUC7*, and *MUC5B* (**Figure 5A**). These results confirmed tissue- and cell-specific expression differences in the different gland types, and are concordant with the available staining data in the Human Protein Atlas (30). One clear example is salivary amylase, an enzyme synthesized by serous acinar cells, that shows abundant expression on the protein level in PAR and SM glandular tissue while being virtually absent in SL. A similar trend was found for STATH and LPO. The lower expression levels of these gene products in the SL are likely due to the lower amount of serous acinar cells in this type of glandular tissue (56). However, the near-complete absence of amylase in the SL serous acinar cells indicates that these cells are distinctly different from their counterparts in the SM and PAR. Our findings confirm the validity of using these proteins as key markers to discern SM and PAR gland-derived tissues from those of the SL.

A different example of gland- and cell-specific expression is MUC7 that shows abundant expression at the protein level in the serous cells of the SL and to a lesser extent in the SM while being absent from PAR gland tissue (**Figure 5A**) which matches what we found on the RNA transcript levels from the respective glandular tissues (**Figure 2A**). Interestingly, in the SM, MUC7 is enriched in acinar cells deficient in amylase (**Figure S6**), suggesting that serous cells within the gland may exist as distinct populations. In support of cell- and tissue-specific expression of specific proteins within glandular tissues, our RNAseq results showed transcript levels for CRISP3, a gene coding for an androgen-regulated cysteine-rich protein that has been suggested as a potential early marker of Sjögren's syndrome (57), to parallel those of MUC7 in that both of these genes are highly transcribed in SL (**Figure 2A)**. Indeed, our immunohistochemistry results confirmed that the CRISP3 protein was highly expressed in the acinar cells of the SL with substantially less expression in the SM, and little to no protein or transcript expression in PAR. This is the first report of this protein's localization in the major salivary glands and suggests that CRISP3 along with MUC7 may be useful as markers typifying SL and SM tissue. A marker clearly typifying SL tissue is MUC5B, as shown by gene and protein expression being exclusive to the SL. Concordantly, our immunofluorescent analysis found abundant MUC5B protein in SL mucous cells, making this protein an appropriate marker for SL mucous cell function. Overall, immunofluorescent imaging results highlight a need for resolving within-tissue variation in gene expression at the single-cell resolution.

Finally, we conducted gel electrophoretic separation and western blot analysis of glandular ductal secretions for AMY1, MUC7, CRISP3, and BPIFA2/SPLUNC2 (**Figure 5B).** As revealed by Coomassie blue and periodic acid Schiff stain, the combined secretions of SM and SL (SM/SL) glands showed strikingly different patterns of protein and glycoprotein bands compared to PAR secretion while whole mixed saliva showed a combination of both. Western blot analysis supported our findings in that abundant saliva proteins are expressed in a gland-specific manner and are regulated at the level of transcription. AMY was predominantly expressed in PAR secretion sample, with less, although still substantial levels, in SM/SL secretions, whereas MUC7 was highly expressed in SM/SL secretion and undetectable in PAR ductal saliva. These results are consistent with previous reports (50, 58–60) as well as with our immunofluorescent analysis of salivary gland tissue (**Figure 5A**). We confirmed BPIFA2, a highly glycosylated protein that may function in the innate immune response, to be expressed in whole saliva (61) and we further found it to be enriched in the SM/SL sample, and weakly expressed in PAR secretion, supporting our transcriptome-based evidence that this protein is predominantly derived from the SM. Finally, a doublet of bands for CRISP3 was detected in whole saliva, as was reported previously (62). Here we show for the first time that the origin for the CRISP3 protein is restricted to SM/SL secretions with no protein detectable in PAR ductal saliva, thus matching our immunofluorescent analysis (**Figure 5A**). Interestingly, the CRISP3 band in SM/SL ductal secretion migrates at a lower molecular weight range than the double bands detected in whole mouth saliva. This outcome suggests that post-secretion enzymatic processing occurs, likely resulting in the alteration of CRISP3 sialylation by oral bacterial sialidases, leading to a loss of negatively charged sialic acid glycans and, thus, retarded mobility of the protein in the electrophoretic field (62, 63).

Together, our integrative analysis of RNA and protein levels support the hypothesis that the gland specificity of the secretion of these proteins is likely regulated at the transcriptional level. Moreover, we have identified saliva proteins that originate in the salivary glands, thereby providing bona fide markers of salivary gland function.

CONCLUSION

Our analysis of the transcriptomes of mature and fetal salivary gland tissues identified hundreds of genes that together define mature salivary glands as specialized secretory organs. In addition, we documented dozens of genes, which mostly code for abundant secreted proteins that specify the three major salivary gland types, parotid, submandibular, and sublingual glands. Integrating our glandular RNAseq and mass spectrometry-derived protein abundance data, we were able for the first time to parse out the origins of proteins present in human saliva. In other words, we were able to determine which proteins in whole saliva are intrinsically produced by the salivary glands and which are derived from other sources, such as epithelial cells and plasma leakage in the mouth. Our current understanding of these connections, as based on our own transcriptome data and on proteome data derived from Human Salivary Proteome wiki schematically summarized in **Figure 6**. Our quantitative transcriptomic and proteomic data also allowed us to define relationships between the level of salivary gland gene transcripts and their corresponding proteins in saliva. Similar to other studies exploring the relationship between high abundance proteins and mRNA levels (46), as well as quantitative studies in other secretory cell systems (46), we found that the abundance of highly expressed saliva proteins is regulated at the transcriptome level. Such a correlation also suggests that these abundant proteins undergo efficient translation, although further tests are needed to confirm if this is the case. In addition, quantitative mass spectrometry-based studies analyzing the proteome of each of the 3 major glands themselves is required in order to understand why the levels of some abundant secreted proteins in saliva did not correlate with glandular transcript levels.

Based on this analysis and our spatial mapping, we were also able to identify two subsets of serous acinar cells in the submandibular gland that appeared to be specialized for expressing either AMY1 or MUC7. Acinar cells have been traditionally defined in a binary categorization as either serous or mucous. However, our data suggest that these cells are more heterogeneous than previously thought and that subsets of these cell types synthesize specific salivary proteins. These insights have major implications for understanding the relationship between glandular secretions and protein content of these bodily fluids and consequently inform saliva diagnostics. Furthermore, our transcriptomic analysis of healthy human gland tissue has major ramifications for *de-novo* engineering of these organs as well as determining the impact of the disease on salivary gland homeostasis and function.

Our results furthermore indicate that the mature salivary glands are highly metabolically active secretory organs and are enriched for pathways facilitating secretory functions. In addition, we discovered that salivary glands serve as hubs of immune activity. All these functions that were found in mature glands were absent in fetal salivary glands. Instead, we found in fetal glands an enrichment of developmentally related genes expressed. Fetal salivary glands, although anatomically distinct from each other, could not be distinguished based on their transcriptional profiles. This indicates that the developmental differentiation of glandular function and functional specialization of the three mature gland types occur relatively late during fetal development. These findings pave the way for future studies dissecting mechanisms of regulation of transcriptome during glandular development and have significant implications for *de novo* organ generation.

METHODS

Human tissue and saliva samples: Adult salivary gland tissues were collected from patients aged 23 to 70 years by clinicians at the University of California, San Francisco Medical Center during routine surgeries with informed consent (UCSF Biorepository, institutional review body approval number 17-22669). The UCSF pathology lab deemed tissues to be healthy. Human fetal salivary glands were harvested from post-mortem fetuses between 22 and 24 weeks gestation obtained from elective legal abortions with the written informed consent of the women undergoing the procedure and the approval of the Institutional Review Board at the University of California San Francisco (IRB# 10-00768). Specimens were donated anonymously at San Francisco General Hospital. Saliva collection was performed as approved by the University at Buffalo Health Science Institutional Review Board (Study Nr. 030-505616). See **Supplementary Methods** for details on sample collection and sex distribution.

RNA isolation and sequencing: RNA extraction, library preparation, and sequencing were conducted using standard methodologies to generate paired-end 150bp reads that passed standard quality control steps. The resulting reads were mapped to the human transcriptome reference (hg19) from Ensembl (64) and quantified using Kallisto (65). Differential expression analysis was performed by DESeq2 (27). All downstream analysis was conducted using custom-bioinformatic pipelines available at https://github.com/GokcumenLab/glabBits (under Saliva - RNAseq). The normalized gene expression levels can be found in **Table S2**. Raw RNAseq datasets are uploaded to the GEO database https://www.ncbi.nlm.nih.gov/geo/ with the project name GSE143702. It will be published on July, 10th, 2020. Detailed methodological details can be found in **Supplementary Methods**.

Immunofluorescent imaging, gel electrophoresis and immunoblotting: The immunofluorescent analysis was performed on 12 μm-thick fixed tissue sections, as previously described (66). Sections were imaged using a line-scanning confocal microscope (Leica sp5). Sample preparation, gel electrophoretic separation, staining of protein and glycoprotein bands by Commassie and periodic acid Schiff stain as well as immunoblotting were performed as previously described (5, 43). Methodological details and a description of the antibodies used can be found in **Supplementary Methods**.

ACKNOWLEDGMENTS

The authors would like to thank William Lau for help with retrieving the protein abundance data from the NIDCR Human Salivary Proteome Wiki, and Tasha Lea (certified pathologist assistant) and Erica Oropeza, along with The Biospecimen Resources Program (BIOS) at UCSF, for facilitating the efficient acquisition, quality control and management of biospecimens. This work was supported by the National Institute of Dental and Craniofacial Research (NIDCR) under the award numbers 1R35DE028255 (SMK) and 2R01DE019807 (SR), and by the Tobacco-Related Disease Research (TRDRP) under award #28IR-0071 (SMK).

FIGURES

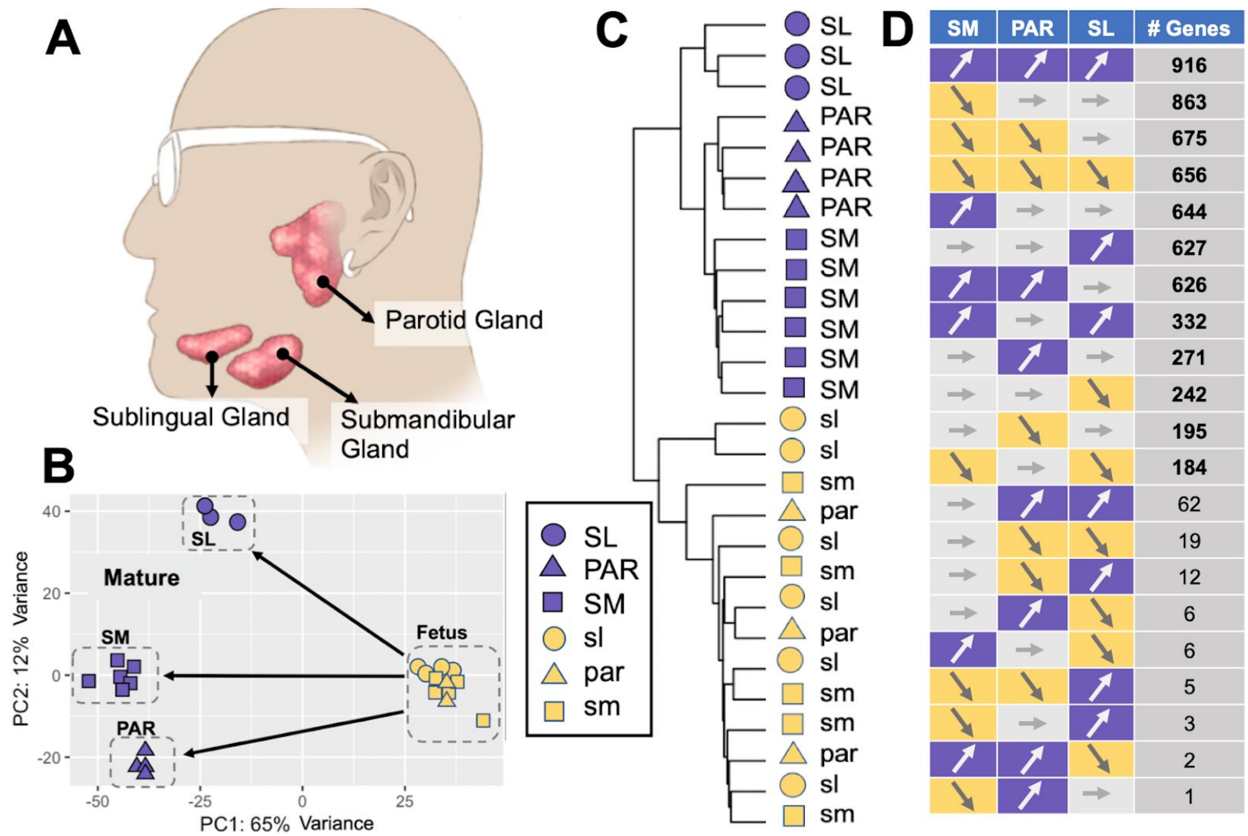


Figure 1. General trends of gene expression in adult and fetal salivary glands. **(A)** Anatomical location of the three major glands in humans. **(B)** Principal component analysis of the gene expression levels in adult and fetal parotid, submandibular, and sublingual glands. Adult gland types show clear separation based on their gene expression without any *a priori* hypothesis, while fetal glands cannot be separated based on their gene expression alone. **(C)** Hierarchical clustering analysis of adult and fetal parotid, submandibular, and sublingual, glands based on all the transcriptome data without *a priori* clustering information. Purple symbols represent adult samples and yellow symbols represent fetal samples. Triangle, square, and circle shapes represent parotid, submandibular, and sublingual glands, respectively. **(D)** The number of genes with transcript levels higher in adult glands (designated by purple background and white up arrow) or lower in adult glands (designated by yellow background and gray down arrow) as compared to the corresponding fetal glands (adjusted p-value < 0.0001, the gray background with horizontal arrows indicate non-significance).

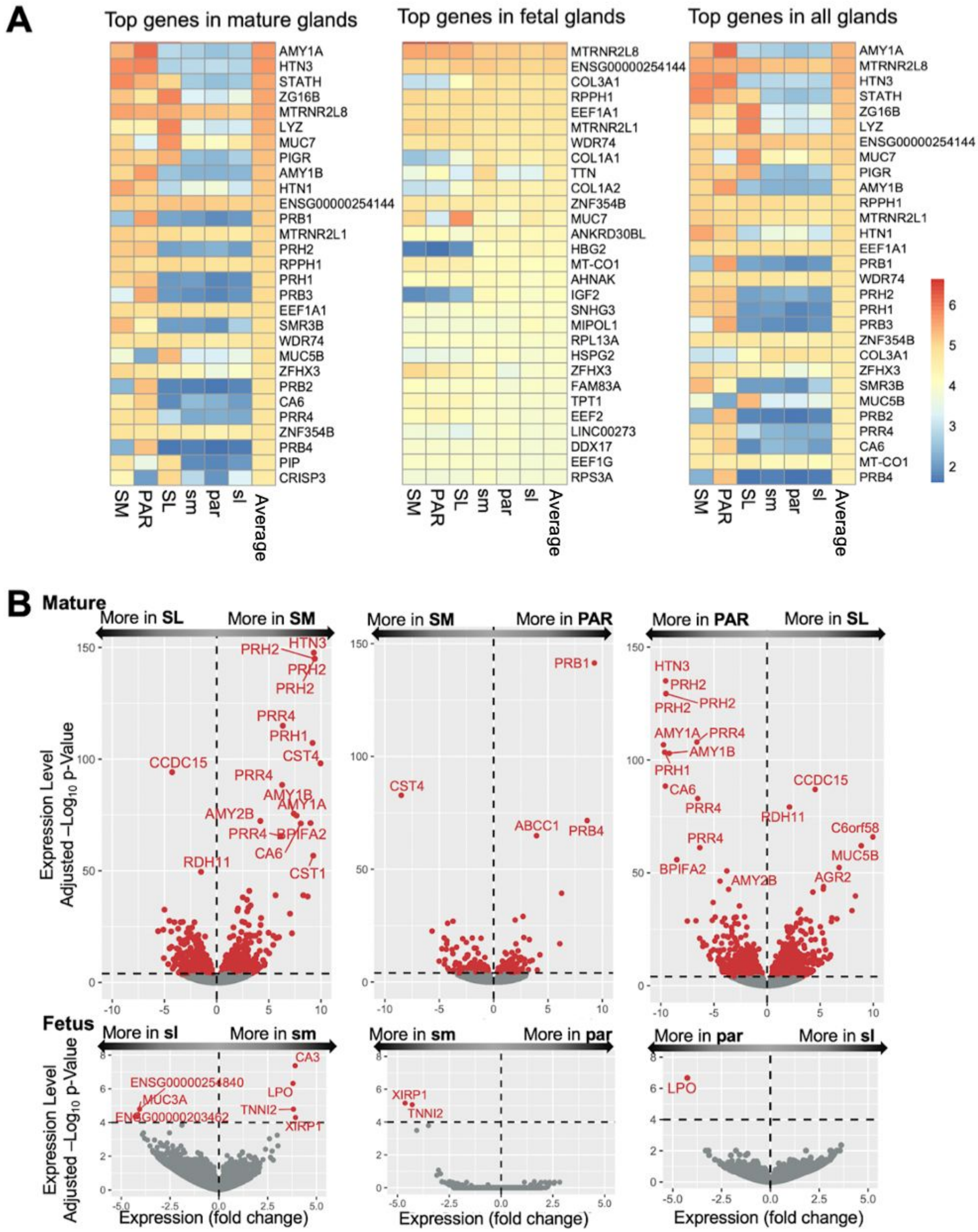

A
Top genes in mature glands
Top genes in fetal glands
Top genes in all glands
SM PAR SL sm par sl Average
B Mature
More in SL
More in SM
More in PAR
Fetus
More in sl
More in sm
More in par
Expression Level
Adjusted −Log10 p-Value
Expression (fold change)

Figure 2. Adult glands can be distinguished from fetal glands and from each other through the expression of secreted proteins. (A) Heatmaps of the $\log_{10}$ normalized expression of gland-specific genes. The top 30 expressed genes in the mature, fetal and both types of glands are shown. Note that the top gene, *RN7SL1,* was excluded because of its role as a housekeeping gene. *MUC7* is expressed in both adult and fetal glands and is enriched in adult sublingual glands. *HTN, AMY*, and *STATH* are highly expressed in adult parotid and submandibular glands. *CST* genes are expressed in the adult submandibular gland, and *PRB* genes are expressed in the adult parotid gland, specifically. **(B)** Transcriptional comparison of adult and fetal major salivary glands. Volcano plots showing the expression differences between glands in a pairwise fashion for adult (top) and fetus (bottom). The x-axis indicates gene expression fold changes and the y-axis indicates $-\log_{10}$ value of adjusted *p*-value. Genes with significantly different expressions between glands are indicated in red on the plots (adjusted *p*-value < 0.0001), others were shown in gray.

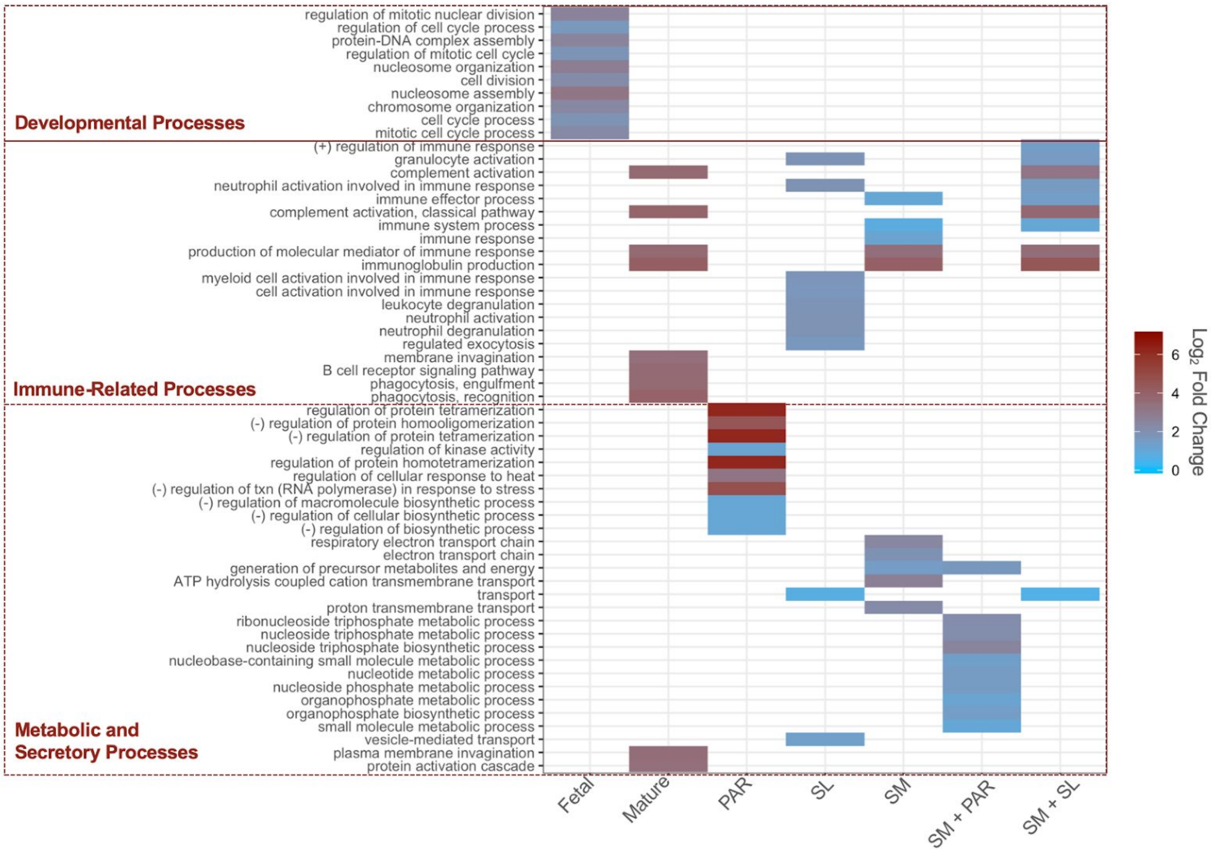


Figure 3. Gene ontology (GO) enrichment analysis of adult and fetal salivary gland transcripts indicates shared and tissue-specific biological functions in adult tissue. The rows represent individual gene ontology categories grouped in broader functional sections by red-dashed rectangles. The columns show the subset of genes used for enrichment analysis. The first column includes the enrichment of genes that are expressed significantly higher in fetal glands as compared to their mature counterparts. The remaining columns include enrichment analysis of genes that are expressed significantly higher in mature tissues as compared to their fetal counterpart. 2nd column: upregulated in all mature glands, 3rd - 5th: upregulated only in PAR, SL, and SM, respectively. 5th: upregulated in SM and PAR, but not in SL. 6th: upregulated in SM and SL, but not in PAR. The colors indicate $\log_2$ fold enrichment of the number of genes from the expected number of genes. Note that we included the top 10 most significant enrichments based on false discovery rate for each gland category. The full results for the GO analysis can be found in **Table S3**.

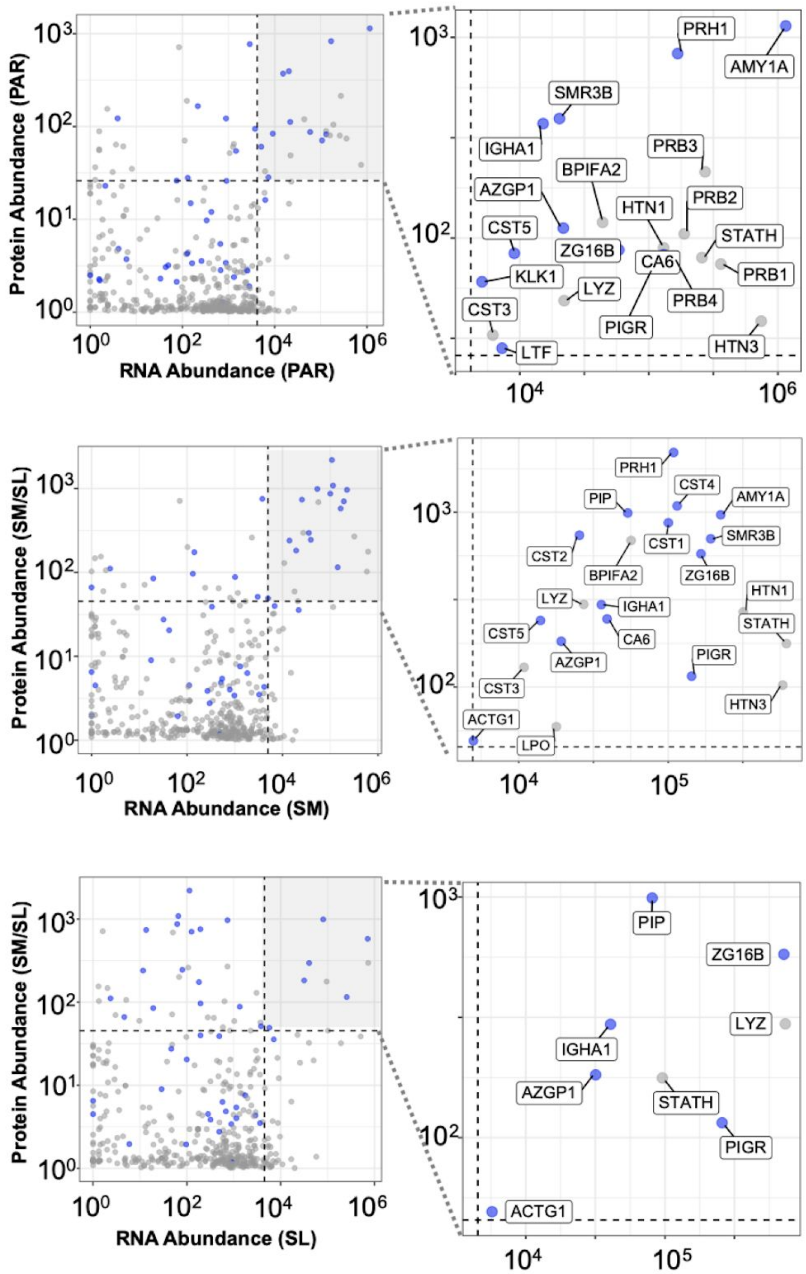


Figure 4. Gland-specific expression of the most abundant secreted saliva proteins is predominantly regulated at the transcriptome level. In order to compare transcript levels of genes expressed by the 3 types of salivary gland encoding secreted proteins to protein levels in saliva, we integrated our transcriptomic data with the mass spectrometry-derived proteome of whole saliva (Human Salivary Proteome Wiki database). Each graph represents a comparison of transcript abundances of a specific gland type with protein abundances for that gland's corresponding ductal saliva. The x-axis represents $\log_{10}$ normalized transcript levels and the y-axis shows $\log_{10}$ normalized protein abundances found in glandular secretions. Genes showing the highest abundance (top 10%) at both the transcript and protein level are

highlighted in the top-right quadrant by a grey background. The areas of the grey quadrants are shown enlarged in the right panels with the proteins labeled. Blue dots indicate genes coding for secreted proteins that are highly abundant at both the transcript and protein levels. Grey dots indicate genes that are highly abundant at the transcript level but not at the protein level.

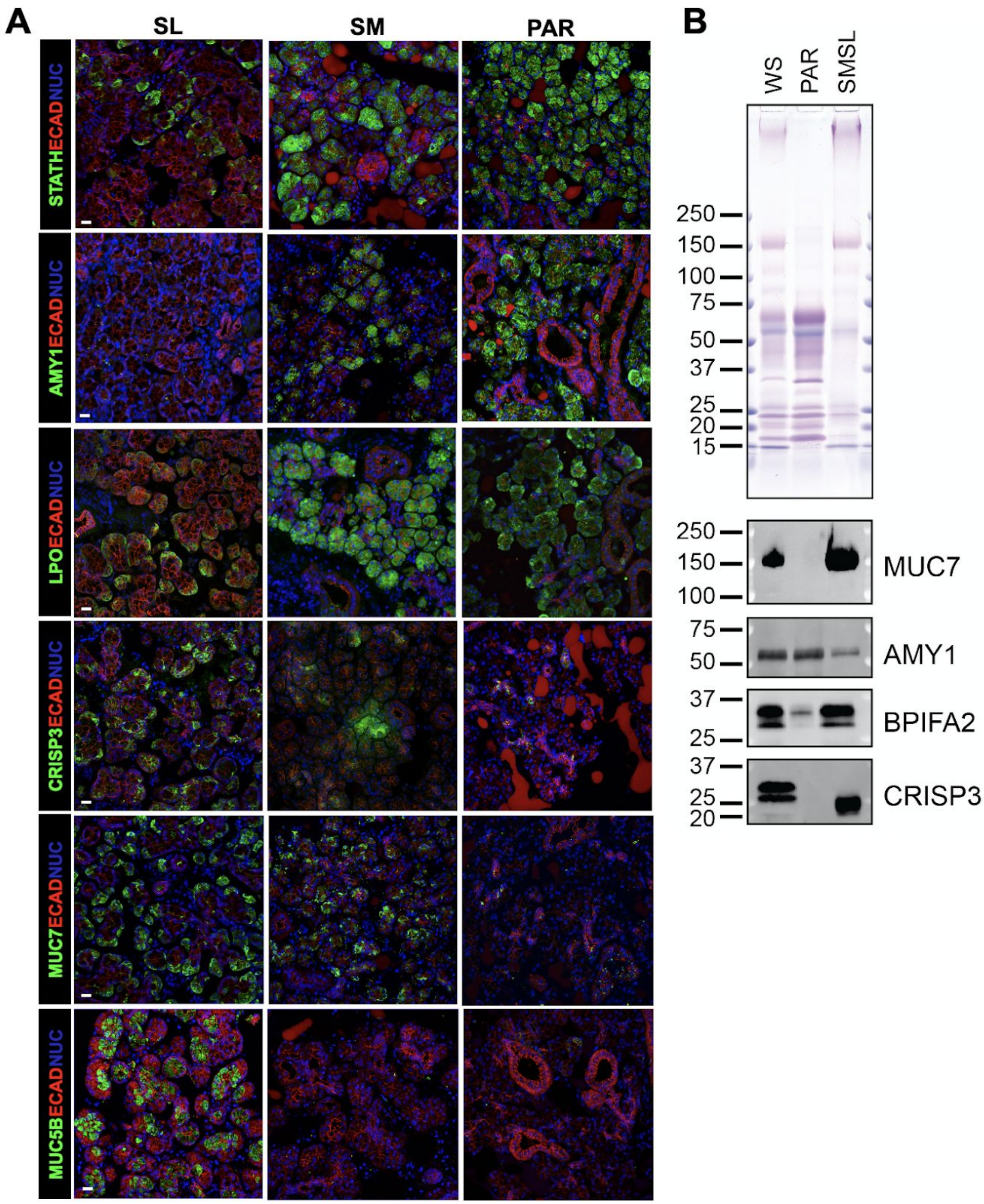


Figure 5. Secreted proteins are differentially expressed among the three major salivary gland types. A. Immunohistochemistry of glandular tissues. The SM and PAR acinar cells are highly enriched for amylase (AMY1), statherin (STATH) and lactoperoxidase (LPO) compared to the SL, consistent with these being markers of serous cells. MUC7 and CRISP3 are expressed by a subset of acinar cells of the sublingual and submandibular gland, with little to no

expression in the PAR. MUC5B is highly expressed by the SL mucous acinar cells but not by the acinar cells of the PAR or SM, indicating it to be a marker of SL function. ECAD = E cadherin; NUC = nuclei. Scale bar = 25 µm. **B**. Gel electrophoretic separation of WS and glandular secretions (PAR, SMSL) followed by staining of proteins with Coomassie blue and of glycoproteins with periodic acid Schiff stain (pink bands in upper panel), and probing of transfers with antibodies against MUC7, AMY, BPIFA2, and CRISP3.

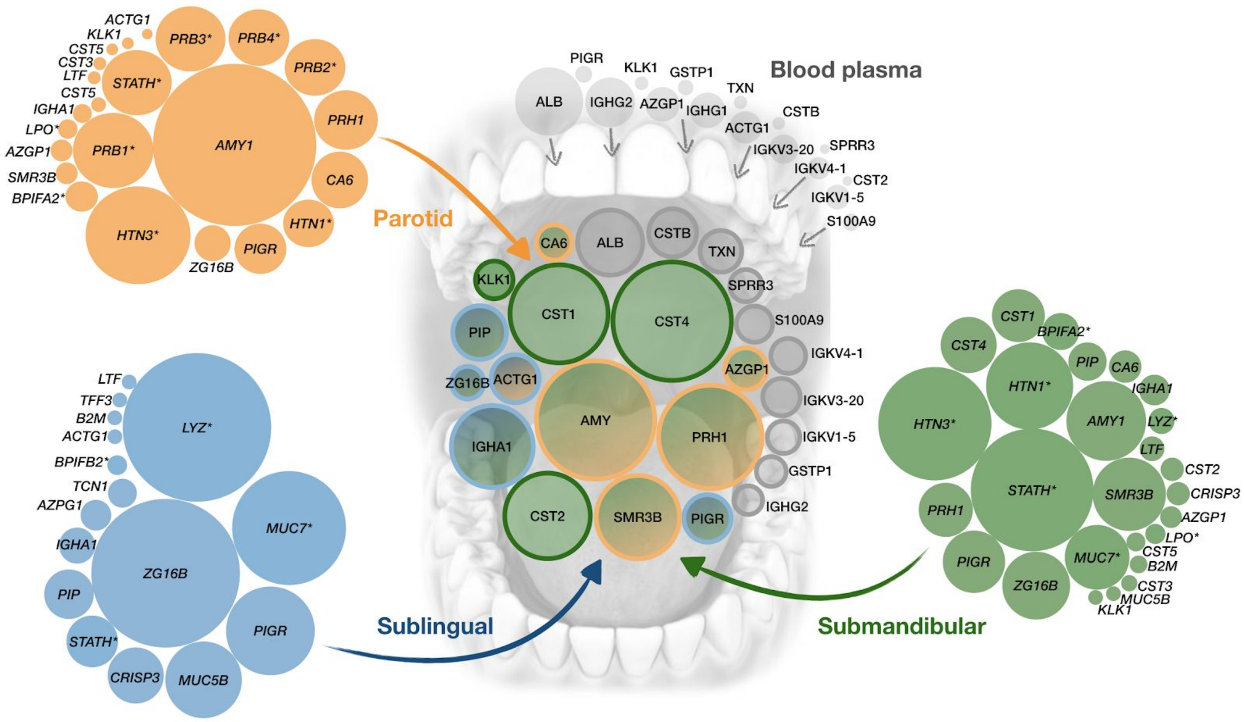


Figure 6. Glandular origins of the most abundant saliva proteins in whole mouth saliva. This schematic diagram summarizing our observations shows the putative origins of the most abundant proteins in whole mouth saliva. The central group of circles represents the most abundant proteins detected in whole mouth saliva by quantitative mass-spectrometric analysis (data source: HSP-Wiki https://salivaryproteome.nidcr.nih.gov/). The areas of the circles approximate the observed protein abundances. Orange, green, blue, and gray colors indicate the putative origin of these proteins from the corresponding major salivary glands or from blood plasma exudate. The groups of circles on the outside represent the gene (RNA) expression levels in PAR (orange), SL (blue), and SM (green) coding for abundant salivary proteins as detected by mass spectrometry in the corresponding glandular secretions (data source: HSP-Wiki https://salivaryproteome.nidcr.nih.gov/). The areas of the circles approximate relative RNA abundances in each gland type. Note that some proteins, indicated by an asterisk (*) near the protein's name are detected at the glandular level as secreted proteins, but were not among the most abundant proteins detected in whole mouth saliva. The gray circles on top indicate the relative protein abundances in blood plasma of only those proteins that were abundantly detected in whole mouth saliva.

SUPPLEMENTARY MATERIAL

Supplementary Tables

Table S1. Summary information on the salivary glands used in the study. The tissue of origin, developmental stage (Mature or Fetal), sex, age, and the sample size for each category. The table is embedded in **Supplementary Methods**.

Table S2. Transcriptome data for each of the 3 major salivary gland types. Tab 1: Directory (Readme) that includes a description of the datasets presented in this table. Tab 2: Processed transcriptome data for each of the 3 major salivary gland types. The non-aligned RNAseq data are available at GEO https://www.ncbi.nlm.nih.gov/geo/ with the project name GSE143702. It will be published on July, 10th, 2020. Tab3: The raw transcriptome data (normalized gene read counts) for each tissue sample derived from the 3 major salivary gland types.

Table S3. The results of the gene ontology analysis. Tab1 is a directory (readme) describing the analysis specifics that produced the data presented in this table. Tabs 2-13 show GOrilla (http://cbl-gorilla.cs.technion.ac.il/) output for differentially expressed gene groups. Each tab contains three rows indicating GO results for "Biological Process", "Cellular Component", and "Molecular Function". Tab 14. Shows enrichment for genes that encode proteins that are abundantly found in saliva (top 10%) but are not highly expressed (bottom 10%) at the RNA level in glandular tissues are defined as "Proteome > RNAseq" and vice versa.

Table S4. The tissues of origins for RNA (GTEx) and Proteome (Human Protein Atlas) comparison databases used in this study. The table is embedded in Supplementary Methods.

Supplementary Figures

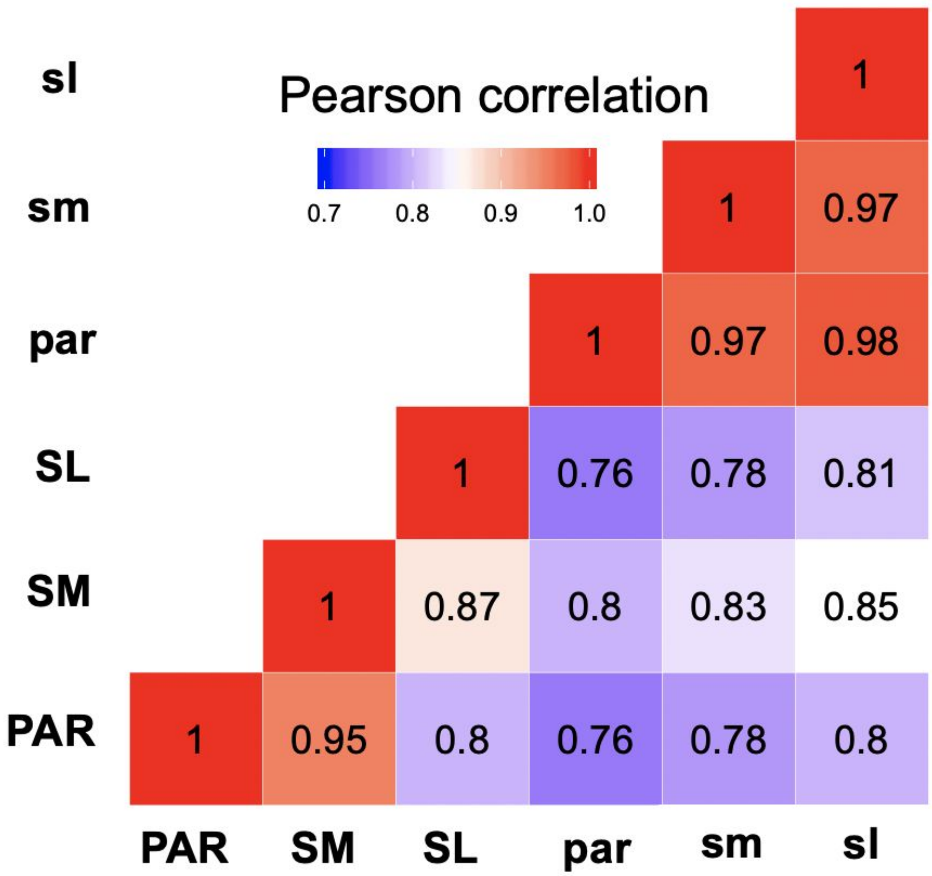
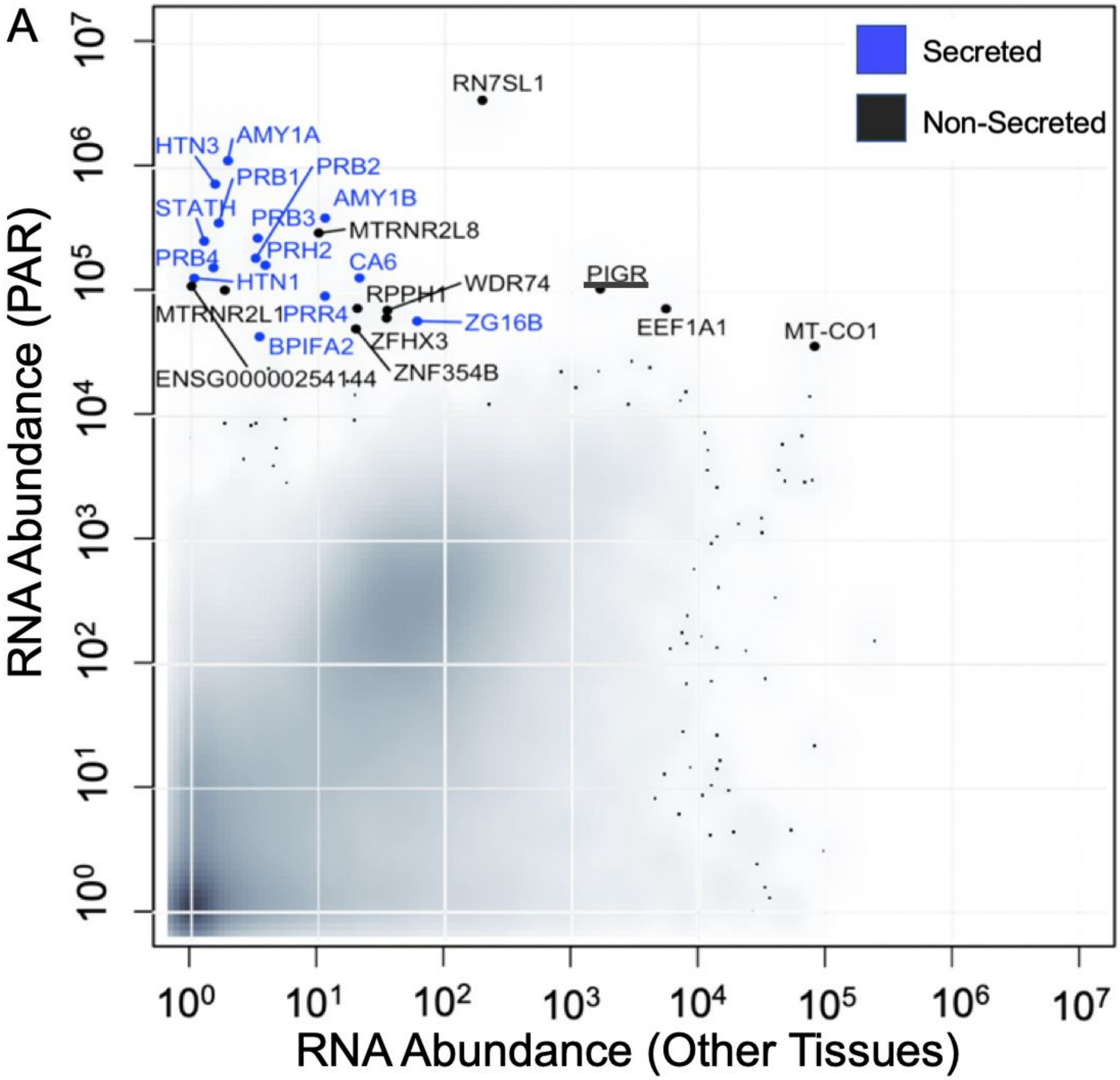
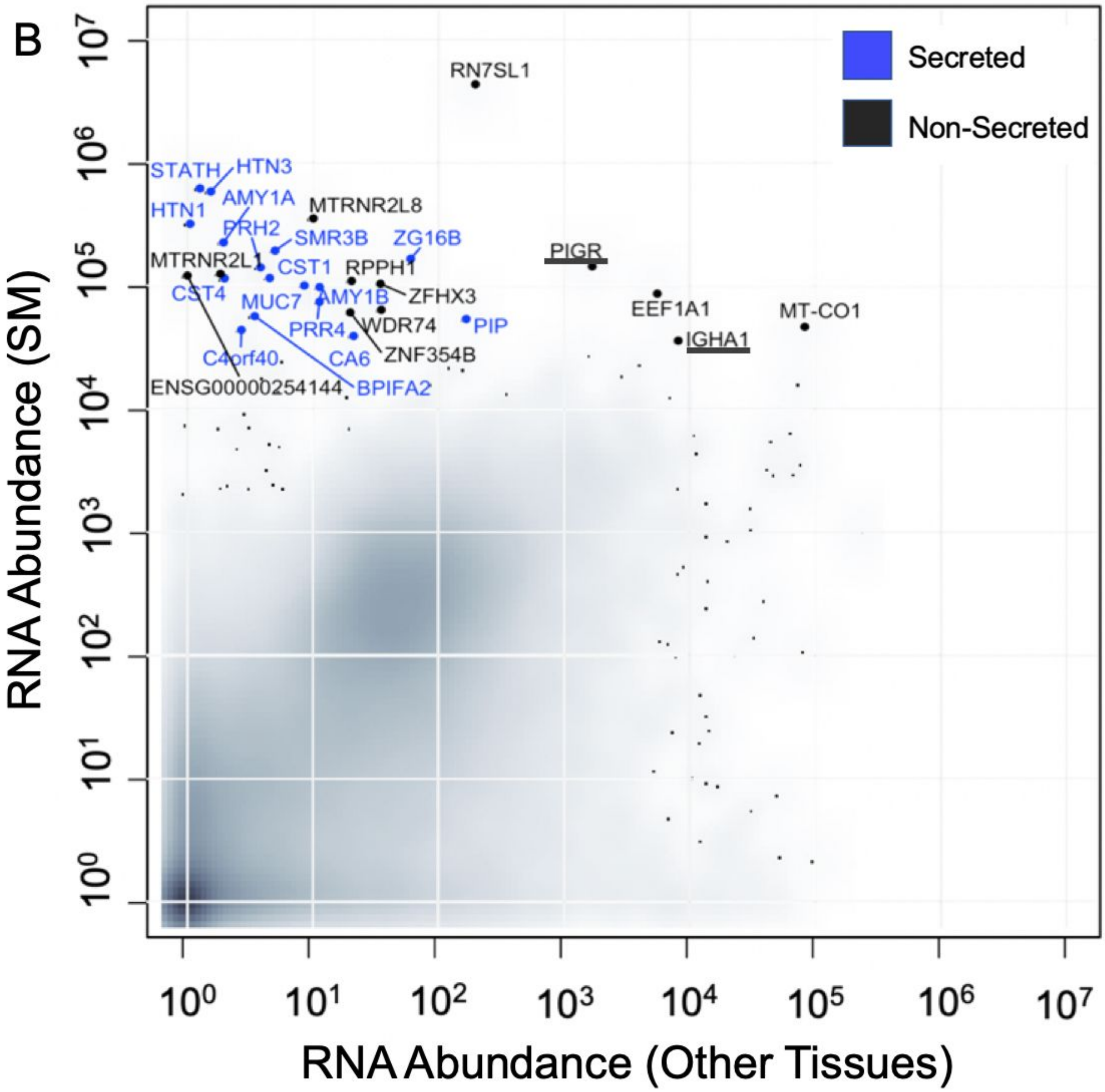


Figure S1. Correlation analysis of the genome-wide transcriptomes of all salivary gland samples included in this study. A heatmap was constructed by the Pearson correlation of gene expression of all salivary gland samples, with the number in each box representing the correlation coefficient among each pair compared.


A
Secreted
Non-Secreted
RN7SL1
HTN3
AMY1A
PRB1
PRB2
AMY1B
STATH
PRB3
MTRNR2L8
PRH2
CA6
PRB4
HTN1
RPPH1
WDR74
PIGR
MTRNR2L1
PRR4
ZG16B
EEF1A1
MT-CO1
BPIFA2
ZFHX3
ZNF354B
ENSG00000254144
RNA Abundance (PAR)
RNA Abundance (Other Tissues)
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7


B
Secreted
Non-Secreted
RN7SL1
STATH
HTN3
AMY1A
MTRNR2L8
HTN1
PRH2
SMR3B
ZG16B
PIGR
MTRNR2L1
CST1
RPPH1
CST4
MUC7
AMY1B
ZFHX3
EEF1A1
MT-CO1
PRR4
WDR74
PIP
IGHA1
C4orf40
CA6
ZNF354B
ENSG00000254144
BPIFA2
RNA Abundance (SM)
RNA Abundance (Other Tissues)

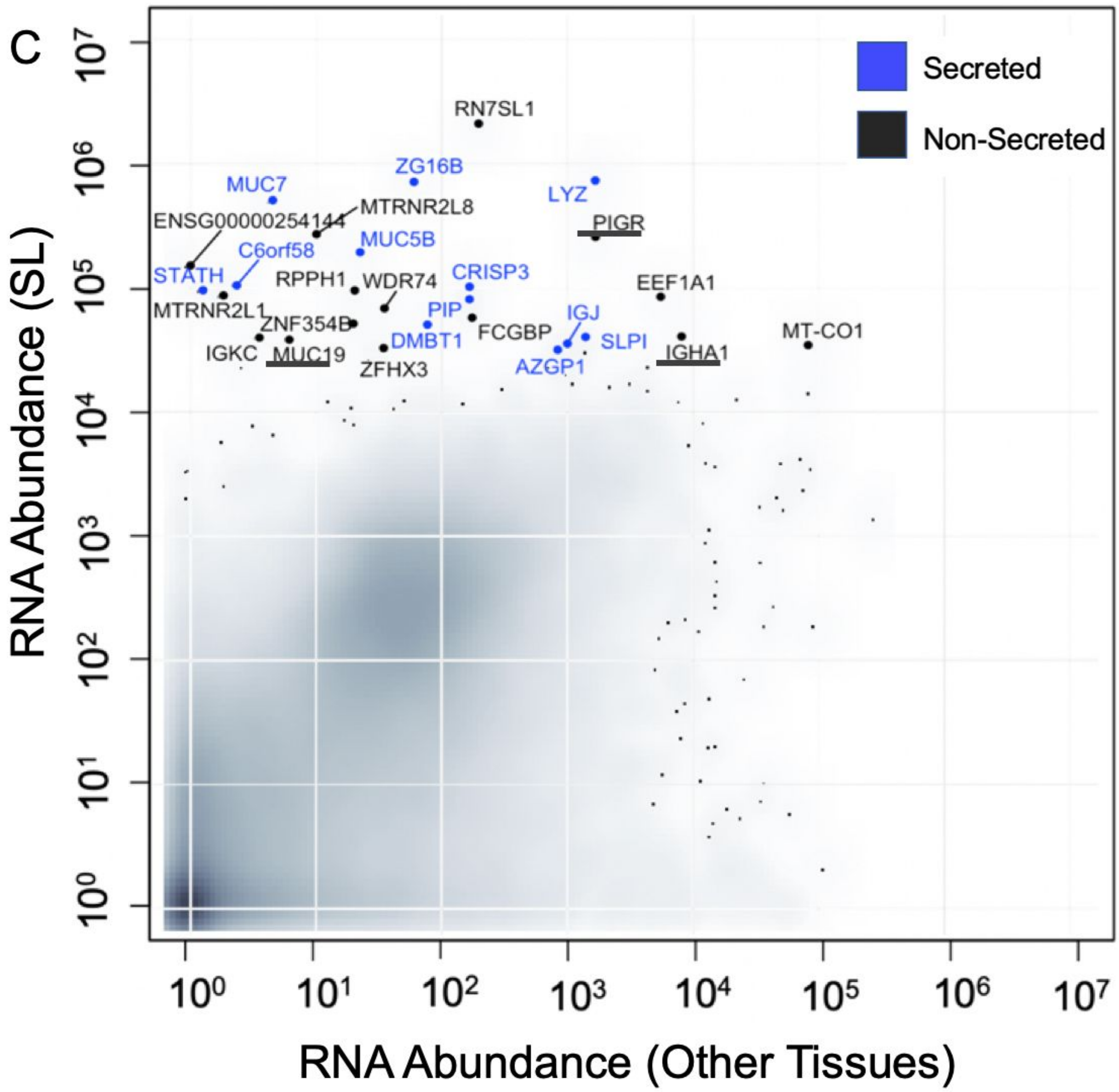


Figure S2. Salivary gland-specific gene expression compared to other tissues. (A-C) The y-axis shows the transcript levels of each gene ($\log_{10}$ (1+normalized read counts)) for each adult major salivary gland type. For comparative purposes, the x-axis shows the maximum gene expression levels in through GTEx Portal except for minor salivary glands **(Table S4)**. Genes coding for secreted proteins are highlighted in blue. Note that also PIGR, IGHA1, and MUC19 (underlined) can occur as both membrane-bound and secreted proteins, although they were not classified as "secreted protein" in the Human Protein Atlas (see **Supplementary Methods**).

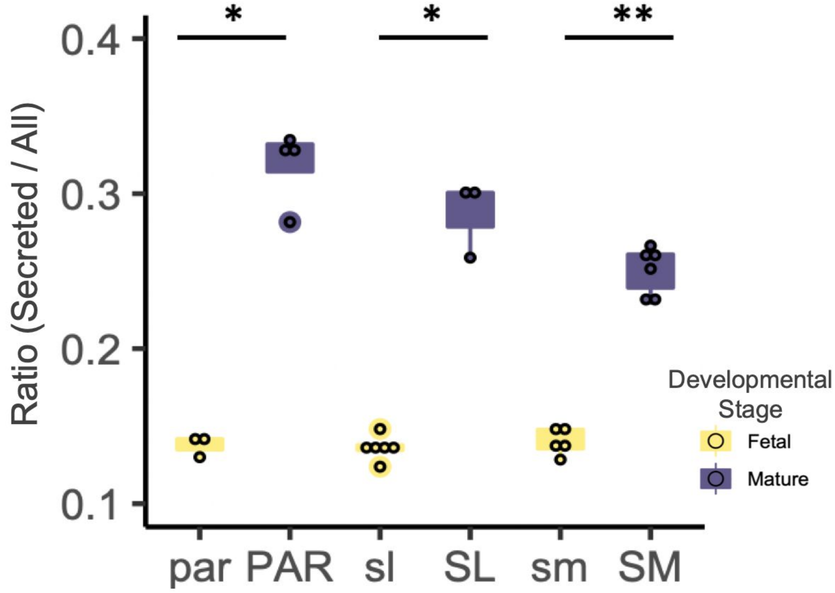


Figure S3. Bar plots showing the ratio of transcripts encoding for secreted proteins versus all proteins in adult and fetal salivary glands. Genes were classified based on whether they encode “secreted” proteins based on protein annotation obtained from the Human Protein Atlas (30). A ratio for each salivary gland was derived by dividing the sum total of all gene transcripts encoding secreted proteins by the sum total of all gene transcripts expressed. Each dot represents data from an individual salivary gland tissue sample. Mature glands have a significantly higher relative expression of genes that encode for secreted proteins than fetal glands (*p < 0.05.**p < 0.01, One-tailed Wilcoxon rank-sum test).

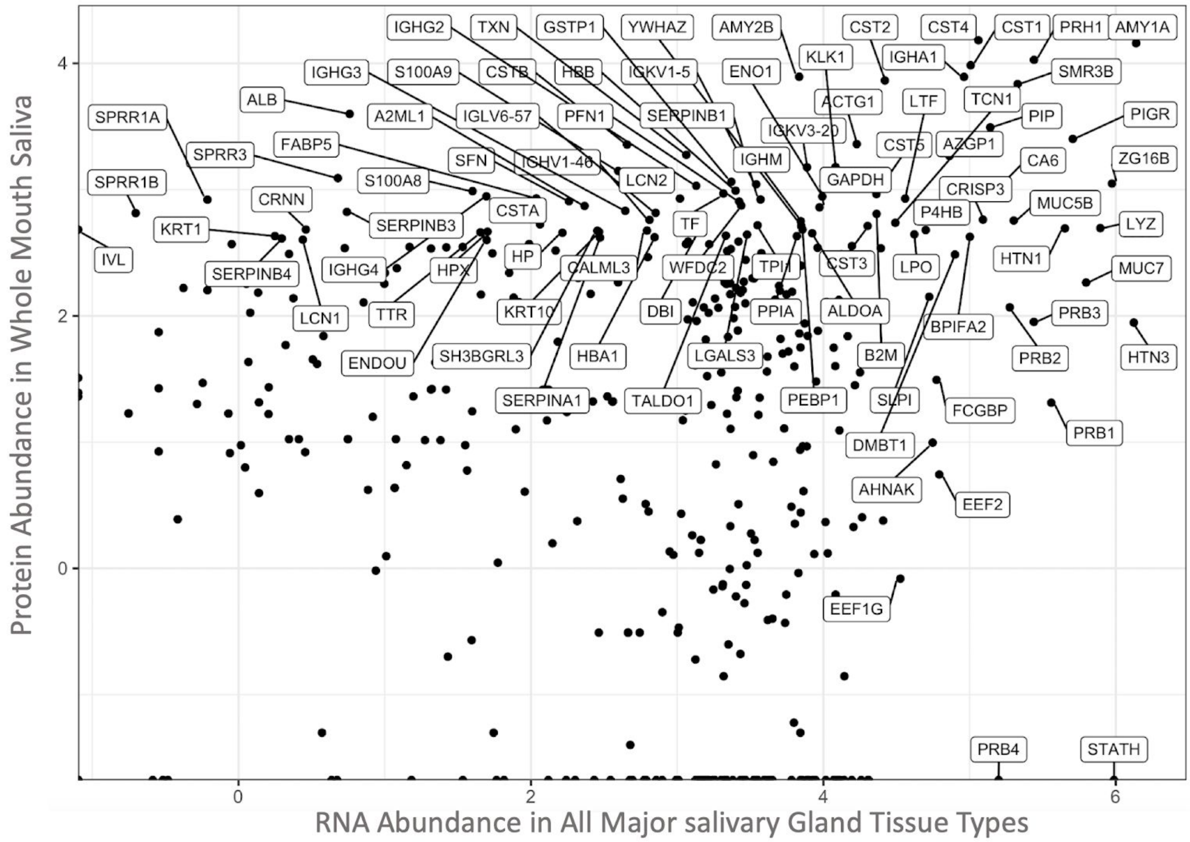


Figure S4. Secreted proteins in the whole saliva identified by RNAseq analysis of major salivary gland tissues. The total sum of transcripts encoding for secretory proteins in all the three types of adult glands (x-axis, $\log_{10}$) was plotted against the secreted protein abundance in whole mouth saliva (x-axis, $\log_{10}$). Well known secreted proteins (*e.g.*, PIGR, AMY1A, CST4) were found highly abundant at both the mRNA and protein levels, indicating that these proteins are likely derived from the salivary glands. In contrast, non-secreted proteins that are abundant in whole saliva, such as KRT1 and SERPINB4, showed negligible transcript levels (<10 reads) in salivary gland tissue, suggesting that they originate from other tissues or organs.

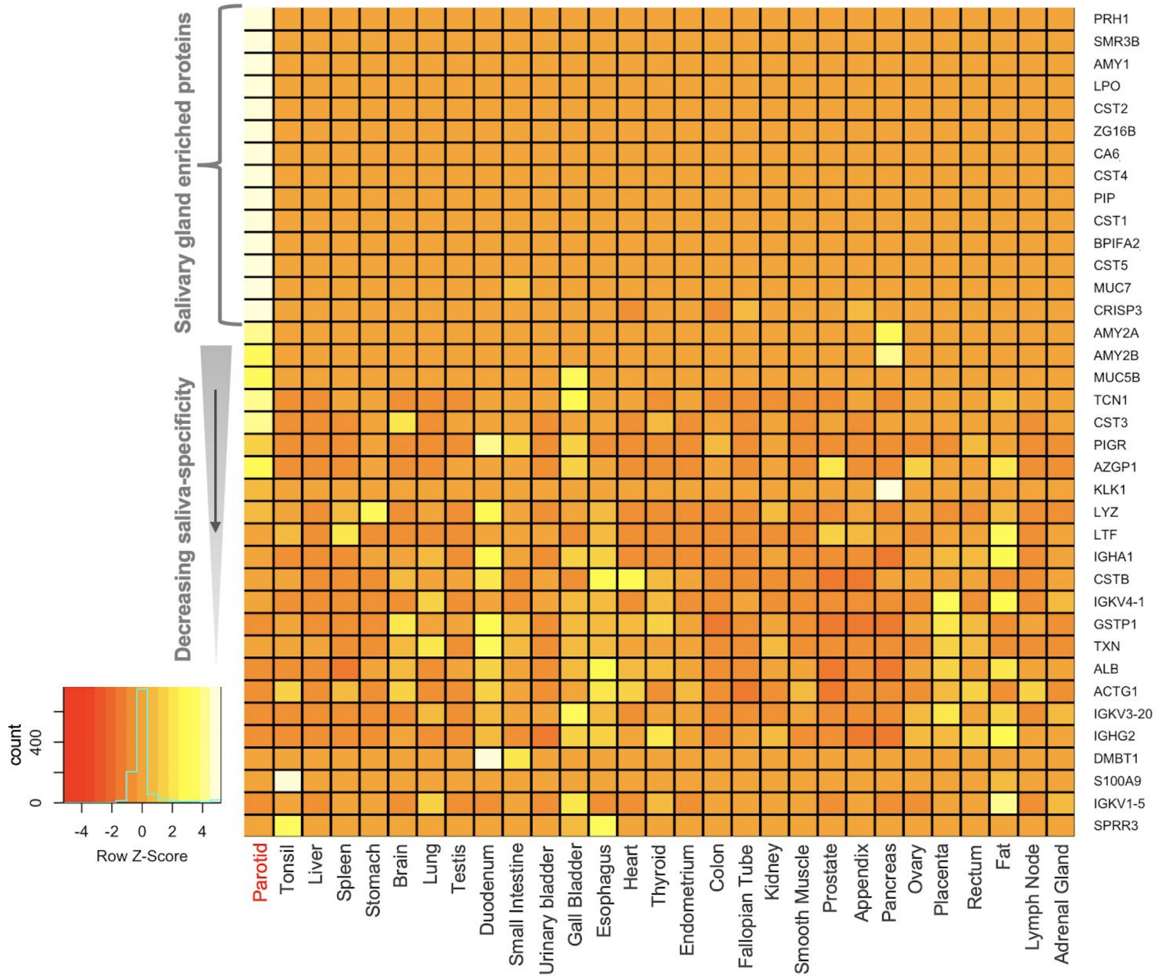


Figure S5. Comparison of the most abundant proteins in human saliva (parotid gland) with the proteomes of 29 human organs from the Human Protein Atlas database (52). Genes were chosen based on their expression levels from the saliva proteome and salivary gland RNAseq analysis. The data is normalized across rows for observing the comparative abundance of proteins across tissue types. The x-axis labels tissue types and the y-axis the gene names. The genes are ordered from top to bottom based on their enrichment in saliva/salivary glands. The colors represent a z-score *i.e.*, the deviation from the mean expression in a given row.

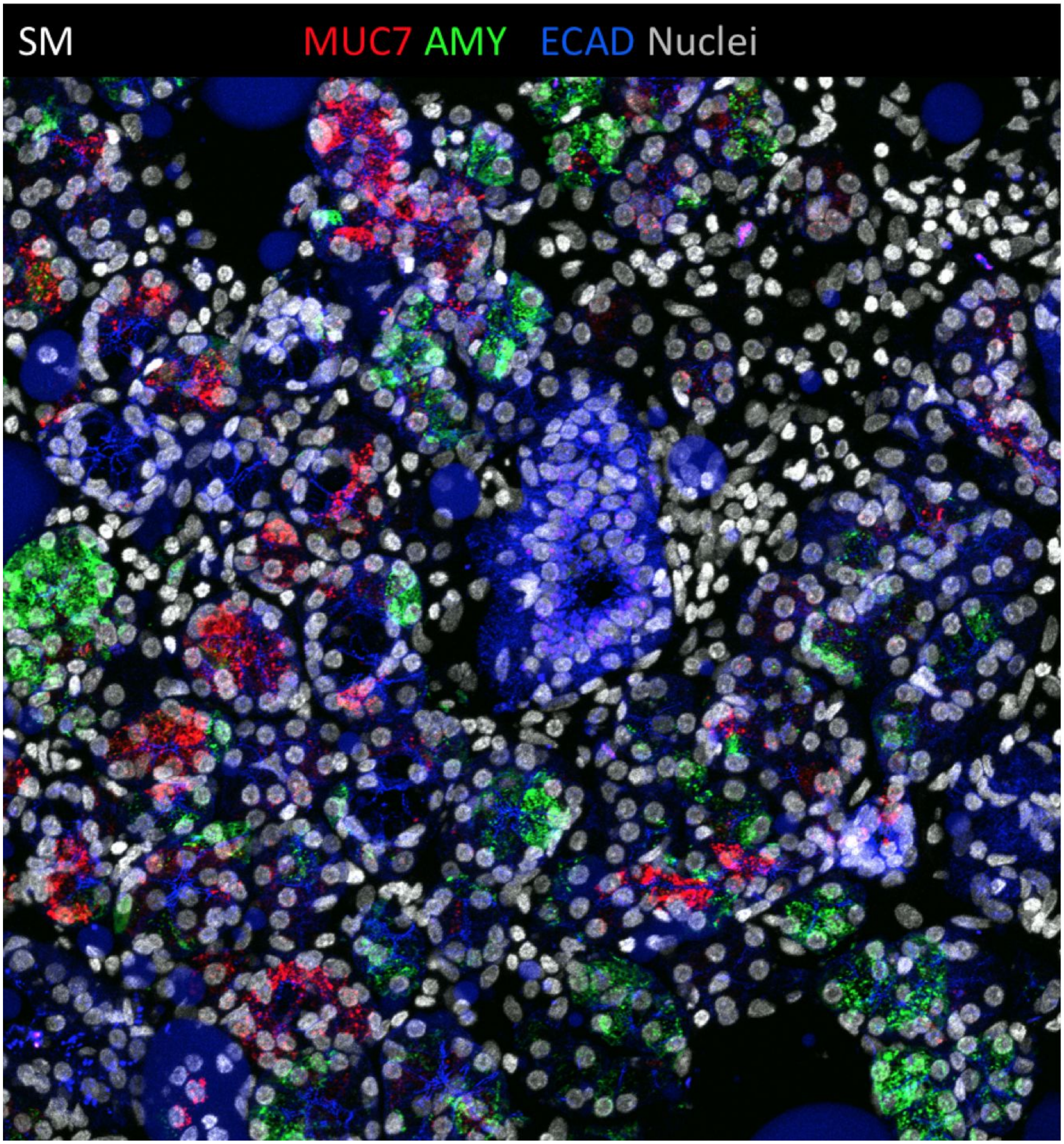


Figure S6. MUC7 and amylase are expressed by distinct subtypes of serous acinar cells. The mature submandibular gland tissue section was immunolabeled for MUC7 (red), AMY (amylase, green) and E-cadherin (blue).

Supplementary Materials and Methods

Salivary gland samples: Adult human salivary gland biopsies (**Table S1**) were collected via the UCSF Biospecimen Resources Program (BIOS) under the institutional review body approval number 17-22669. Sample collection for the adult tissues took place during oral surgery procedures that were performed independent of this project. SM and PAR tissue samples were taken during oral surgeries from individuals suffering from cancers of the head and neck. The sample collection was limited to those patients who had not received radiotherapy, chemotherapy or immunotherapy. SL samples were derived from patients with salivary duct stones. Healthy tissue regions were identified and separated from inflamed or cancerous tissues by the UCSF pathology lab. Our immunofluorescent analysis further confirmed tissue health, as determined by cell and tissue morphology and the absence of lymphocytic infiltrates. It is unlikely but remains plausible that the disease status of the patients may have altered transcript levels. For example, it is possible that certain diseases, including oral cancers, can generate widespread inflammation, biasing our results for detecting higher levels of immune-system related genes. In addition, all adult sublingual glands were derived from female donors. To test for any potential biases, we investigated the variation of gene expression among samples of the same gland type as well as sex-specific expression differences. We observed extremely small variation in gene expression abundances of samples of the same gland type (**Figure 1B** and **Figure 1C**) and no sex-specific trends at the global transcriptome level (data not shown). This indicates that the differences among individual samples of the same gland type are much smaller than the gland-specific transcriptome trends we are reporting. Human fetal salivary glands were harvested from post-mortem fetuses between 22 and 24 weeks of gestation with the approval of the Institutional Review Board at the University of California San Francisco (IRB# 10-00768). Tissue was identified by location and glandular appearance. Sex was confirmed through analysis of transcript levels of male-specific genes, namely, *UTY* and *KDM5D (67)*.

Table S1. Summary information on the salivary glands used in the study.

	Number of glands and sex distribution		
Gland type	Total number	Male	Female
Adult parotid	4	2	2
Adult submandibular	6	3	3
Adult sublingual	3	0	3
Fetal parotid	3	2	1
Fetal submandibular	5	4	1
Fetal sublingual	6	4	2

Preparation of tissue samples: For RNA analysis, tissue was frozen in liquid nitrogen and stored at -80°C. RNA isolation was conducted as described previously (66, 68). For immunofluorescent analysis, tissue was either flash-frozen in optimal cutting temperature compound (OCT) (Tissue-Tek) and stored at -80°C, or immediately fixed with 4% paraformaldehyde (PFA) overnight at 4°C. Fixed samples were washed with PBS, cryoprotected by immersion in a 12.5 – 25% stepwise sucrose gradient, and then embedded in OCT for storage at -80°C. Tissue was sectioned (12 μm thickness) immediately before immunofluorescent analysis using a cryostat (Leica).

Saliva sample collection: Saliva from healthy humans was collected following the protocol approved by the University at Buffalo Human Subjects IRB board (study # 030–505616). Informed consent was obtained from all human participants. Stimulated whole saliva was collected while chewing on parafilm. Clarified whole saliva supernatant was obtained by centrifugation at 12,000 × g for 15 min at 4°C to remove particulate matter. Parotid and submandibular-sublingual ductal salivary secretions were collected following the stimulation of salivary flow by application of 2% citric acid to the dorsum of the tongue. Parotid saliva was collected from the orifice of Stensons's duct using a modified Carlsen-Crittenden device, and submandibular-sublingual saliva was collected from the floor of the mouth after isolation with absorbent cotton rolls using disposable plastic pasteur pipettes (VWR International, Radnor, PA) as previously described (43). Protein concentrations in saliva samples were determined using the Pierce bicinchoninic acid (BCA) protein assay kit (Thermo Scientific, Rockland, IL), with bovine serum albumin as the standard.

RNA sequencing and analysis: Human adult and fetal tissue samples were mechanically homogenized using a hand-held homogenizer (Thermo Fisher Scientific) and lysed in 500 μL RNA lysis buffer (Ambion) by sonication (1 x 2-4 second pulse, Branson SFX150). RNA was isolated from 3 x 30 μm sections of human adult and fetal tissue using the RNAqueous Micro Kit (Ambion), and total RNA samples were DNase-treated (Ambion). Sample yield and integrity was analyzed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA sequencing was performed by standard operating procedure by GENEWIZ (https://www.genewiz.com/en) using Illumina HiSeq with a 2 x 150 bp configuration. Quality control of the obtained sequences was performed in the Gokcumen laboratory using FastQC (69) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 12/10/2017). The results were further reviewed by MultiQC (70). Adaptor sequences, low-quality bases from both sides of the read (3 bases or smaller), and reads with a length smaller than 36 bp were discarded by Trimmomatic (71) (**Figure S3**). Filtered reads were mapped to the human transcriptome reference (hg19) from Ensembl (64) and biomaRt (72) and quantified using Kallisto (65). Differential expression analysis was performed by DESeq2 (27), which calculates the fold change of transcription of each gene using the Wald test and a correction for multiple hypotheses. We used an adjusted (i.e., multiple hypotheses corrected p-value of < 0.0001) to identify genes that were upregulated (fetus < adults) and downregulated (adults < fetus) during development in each type of salivary gland. Since all adult sublingual glands were of female background, we excluded the Y chromosome from the analyses. The remaining 40,882 genes were used for the following analyses. The RNA abundances for 40,882 genes that we interrogated as well as comparative results are provided in **Table S2**. The RNA-seq data (fastq files) have been submitted to GEO https://www.ncbi.nlm.nih.gov/geo/ with the project name GSE143702. It will be published on July, 10th, 2020.

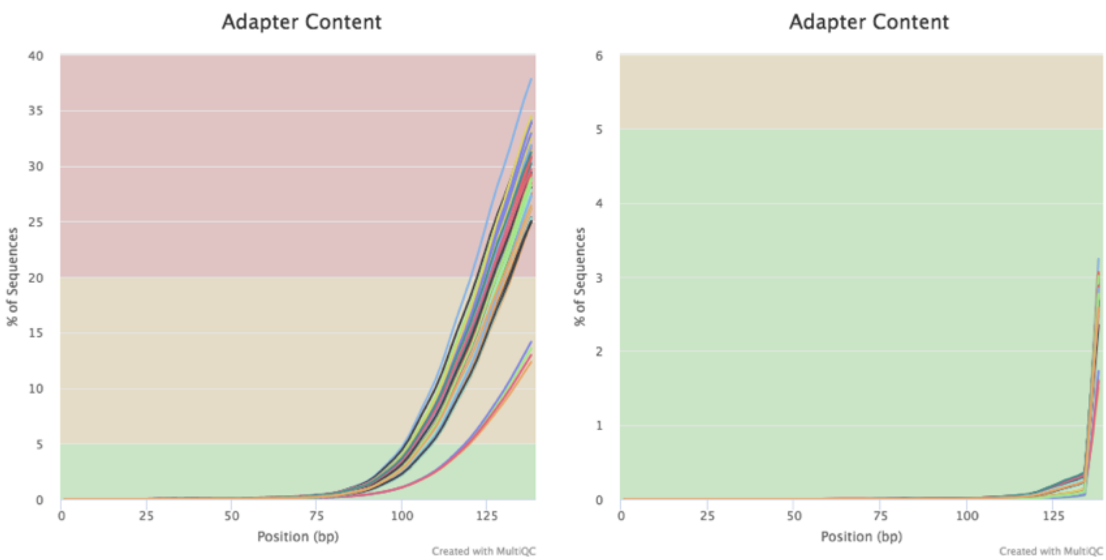


Figure S6 Adapter trimming of RNAseq data by using the sequence analysis tool trimmomatic. Adapter contents before (left panel) and after trimming (right panel) are shown.

Immunofluorescent Analysis: Tissue section immunofluorescence analysis was performed as previously described (66). Frozen adult human salivary sections were fixed with 4 % PFA at room temperature (RT) for 20 min and subsequently washed in PBS followed by permeabilization with 0.5 % Triton-X in PBS for 10 min. Tissue sections were blocked with 10% donkey serum (Jackson Laboratories) and 1% BSA (Sigma-Aldrich) in 0.05% PBS-Tween 20 for 1 h at RT. Tissue sections were incubated with primary antibodies overnight at RT: rabbit anti-MUC7 (1:200; Sigma HPA006411), mouse anti-MUC7 (1:500; Abcam ab105466), mouse anti-MUC5B (1:500; Abcam ab105460), rat anti-E cadherin (1:300; Sigma U3254), rabbit anti-CRISP3 (1:200; Sigma HPA054392), rabbit anti-AMY1A (1:200; Sigma HPA045394), rabbit anti-AQP5 (1:400; Millipore AB3559), rabbit anti-statherin (1:500, Ruhl laboratory*),* and rabbit anti-LPO1 (1:200; Sigma HPA028688). Antibodies were detected by incubating samples with Cy2-, Cy3- or Cy5-conjugated secondary Fab fragments (1:300 in 0.05% PBS-Tween-20;

Jackson Laboratories) for 2 hours at RT. Nuclei were detected using Hoechst 33342 (1:1000, Sigma Aldrich). Fluorescence images were obtained using a Leica Sp5 line-scanning confocal microscope.

SDS-PAGE and immunoblotting: Saliva samples were denatured under reducing conditions. Equal amounts of total protein (15 µg per lane for Coomassie and periodic acid Schiff stain) were subjected to separation by SDS-PAGE using 8-16% gradient Tris-glycine mini gels (Novex, Invitrogen, Carlsbad, CA), and stained for proteins and glycans as previously described (5). Stained gels were imaged using a flat-bed scanner in the transparent mode (ImageScanner III, GE Healthcare). Immunoblotting was performed as previously described (43) except that for electrotransfer a BioRad Trans Turbo Blot apparatus was used. Blots were probed with the following antibodies diluted in tris-buffered saline containing 2% milk (TBS-milk): mouse monoclonal anti-human mucin 7 (MUC7) diluted 1:500 (4D2-1D7, Abcam), rabbit polyclonal anti-human alpha amylase 1A (AMY1) diluted 1:1,000 (HPA045394, Sigma), rabbit polyclonal anti-human parotid secretory protein (SPLUNC2B, BPIFA2) diluted 1:500 (C-20, a gift from Dr. Colin Bingle at the University at Sheffield (61), and rabbit polyclonal anti-CRISP3 diluted 1:500 (HPA054392, Sigma). As secondary antibodies, Alexa Fluor 488 tagged goat anti-rabbit or anti-mouse IgG (Life Technologies) were used diluted 1:1,000 in TBS-milk. Fluorescent bands were detected using a BioRad ChemiDoc imaging system.

Data analysis and integration of datasets: All the downstream data analysis and comparison of our dataset with publicly available databases were conducted using custom bioinformatic pipelines written on Rstudio (v1.2.1335), R(v3.5.3) and ggplot2 (73). These codes were made available through https://github.com/GokcumenLab/glabBits/tree/master/saliva_RNAseq.

To identify genes that are expressed in a salivary gland-specific manner, we compared our RNAseq results from the 3 major salivary glands to RNAseq data available through the GTEx database that were obtained from 53 tissues collected from approximately 1,000 individuals (https://gtexportal.org/home/documentationPage#AboutSamples) (28, 29), (**Table S4**). These organs include those that synthesize secreted proteins such as the pancreas, mammary gland, thyroid, and intestine. To define whether a gene was specific to the salivary glands, we first identified among the non-salivary gland tissues or organs those in which the gene in question was highest expressed. We then compared this level of gene expression to the gene expression levels in each major salivary gland. To account for potential differences in the RNAseq datasets from the GTEx database due to different experimental platforms and bioinformatics data analysis processes, we compared “relative expression level ranks” by using a ($\log_{10}$ (1+normalized read counts)) transformation for each organs RNAseq dataset.

Functional categorization of genes was determined by cross-checking using Gene Ontology Resources (http://geneontology.org/) (32). Then, an enrichment analysis was conducted using GOrilla (74) using default settings that enable multiple hypothesis testing (**Table S3**). Secreted protein gene annotations were obtained from the Human Protein Atlas (https://www.proteinatlas.org/) with the query: “protein_class:Predicted secreted proteins NOT protein_class:Predicted membrane proteins” (30).

We used the Human Salivary Proteome Wiki (https://salivaryproteome.nidcr.nih.gov/) to obtain proteome data for the whole saliva, ductal saliva from the parotid gland, submandibular gland, and sublingual gland, as well as from blood plasma. The database provides mass-spectrometry-based abundance data for approximately 3,000 proteins compiled from multiple studies (20, 75). Similar to the comparison approach that we used for integrating GTEx data, we used log transformation ($\log_{10}$ (1+normalized abundance)) for each dataset. In addition, we used mass-spectrometry-based protein abundances available through Human Protein Atlas from 29 tissues curated recently by Wang et al. (52) (**Table S4**). This dataset contains abundance information on 13,000 proteins. In addition, it provides conformational analyses on antibodies used in this study.

Table S4. The tissues of origins for RNA (GTEx) and proteome (Human Protein Atlas (52)**) comparison databases used in this study.**

GTEx (RNA)	Human Protein Atlas
Adipose - Subcutaneous	Adipose Tissue
Adipose - Visceral (Omentum)	
Adrenal Gland	Adrenal Gland
	Appendix
Artery - Aorta	
Artery - Coronary	
Artery - Tibial	
Bladder	Urinary Bladder
Brain - Amygdala	
Brain - Anterior cingulate cortex (BA24)	Brain
Brain - Caudate (basalganglia)	
Brain - Cerebellar Hemisphere	
Brain - Cerebellum	
Brain - Cortex	
Brain - Frontal Cortex(BA9)	

Brain - Hippocampus	
Brain - Hypothalamus	
Brain - Nucleus accumbens (basalganglia)	
Brain - Putamen (basalganglia)	
Brain - Spinalcord (cervicalc1)	
Brain - Substantia nigra	
Breast - Mammary- Tissue	
Cells - EBV - transformed lymphocytes	
Cells - Transformed - fibroblasts	
Cervix - Ectocervix	
Cervix - Endocervix	
Colon - Sigmoid	Colon
Colon - Transverse	
	Duodenum
Esophagus - Gastroesophageal Junction	
Esophagus - Mucosa	Esophagus
Esophagus - Muscularis	
Fallopian Tube	Fallopian Tube
	Gallbladder
Heart - Atrial Appendage	Heart
Heart - Left Ventricle	
Kidney - Cortex	Kidney
Liver	Liver
Lung	Lung

	Lymph Node
Minor Salivary Gland	Parotid
Muscle Skeletal	
Nerve Tibial	
Ovary	Ovary
Pancreas	Pancreas
Pituitary	Pituitary
	Placenta
Prostate	Prostate
	Rectum
Skin - Not Sun Exposed (Suprapubic)	
Skin - Sun Exposed (Lower leg)	
Small Intestine - Terminal Ileum	Small Intestine
	Smooth Muscle
Spleen	Spleen
Stomach	Stomach
Testis	Testis
Thyroid	Thyroid
	Tonsil
Uterus	Uterine Endometrium
Vagina	
Whole Blood	

REFERENCES

1. S. Ruhl, The scientific exploration of saliva in the post-proteomic era: from database back to basic function. *Expert Rev. Proteomics* **9**, 85–96 (2012).

2. C. Dawes, *et al.*, The functions of human saliva: A review sponsored by the World Workshop on Oral Medicine VI. *Arch. Oral Biol.* **60**, 863–874 (2015).

3. C. Ragunath, *et al.*, Probing the role of aromatic residues at the secondary saccharide-binding sites of human salivary alpha-amylase in substrate hydrolysis and bacterial binding. *J. Mol. Biol.* **384**, 1232–1248 (2008).

4. E. C. Moreno, K. Varughese, D. I. Hay, Effect of human salivary proteins on the precipitation kinetics of calcium phosphate. *Calcif. Tissue Int.* **28**, 7–16 (1979).

5. S.-M. Heo, *et al.*, Host defense proteins derived from human saliva bind to *Staphylococcus aureus*. *Infect. Immun.* **81**, 1364–1373 (2013).

6. B. W. Cross, S. Ruhl, Glycan recognition at the saliva - oral microbiome interface. *Cell. Immunol.* **333**, 19–33 (2018).

7. L. A. Tabak, M. J. Levine, I. D. Mandel, S. A. Ellison, Role of salivary mucins in the protection of the oral cavity. *J. Oral Pathol.* **11**, 1–17 (1982).

8. E. S. Frenkel, K. Ribbeck, Salivary mucins in host defense and disease prevention. *J. Oral Microbiol.* **7**, 29759 (2015).

9. A. C. Poole, *et al.*, Human salivary amylase gene copy number impacts oral and gut microbiomes. *Cell Host Microbe* **25**, 553–564.e7 (2019).

10. D. Xu, *et al.*, Archaic hominin introgression in Africa contributes to functional salivary MUC7 genetic variation. *Mol. Biol. Evol.* **34**, 2704–2715 (2017).

11. P. Pajic, *et al.*, Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *Elife* **8** (2019).

12. E. J. Helmerhorst, F. G. Oppenheim, Saliva: a dynamic proteome. *J. Dent. Res.* **86**, 680–693 (2007).

13. C. P. Mavragani, H. M. Moutsopoulos, Sjögren's syndrome: Old and new therapeutic targets. *J. Autoimmun.*, 102364 (2019).

14. A. Vissink, P. van Luijk, J. A. Langendijk, R. P. Coppes, Current ideas to reduce or salvage radiation damage to salivary glands. *Oral Dis.* **21**, e1–10 (2015).

15. D. Heller, E. J. Helmerhorst, F. G. Oppenheim, Saliva and serum protein exchange at the tooth enamel surface. *J. Dent. Res.* **96**, 437–443 (2017).

16. W. Yan, *et al.*, Systematic comparison of the human saliva and plasma proteomes.

Proteomics Clin. Appl. **3**, 116–134 (2009).

17. R. J. Lamont, H. Koo, G. Hajishengallis, The oral microbiota: dynamic communities and host interactions. *Nat. Rev. Microbiol.* **16**, 745–759 (2018).

18. J. L. Mark Welch, B. J. Rossetti, C. W. Rieken, F. E. Dewhirst, G. G. Borisy, Biogeography of a human oral microbiome at the micron scale. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E791–800 (2016).

19. N. Grassl, *et al.*, Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome. *Genome Med.* **8**, 44 (2016).

20. P. Denny, *et al.*, The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions. *J. Proteome Res.* **7**, 1994–2006 (2008).

21. Y. Ogawa, *et al.*, Proteomic analysis of two types of exosomes in human whole saliva. *Biol. Pharm. Bull.* **34**, 13–23 (2011).

22. M. Castagnola, *et al.*, Salivary biomarkers and proteomics: future diagnostic and clinical utilities. *Acta Otorhinolaryngol. Ital.* **37**, 94–101 (2017).

23. J. A. Loo, W. Yan, P. Ramachandran, D. T. Wong, Comparative human salivary and plasma proteomes. *J. Dent. Res.* **89**, 1016–1023 (2010).

24. K. Thomadaki, *et al.*, Whole-saliva proteolysis and its impact on salivary diagnostics. *J. Dent. Res.* **90**, 1325–1330 (2011).

25. S. Robinson, *et al.*, A mass spectrometry-based strategy for detecting and characterizing endogenous proteinase activities in complex biological samples. *Proteomics* **8**, 435–445 (2008).

26. M. Uhlen, *et al.*, Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).

27. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

28. GTEx Consortium, *et al.*, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

29. GTEx Consortium, The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

30. M. Uhlén, *et al.*, Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

31. M. Magacz, K. Kędziora, J. Sapa, W. Krzyściak, The significance of lactoperoxidase system in oral health: Application and efficacy in oral hygiene products. *Int. J. Mol. Sci.* **20** (2019).

32. M. Ashburner, *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology

Consortium. *Nat. Genet.* **25**, 25–29 (2000).

33. The Gene Ontology Consortium, Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017).
34. N. McGovern, *et al.*, Human fetal dendritic cells promote prenatal T-cell immune suppression through arginase-2. *Nature* **546**, 662–666 (2017).
35. A. Le, M. Saverin, A. R. Hand, Distribution of dendritic cells in normal human salivary glands. *Acta Histochem. Cytochem.* **44**, 165–173 (2011).
36. N. Takahashi, T. Horie, Heerfordt syndrome. *Nihon Rinsho* **60**, 1822–1826 (2002).
37. R. Gupta, D. Balasubramanian, J. R. Clark, Salivary gland lesions: recent advances and evolving concepts. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **119**, 661–674 (2015).
38. N. S. Goud, R. K. Joshi, R. D. Bharath, P. Kumar, Fluorine-18: A radionuclide with diverse range of radiochemistry and synthesis strategies for target based PET diagnosis. *Eur. J. Med. Chem.* **187**, 111979 (2020).
39. S. Ahmad Sarji, Physiological uptake in FDG PET simulating disease. *Biomed. Imaging Interv. J.* **2**, e59 (2006).
40. B. S. Purohit, *et al.*, FDG-PET/CT pitfalls in oncological head and neck imaging. *Insights Imaging* **5**, 585–602 (2014).
41. F. N. Nogueira, R. A. Carvalho, Metabolic remodeling triggered by salivation and diabetes in major salivary glands. *NMR Biomed.* **30** (2017).
42. J. M. Gutierrez, *et al.*, Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nat. Commun.* **11**, 68 (2020).
43. S. Thamadilok, *et al.*, Human and non-Human primate lineage-specific footprints in the salivary proteome. *Mol. Biol. Evol.* (2019).
44. G. H. Perry, *et al.*, Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260 (2007).
45. D. Xu, *et al.*, Recent evolution of the salivary mucin MUC7. *Sci. Rep.* **6**, 31791 (2016).
46. Y. Liu, A. Beyer, R. Aebersold, On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
47. D. I. Hay, The isolation from human parotid saliva of a tyrosine-rich acidic peptide which exhibits high affinity for hydroxyapatite surfaces. *Arch. Oral Biol.* **18**, 1531–1541 (1973).
48. C. Hannig, M. Hannig, T. Attin, Enzymes in the acquired enamel pellicle. *Eur. J. Oral Sci.* **113**, 2–13 (2005).
49. J. Li, *et al.*, Statherin is an in vivo pellicle constituent: identification and immuno-quantification. *Arch. Oral Biol.* **49**, 379–385 (2004).

50. A. Walz, *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* **9**, 1582–1592 (2009).

51. A. Walz, *et al.*, Proteome analysis of glandular parotid and submandibular-sublingual saliva in comparison to whole human saliva by two-dimensional gel electrophoresis. *Proteomics* **6**, 1631–1639 (2006).

52. D. Wang, *et al.*, A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).

53. E. J. Helmerhorst, C. Dawes, F. G. Oppenheim, The complexity of oral physiology and its impact on salivary diagnostics. *Oral Dis.* **24**, 363–371 (2018).

54. P. Sharma, *et al.*, MUC5B and MUC7 are differentially expressed in mucous and serous cells of submucosal glands in human bronchial airways. *Am. J. Respir. Cell Mol. Biol.* **19**, 30–37 (1998).

55. J. H. Jung, *et al.*, Proteomic analysis of human lacrimal and tear fluid in dry eye disease. *Sci. Rep.* **7**, 13363 (2017).

56. O. Amano, K. Mizobe, Y. Bando, K. Sakiyama, Anatomy and histology of rodent and human major salivary glands: -overview of the Japan salivary gland society-sponsored workshop-. *Acta Histochem. Cytochem.* **45**, 241–250 (2012).

57. N. I. Tapinos, M. Polihronis, G. Thyphronitis, H. M. Moutsopoulos, Characterization of the cysteine-rich secretory protein 3 gene as an early-transcribed gene with a putative role in the pathophysiology of Sjögren's syndrome. *Arthritis & Rheumatism* **46**, 215–222 (2002).

58. E. C. Veerman, P. A. van den Keybus, A. Vissink, A. V. Nieuw Amerongen, Human glandular salivas: their separate collection and analysis. *Eur. J. Oral Sci.* **104**, 346–352 (1996).

59. A. D. Merritt, M. L. Rivas, D. Bixler, R. Newell, Salivary and pancreatic amylase: electrophoretic characterizations and genetic studies. *Am. J. Hum. Genet.* **25**, 510–522 (1973).

60. S. Thamadilok, H. Roche-Håkansson, A. P. Håkansson, S. Ruhl, Absence of capsule reveals glycan-mediated binding and recognition of salivary mucin MUC7 by *Streptococcus pneumoniae*. *Mol. Oral Microbiol.* **31**, 175–188 (2016).

61. L. Bingle, *et al.*, Characterisation and expression of SPLUNC2, the human orthologue of rodent parotid secretory protein. *Histochem. Cell Biol.* **132**, 339–349 (2009).

62. L. Udby, *et al.*, An ELISA for SGP28/CRISP-3, a cysteine-rich secretory protein in human neutrophils, plasma, and exocrine secretions. *J. Immunol. Methods* **263**, 43–55 (2002).

63. Y. Zhou, *et al.*, Differential utilization of basic proline-rich glycoproteins during growth of oral bacteria in saliva. *Appl. Environ. Microbiol.* **82**, 5249–5258 (2016).

64. D. R. Zerbino, *et al.*, Ensembl 2018. *Nucleic Acids Res.* (2017)

https:/doi.org/10.1093/nar/gkx1098.

65. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

66. E. Emmerson, *et al.*, Salivary glands regenerate after radiation injury through SOX2-mediated secretory cell replacement. *EMBO Mol. Med.* **10** (2018).

67. F. Staedtler, *et al.*, Robust and tissue-independent gender-specific transcript biomarkers. *Biomarkers* **18**, 436–445 (2013).

68. E. Emmerson, *et al.*, SOX2 regulates acinar cell development in the salivary gland. *Elife* **6** (2017).

69. S. W. Wingett, S. Andrews, FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **7**, 1338 (2018).

70. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

71. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

72. S. Durinck, *et al.*, BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).

73. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis*, 2nd Ed. (Springer Publishing Company, Incorporated, 2009).

74. E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).

75. A. Murr, *et al.*, Cross-sectional association of salivary proteins with age, sex, body mass index, smoking, and education. *J. Proteome Res.* **16**, 2273–2281 (2017).