Gene Expansion and Regulatory Rewiring Shape Sex-Biased Evolution of the Mouse Submandibular Gland Secretome

AUTHORS

Luane Jandira Bueno Landau[1], Shikha Jain[2], Nathan Griffin[3], Achisha Saikia[4], Jill M. Kramer[2], Sarah Knox[3], Stefan Ruhl[2], Omer Gokcumen[1]*

AFFILIATIONS

[1]Department of Biological Sciences, University at Buffalo, The State University of New York, NY 14260, USA.

[2]Department of Oral Biology, School of Dental Medicine, University at Buffalo, The State University of New York, NY, 14214, USA.

[3]Department of Cell & Tissue Biology, University of California, San Francisco, CA, 94143, USA.

[4]Department of Genetics, Genomics and Bioinformatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, The State University of New York, NY 14203, USA.

*correspondence:

Omer Gokcumen: omergokc@buffalo.edu

ABSTRACT

Mammalian saliva plays essential roles in digestion, immunity, and host-microbiome interactions, yet its protein composition varies across species and sexes. The evolutionary mechanisms underlying this molecular diversity remain poorly understood. Here, we compared mouse and human salivary gland secretomes at genomic, transcriptomic, and proteomic levels to understand how saliva composition evolves. We performed RNA-seq analysis of the major mouse salivary glands (parotid, submandibular, and sublingual), liver and pancreas from both sexes, compared them with reanalyzed previously published human salivary gland transcriptomes, and integrated them with proteome data of mouse and human whole saliva. We found that evolution of gene expression in mouse salivary glands is driven by rapid gene turnover and sexual dimorphism. In the submandibular and sublingual glands, respectively, 68% and 73% of expression from genes encoding secreted proteins derives from lineage-specific genes that lack one-to-one human orthologs. Mouse submandibular gland shows striking sexual dimorphism, with 1537 tissue specific sex-biased genes, five times higher than in the liver, a classic model of sex-biased expression. These genes cluster in regions shaped by recent gene duplication, such as the kallikrein gene cluster, a mouse-specific expansion that accounts for ~16.4% of male-biased submandibular expression. Our analyses suggest that this bias arises through regulatory changes that are expanded by gene duplication, including the spread of a testosterone-associated regulatory motif and the expansion of a shared chromatin domain that promotes coordinated gene regulation. Our results reveal how lineage-specific gene duplication and regulatory rewiring drive rapid, sex-specific evolution of the mammalian salivary gland secretome.

INTRODUCTION

Diet and pathogen exposure are major environmental drivers of mammalian evolution (Price et al. 2012; Enard et al. 2016). Because the oral cavity represents the body's first interface with the external environment, saliva plays a central role in mediating these pressures, balancing protection against pathogens with responsiveness to dietary variation. Saliva also contributes to oral homeostasis by lubricating and cleansing the mouth, protecting tooth enamel, enabling taste perception, and initiating digestion (Ruhl 2012; Pajic et al. 2022). These functions are carried out by a complex mixture of proteins secreted primarily by the three major salivary glands, parotid, submandibular, and sublingual glands (Dawes et al. 2015; Chibly et al. 2022). Importantly, the salivary gland secretome, i.e., the entirety of proteins secreted by the salivary glands, reflects adaptive tuning to species-specific ecological contexts, with protein repertoires shaped by dietary specialization (Pajic et al. 2019) and pathogen exposure (Xu et al. 2016). Together, the integration of immune, digestive, and sensory functions makes saliva an exceptional system for investigating how environmental pressures drive the evolution of mammalian genomes.

Previous studies suggested that saliva-related genes undergo rapid evolution, mainly to accommodate environmental and dietary changes which accompany speciation (Pajic et al. 2019). Recent locus-specific comparative genomic studies showed that genes expressed in the salivary glands that encode secreted salivary proteins differ substantially between mice and humans (Pajic et al. 2025). For instance, analyses of the secretory calcium-binding phosphoprotein (SCPP) locus indicated that multiple human salivary genes lack orthologs in mice. Proteomic analyses also have revealed that the composition of saliva proteins differs markedly between humans and mice (Blanchard et al. 2015; De Sousa-Pereira et al. 2015), and comparative studies further show substantial interspecies differences in their major components (Karn et al. 2013). These findings suggest that human and mouse salivary glands differ greatly in both gene content and expression profiles, supporting the view that the salivary gland secretome evolves rapidly in response to environmental and pathogenic pressures. Despite these insights, a systematic, secretome-wide investigation of the evolutionary dynamics and regulatory mechanisms underlying these differences has not yet been conducted.

In contrast to humans, the salivary glands in many other mammalian species are different between sexes. Sexual dimorphism in salivary glands has been well documented in rodents (Ikemoto and Matsushima 1984), and pronounced sex differences have been shown in salivary glands of mice (Chai et al. 1993), which are not known to occur in humans. Generally, sex-biased genes across multiple tissues in mammals are poorly conserved across species (Naqvi et al. 2019). In particular, the mouse submandibular gland exhibits striking sex-specific differences in cellular composition, secretory activity, and hormonal regulation (Hatakeyama et al. 1987; Konttinen et al. 2010; Mukaibo et al. 2019). Although substantial progress has been

made in understanding saliva composition and its role in oral biology in both species, the evolutionary basis of lineage- and sex-specific variation in the salivary gland secretome remains underexplored.

Here, we investigate the mechanisms underlying the rapid diversification of the salivary gland secretome. Using comparative analyses between humans and mice, we assess the conservation of genes highly expressed in salivary glands and determine the contribution of lineage-specific genes to the salivary gland secretome. We further compare male and female mice to identify the genomic basis of sex-biased expression in the submandibular gland. By integrating transcriptomic, proteomic, and genomic data, we delineate the evolutionary processes driving salivary diversity and sexual dimorphism.

RESULTS

The highest proportion of gene expression in mouse submandibular and sublingual glands derives from lineage-specific genes

To investigate differences in salivary gene expression across commonly used mouse models and compare them to humans, we performed transcriptomic profiling of the parotid, submandibular, and sublingual glands, along with the pancreas and liver as reference tissues, in CD1 and C57BL/6 strains (n = 3 per sex per strain; **Tables S1**). Both CD1 and C57BL/6 strains are widely used in laboratory studies of salivary glands, and were included as a representative of an inbred and outbred mouse strain, respectively, to strengthen the broader applicability of our findings (Nguyen et al. 2009; Gervais et al. 2015; Hsieh et al. 2017). In parallel, we conducted mass spectrometry-based proteomic profiling of whole saliva from ten C57BL/6 mice, five males and five females (**Table S2**). Principal component analysis (PCA) of the transcriptomic data shows that samples cluster strongly by tissue of origin, with clear separation among the three salivary glands (**Figure S1**), supporting the robustness of the dataset.

To characterize gland-specific expression patterns and contribution to saliva proteome, we calculated mean transcript abundance herein estimated in transcripts per million (TPM). Each gland displayed a distinct transcriptional signature, typically dominated by a small subset of highly expressed, gland-specific genes that code for secreted proteins, consistent with previous work (Gao et al. 2018). We then identified the 30 most highly expressed genes encoding secreted proteins in each mouse gland (see **Methods**; **Figure 1a, Table S3**). These genes account for 69.3%, 64.9%, and 35.6% of total transcript abundance in the parotid, submandibular, and sublingual glands, respectively. When considering only transcripts encoding secreted proteins, these top 30 genes explain an even larger fraction of expression: 95.7%, 95.9%, and 87.0% in the respective glands.

Next, we categorized mouse salivary gland genes based on their evolutionary relationships with human genes, either as one-to-one orthologs, other orthologous relationships (i.e., one-to-many or many-to-many; see **Methods**), or lineage-specific genes (**Figure 1a**). We applied the same classification to human salivary gland transcriptomes from our previous study (Saitou et al. 2020), enabling direct comparison between species (**Figure 1b, Table S4**). The proportion of expression originating from one-to-one orthologs that code for secreted proteins are remarkably low for sublingual and submandibular glands, approximately 8% and 11%, respectively (**Figure 1c**). In contrast, the parotid gland displays a higher proportion of expression of one-to-one orthologs (77.58%), suggesting greater evolutionary conservation in its core secretory functions.

We recently hypothesized, based on a locus-specific analysis, that salivary glands evolve faster than other tissues because of shifting dietary and pathogenic evolutionary pressures (Pajic et al. 2025). Here, we tested this hypothesis more

136 comprehensively at the transcriptome-wide level by comparing the expression levels
137 of one-to-one orthologs between human and mouse, which are expected to maintain
138 comparable expression levels across species (Kryuchkova-Mostacci and Robinson-
139 Rechavi 2016). This expectation was met for genes that code for secreted proteins in
140 pancreas and liver, which showed strong correlation of expression between human
141 and mouse transcriptomes (Spearman's $\rho = 0.64$, $p = 1.68 \times 10^{-147}$ and $\rho = 0.78$, $p =$
142 8.55×10^{-267}, respectively). In contrast, genes that code for secreted proteins in the
143 salivary glands exhibited weaker mouse-human correlations (parotid: $\rho = 0.575$, $p =$
144 5.90×10^{-112}; submandibular: $\rho = 0.58$, $p = 2.71 \times 10^{-114}$; sublingual: $\rho = 0.606$, $p =$
145 2.27×10^{-127}; **Figure S2**).

146 To determine whether these transcriptomic differences for genes encoding secreted
147 proteins are also reflected at the proteomic level, we performed gel electrophoresis of
148 whole saliva (**Figure 1d**; **Figure S3**). The resulting protein and glycoprotein profiles
149 revealed striking qualitative and quantitative differences in banding patterns between
150 human and mouse saliva, with only a few bands appearing to match between the two
151 species. Particularly noteworthy is the observed shift in gel electrophoretic mobility of
152 the high-molecular-weight glycoprotein band from 150 kDa in males to 130 kDa in
153 females. This protein band was identified by mass spectrometry (MS/MS) analysis as
154 Muc19 (**Table S5**).

155 In parallel, we asked the extent to which the transcriptomic variation observed
156 translates into variation in the salivary proteome. To assess this, we compared gene
157 expression in the salivary glands of each mouse strain with the abundance of the
158 corresponding proteins in their saliva for 467 gene-protein pairs (see **Methods**). We
159 observed a significant positive correlation for each gland (parotid: $\rho = 0.362$, $p = 7.29$
160 $\times 10^{-16}$; sublingual: $\rho = 0.268$, $p = 3.76 \times 10^{-9}$; submandibular: $\rho = 0.304$, $p = 1.81 \times$
161 10^{-11}) (**Figure S4**). Notably, correlation strength increased among the most highly
162 expressed genes that code for secreted proteins, with stronger associations observed
163 when restricting the analysis to the top 100 transcripts per gland (parotid: $\rho = 0.492$, p
164 $= 2.06\times10^{-7}$; sublingual: $\rho = 0.528$, $p = 1.65\times10^{-8}$; submandibular: $\rho = 0.582$, $p =$
165 2.10×10^{-10}) (**Figure 1e**). These protein-level analyses support the notion that transcript
166 abundance in the salivary glands directly affects protein levels in saliva, and that highly
167 expressed genes encoding secreted proteins have an outsized influence on shaping
168 the oral proteome.

169 Our combined transcriptomic and proteomic analyses support a model in which
170 mammalian salivary glands function as specialized organs that produce a limited set
171 of highly abundant secreted proteins. Our data further support the hypothesis that the
172 genes coding for secreted proteins in the salivary glands show faster turnover than
173 those in liver and pancreas, a phenomenon more pronounced among the top
174 expressed genes encoding secreted proteins. Transcriptomic data from individual
175 salivary glands accurately predict their relative contributions to the oral proteome,
176 consistent with previous observations in humans (Saitou et al. 2020). Overall, our

177 results reveal fundamental differences in expression profiles of genes coding for
178 secreted proteins in the salivary glands of humans and mice, consistent with a rapid
179 evolution of the salivary gland secretome.

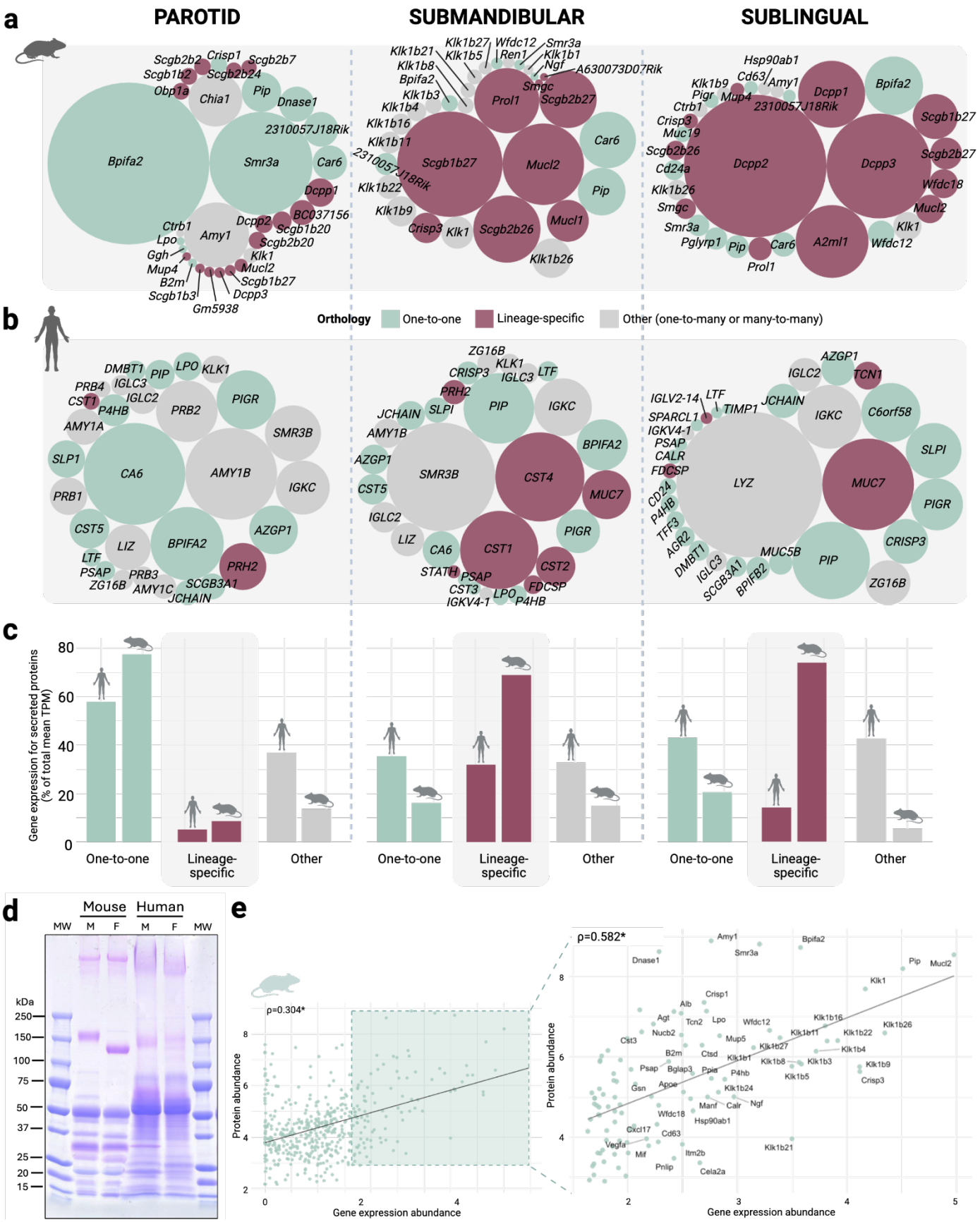

Figure 1. Most mouse genes coding for secreted proteins in salivary glands are lineage-specific. **(a)** Bubble plots showing the top 30 most highly transcribed genes that code for secreted proteins in each

major mouse salivary gland. Bubble size represents mean transcript abundance herein estimated in transcripts per million (TPM), and colors denote the type of orthology: one-to-one (green), lineage-specific (maroon), and other orthologous relationships (including many-to-many and one-to-one) (gray). **(b)** Same analysis as in A, here for human salivary glands. (**c**) Proportion of total gene expression (mean TPM) of genes that code for secreted proteins attributed to each orthology category in mouse and human. **(d)** Saliva proteins separated by SDS-PAGE and stained with Coomassie blue and Periodic acid Schiff's stain for pooled whole saliva of males (M, n=5) and females (F, n=5) for mice and humans. Pink bands are glycosylated proteins revealed by periodic acid Schiff stain. Note the shift in gelelectrophoretic mobility of the high-molecular-weight glycoprotein band from 150 kDa in males to 130 kDa in females. MW denotes molecular weight markers. **(e)** Correlation between gene expression (Log_{10}(1+Mean TPM)) and protein abundance (Log_{10}(1+Mean iBAQ)) for genes that code for secreted proteins in the submandibular gland. The right panel highlights the top 100 most highly expressed genes, for which RNA and protein expression are more concordant. ρ denotes Spearman's correlation. *denotes p<0.05.

The submandibular gland is the major contributor to sexual dimorphism in saliva

Previous studies have documented sexual dimorphism in the postnatal development and cellular composition of the mouse submandibular gland, most notably the male-specific expansion of granular convoluted tubules (ducts) (Caramia 1966; Pardini and Taga 1996). These morphological differences are known to influence salivary protein output and gene expression (Treister et al. 2005; Mukaibo et al. 2019). Thus, we hypothesize that the submandibular gland contributes to sexual dimorphism in the salivary proteome. Analysis of our mass spectrometry dataset identified 368 and 152 proteins with significantly male- and female-biased protein abundances, respectively (p_{adj}<0.05 and |log_2FoldChange| > 0.5, **Table S6**). To determine the transcriptional basis of this protein-level dimorphism, we next examined whether sex-biased gene expression across individual salivary glands contributes to the observed differences in salivary protein abundance.

To systematically address this hypothesis, we performed comparative transcriptomic analyses of sex-biased gene expression across mouse salivary glands, liver, and pancreas (**Table S7**). The submandibular gland exhibited the largest number of tissue-specific differentially expressed genes (DEGs, FoldEnrichment>1, p_{adj}<0.05) between sexes (1,537 genes) with other salivary glands exhibiting markedly fewer tissue-specific DEGs (**Figure 2a**), consistent with previous reports (Mukaibo et al. 2019). What distinguishes our analysis is the direct cross-organ comparison. We found that the submandibular gland, analyzed within the same individuals, harbors more than fivefold sex-biased DEGs than liver and ninety-sixfold more than pancreas. These results place the submandibular gland among the most sexually dimorphic tissues in mice, clearly exceeding even the liver, which is widely regarded as the canonical model for studying sexual dimorphism in mammalian gene expression (Lai et al. 2016; Nikkanen et al. 2022). Further, a direct comparison of the extent of sexual differentiation in submandibular transcriptomes shows a dramatic difference between mice and humans (**Figure 2b, Table S8**).

228 Next, we tested if most sexually dimorphic saliva proteins originate from genes with
229 sex-biased expression in the submandibular gland. Protein abundance in saliva is
230 influenced by many factors, including RNA abundance, translational efficiency,
231 posttranslational modifications, and peptide degradation. Despite this complexity, we
232 found that 70 of the 507 sexually dimorphic salivary proteins are encoded by genes
233 that are differentially expressed in the submandibular gland, compared with only one
234 gene in the parotid gland and one in the sublingual gland (**Figure 2c**). These results
235 show that the submandibular gland is the primary contributor to the sexual dimorphism
236 observed in the protein composition of mouse saliva.
237
238 To further validate these results, we used immunohistochemical staining with well-
239 characterized antibodies to examine the expression of nerve growth factor (*Ngf*), one
240 representative sex-biased protein, which showed pronounced sexual dimorphism in
241 both submandibular RNA-seq and salivary proteome analyses (Bie et al. 2010; Wong
242 et al. 2019). Examination of submandibular gland sections from male and female mice
243 revealed clear immunohistochemical differences between sexes (**Figure 2d**),
244 supporting that pronounced sexual dimorphism at the gene expression and secreted
245 protein levels, are also manifested at the cellular level.
246
247 In addition to transcriptional differences, post-translational modifications, such as
248 glycosylation, may further contribute to sex-specific differences in the biochemical and
249 functional properties of salivary proteins. In particular, sialic acids are negatively
250 charged terminal glycan residues on most mammalian glycan chain and are commonly
251 present on glycans of secreted glycoproteins, most prominently on mucins, where they
252 influence the biophysical and tribological properties of these molecules and play
253 important roles in their interaction with microbes and viruses (Varki and Schauer 2009;
254 Hiremath and Wagner 2026). Consistent with this possibility, we observed differential
255 expression of several genes involved in sialylation pathways, including the
256 sialyltransferases *St3gal1*, *St3gal4*, *St3gal5*, *St6galnac1*, and *St6galnac2*, between
257 males and females (**Figure S5**). Notably, in contrast to mice, in humans none of these
258 sialyltransferases are differentially expressed between sexes in the submandibular
259 gland.
260
261 To test whether these transcriptional differences correspond to differences in protein
262 glycosylation, we investigated the prominent glycoprotein band in mouse saliva that
263 shifts in electrophoretic mobility from ~150 kDa in male saliva to ~130 kDa in female
264 saliva (see **Figure 1d**). Analysis by MS/MS of the major protein present in these bands
265 excised from the gel identified it as mucin Muc19. Lectin blotting using PNA, revealed
266 strong binding to the band in male saliva, but lack of binding to the band in females,
267 whereas lectin blotting with MAL-II showed strong binding to the band in females and
268 weak binding to the band in male saliva. Together, these results indicate presence of
269 unsialylated core O-glycans (Galβ1-3GalNAc) on male Muc19 and a higher proportion
270 terminal α2,3-linked sialylation on female Muc19 (**Figure 2e**). The higher degree of
271 Muc19 sialylation in female mice is consistent with the observed electrophoretic shift

of the Muc19 band from ~150 kDa in male saliva to ~130 kDa in female saliva due to the added electronegativity conferred by the higher number of negatively charged sialic acid molecules on the female mucin glycan chains. These results suggest that, in addition to sex-specific differences affecting the transcription of genes encoding secreted proteins in the submandibular gland directly, sex-specific post-translational modifications also contribute substantially to the sexual dimorphism observed in mouse saliva, and are caused by sex-specific transcription differences in modifying enzymes, thereby affecting secreted protein products indirectly in a sex-specific manner.

Together, these findings demonstrate that the submandibular gland, by far the most sexually dimorphic tissue in our dataset, is the primary contributor to the strong sexual dimorphism observed in the salivary proteome, with additional contributions from sex-specific post-translational modifications.

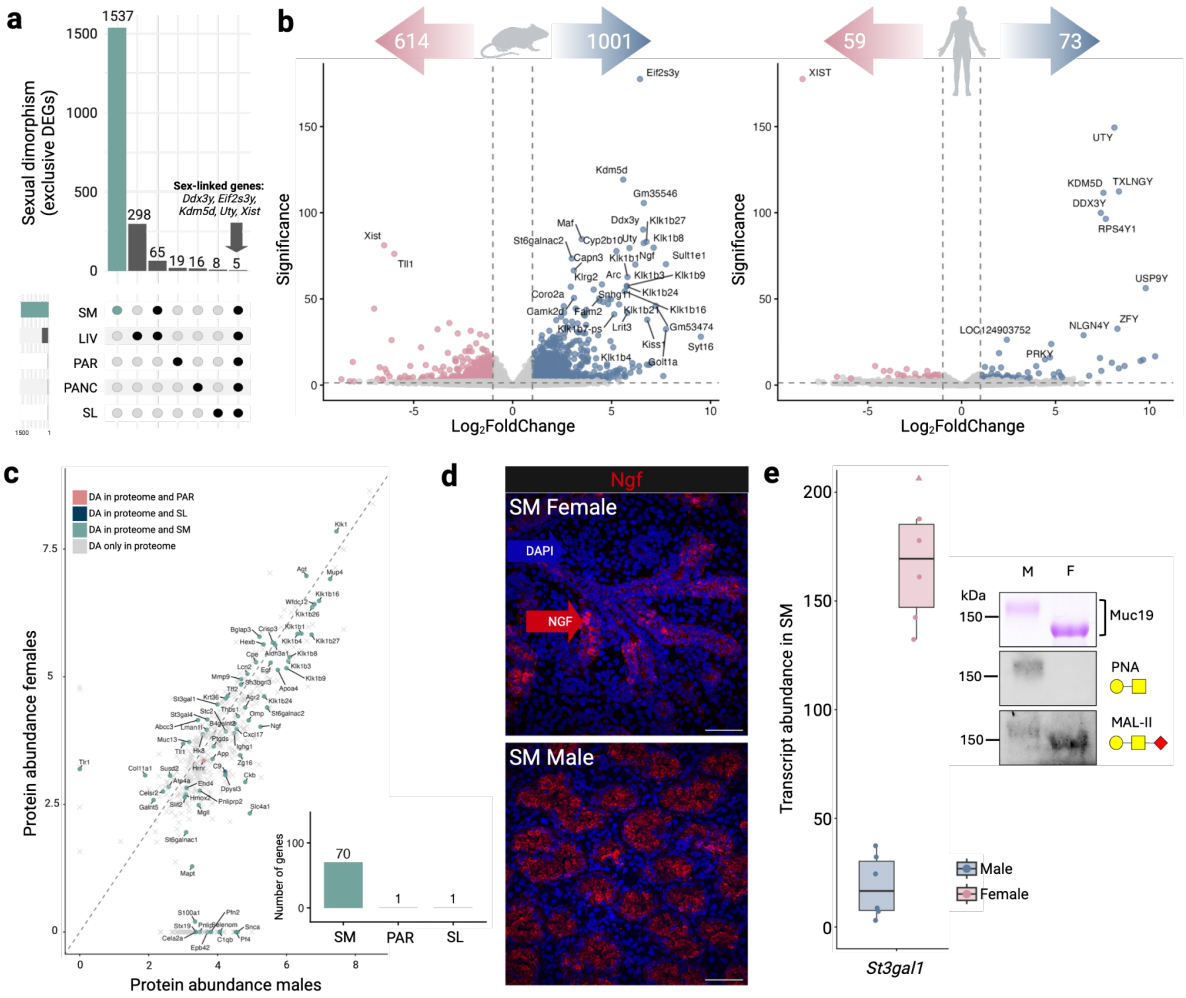

Figure 2. The submandibular gland is the major contributor to sexual dimorphism of the salivary gland secretome. **(a)** Upset plot showing that the submandibular (SM) gland harbors more exclusive DEGs between mouse sexes than other major salivary glands (parotid [PAR], sublingual [SL]) or non-salivary tissues (liver [LIV], pancreas [PANC]). Sex-linked genes (*Ddx3y*, *Eif2s3y*, *Kdm5d*, *Uty*, *Xist*) that are differentially expressed in all tissues are indicated with an arrow. **(b)** Volcano plot of differentially

291 expressed genes in the mouse (left) and human (right) submandibular gland. Arrows indicate the total
292 number of significantly upregulated genes in females (pink) and males (blue). Pink dots represent genes
293 significantly upregulated in females, and blue dots represent genes significantly upregulated in males.
294 Significance is measured in $Log_{10}(p_{adj})$. **(c)**. Correlation plot of protein abundances in males and females
295 quantified in Log_{10}(1+Mean iBAQ). Each dot represents a differentially abundant protein. Female protein
296 abundance is shown on the y-axis and male protein abundance on the x-axis. The dashed line indicates
297 one-to-one correspondence. Colored points denote sexually dimorphic proteins that are also
298 differentially expressed genes in each gland. The inset panel shows the number of differentially
299 abundant proteins whose genes are also differentially expressed in each gland, with the highest
300 contribution observed in the submandibular gland. **(d)** Immunofluorescence staining showing
301 differences in expression patterns between female (top) and male (bottom) submandibular glands. NGF
302 (red). DAPI marks nuclei (blue). **(e)** Sex-biased expression of the sialyltransferase *St3gal1* is associated
303 with differences in Muc19 glycosylation. Left: Boxplot showing transcript abundance of *St3gal1* (TPM)
304 in male and female mice. One female outlier (TPM = 443) is shown as a triangle and was kept for
305 statistical calculations. Right: SDS-PAGE and lectin blot analysis of Muc19 in pooled male (M) and
306 female (F) saliva samples. PNA detects unsialylated core O-glycans, whereas MAL-II detects terminal
307 α2,3-linked sialic acids. Differences in lectin binding and electrophoretic mobility indicate higher
308 sialylation of female Muc19, consistent with increased *St3gal1* expression.

309 ***Klk* gene family expansion underlies sex-biased expression in mice**

310 The profound sexual dimorphism in the mouse submandibular gland is not
311 recapitulated in humans (**Figure 2b**). This observation motivates a closer examination
312 of the evolutionary processes shaping sex-biased gene expression in the mouse
313 submandibular gland. In particular, we asked whether this extensive sex bias might
314 reflect coordinated regulation of clusters of genes within specific genomic regions or a
315 broad, genome-wide effect.

316 To address this, we first searched for physically clustered genes that may be co-
317 regulated to express in a sex-specific manner. Specifically, we performed a genomic
318 clustering analysis of sex-biased DEGs in the mouse submandibular gland using 100-
319 kb windows (see **Methods**; **Table S9**). We found that the Klk gene family forms the
320 largest cluster of sex-biased genes in the genome and is a significant outlier relative
321 to other clusters ($p < 0.001$; **Figure 3a**). Fourteen members of the kallikrein (Klk) family
322 (*Klk15, Klk1b8, Klk1b1, Klk1b9, Klk1b11, Klk1b26, Klk1b27, Klk1b21, Klk1b22,*
323 *Klk1b16, Klk1b24, Klk1b3, Klk1b4, and Klk1b5*) exhibit strong male-biased expression,
324 whereas one member (*Klk1*) shows female-biased expression. Altogether, *Klk* genes
325 make up ~16.4% of total gene expression in male submandibular glands, compared
326 with only 2.5% in female ones (**Figure 3b**).

327 We then asked whether the clustering of sex-biased genes can also be observed at
328 the proteomic level. Of the Klk genes showing sex-biased expression, 11 out of 15
329 also display concordant sex-biased differences in protein abundance in saliva. Using
330 genomic coordinates of genes encoding differentially abundant proteins, we observed
331 that the *Klk* genes exhibiting sex-biased protein abundance are clustered on
332 chromosome 7 (**Figure 3a**), consistent with the transcriptomic clustering pattern.
333 Together, these results indicate that the kallikrein locus represents a genome-wide

334 outlier cluster of coordinated, sex-biased regulation at both the transcriptomic and
335 proteomic levels.

336 Previous studies have already shown that *Klk* genes exhibit male-biased expression
337 in the mouse submandibular gland (Stopka et al. 2016; Mukaibo et al. 2019) and in
338 saliva (Penschow et al. 1991; Karn and Laukaitis 2011; Blanchard et al. 2015;
339 Moustardas et al. 2021). Single-cell transcriptomic analyses further indicate that *Klk*
340 genes are specifically expressed in granular convoluted tubule (GCT) and acinar cells
341 (Horeth et al. 2021), which are approximately 5- and 1.5-fold more abundant in males,
342 respectively (Pardini and Taga 1996). This raises the question of whether differences
343 in cell abundance between male and female submandibular glands alone can explain
344 the observed sex-biased expression. If expression per cell were equivalent between
345 sexes, differences in cell-type abundance could, in principle, account for higher *Klk*
346 expression in males. However, when we quantified overall expression levels, we found
347 that male-female differences exceeded by at least 5.5-fold for all differentially
348 expressed *Klk*s, indicating that the observed expression differences cannot be
349 explained by cell-type proportions alone (**Figure 3c**).

350 One remarkable exception is that the observed male-biased pattern for *Klk1b* genes
351 is not observed in *Klk1*. In fact, we found that *Klk1* is expressed significantly higher in
352 female submandibular glands despite the fact that *Klk1b* genes were shown to be
353 duplicates of *Klk1* (Lawrence et al. 2010; Pavlopoulou et al. 2010), and that *Klk1* is
354 also known to be expressed in acinar and GCT cells which are much more abundant
355 in males. To verify this transcriptional level observation at the protein and cellular
356 levels, we performed immunohistochemical staining of both male and female
357 submandibular gland sections for the female-biased *Klk1* gene, one of the few *Klk*
358 family members for which a robust antibody is available (**Figure 3d)**. Consistent with
359 the gene expression pattern, protein abundance was significantly higher in female
360 granular convoluted tubules (**Figure S6**), resulting in greater overall *Klk1* expression
361 in females despite the higher abundance of these cells in males. This observation
362 suggests evolution of complex regulatory mechanisms in the Klk locus, distinct for *Klk1*
363 and its paralogs.

364 Together, these results indicate that differences in cellular composition alone are
365 insufficient to explain sex-biased *Klk* expression and suggest additional genetic or
366 epigenetic regulation that drive this dimorphism.

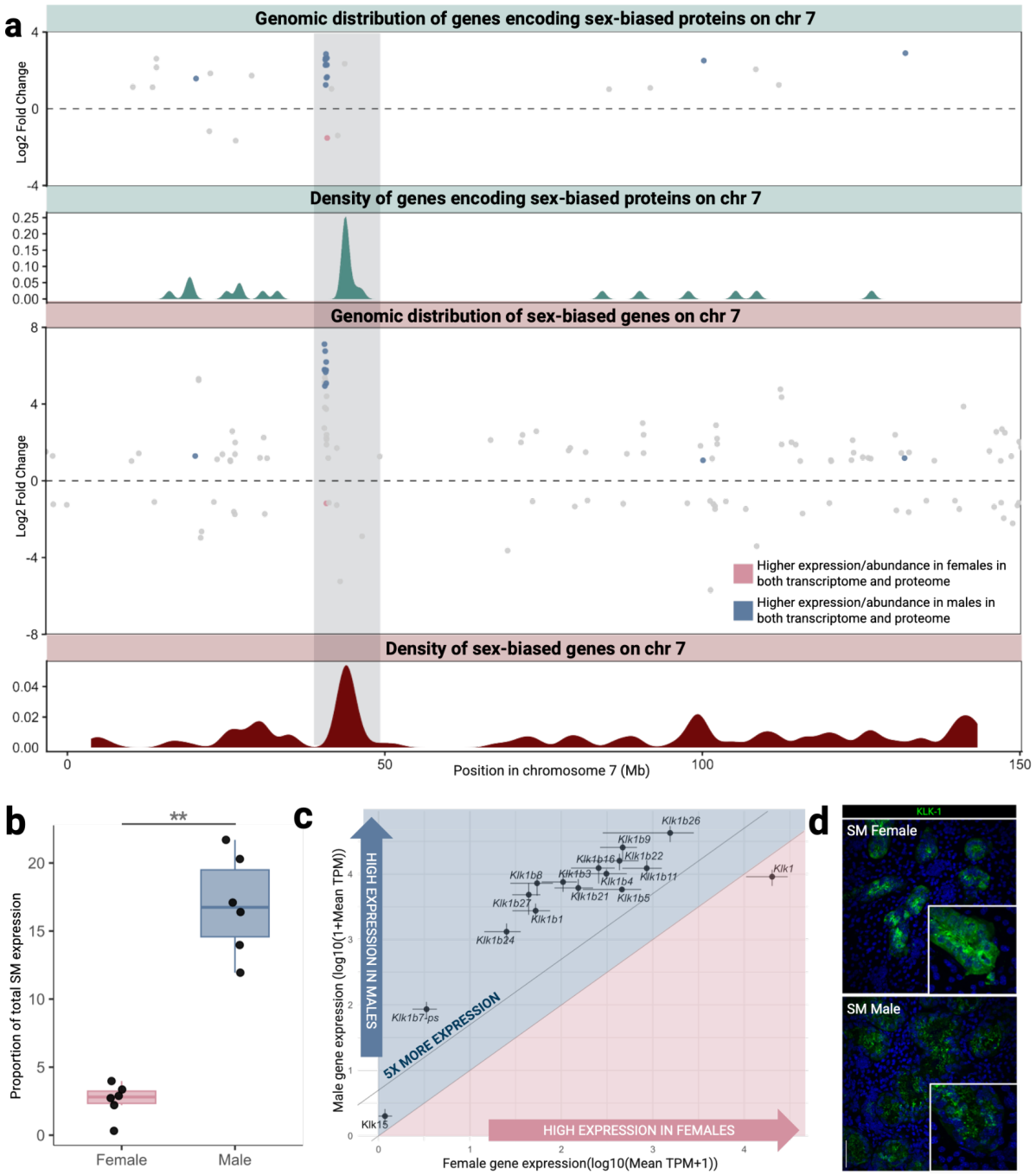

Figure 3. The *Klk* gene family forms the largest cluster of sex-biased transcripts and proteins in mouse submandibular glands, independent of differences in cell abundance. **(a)** Genomic distribution of genes encoding for sex-biased proteins (top, green) and sex-biased transcripts (bottom, magenta) along mouse chromosome 7 in the submandibular gland. Points represent differentially abundant proteins or differentially expressed genes, plotted by genomic position and log2 fold difference between sexes. The shaded region highlights the *Klk* locus. Density tracks below each scatter plot show kernel-smoothed spatial density of significant genes or proteins along the chromosome 7. Blue and pink points indicate genes whose sex bias at the transcript level is also supported by proteomic differences in males and females, respectively. **(b)** Relative expression of *Klk* genes in male and female submandibular glands. Boxes show distribution across biological replicates; points represent individual samples. ** indicates $p < 0.001$. **(c)** Comparison of female versus male *Klk* gene expression levels. All *Klk1b* paralogs show >5-fold higher expression in males, indicating that sex-biased expression is not explained by differences

in glandular cell composition. **(d)** Immunofluorescence staining of submandibular glands from female (top) and male (bottom). Klk1 protein is shown in green; nuclei are counterstained with DAPI (blue), illustrating stronger Klk1 expression in females.

Sex-specific *Klk* expression is shaped by hormonal regulation, and lineage-specific 3D genomic architecture

To investigate the genetic basis of male-biased expression at the *Klk* locus, we first generated a gene tree incorporating gene family expansion and contraction estimates inferred by CAFE 5 and obtained from Ensembl (Mendes et al. 2021). We found that placental mammals, including other rodents, harbor up to 15 *Klk* genes, whereas mice and rats harbor approximately 17-25, indicating a burst of duplications in the murid lineage (**Figure S7**). Specifically, there are 13 additional tandemly duplicated copies in mice compared to the human genome (**Figure 4a**), all duplicated from *Klk1* (Lawrence et al. 2010; Pavlopoulou et al. 2010) and with male-biased expression.

Next, we asked whether the large cluster of differentially expressed genes containing *Klk* genes in the submandibular gland share common regulatory sequences that could underlie their coordinated male-specific expression. To address this, we performed a *de novo* motif enrichment analysis on this cluster, identifying overrepresented sequence motifs located near genes (see **Methods**). We found an enrichment ($p = 1 \times 10^{-42}$) for a conserved CTGATCCTGTTC motif located within 30 bp of the transcription start site in all sex-biased *Klk* genes except *Klk15* (**Figure 4a**; see Methods; **Tables S10** and **S11**). Previous work showed that this motif is a binding site for Ghrl2, a developmental transcription factor whose activity is sex-biased and correlated with testosterone levels (Paltoglou et al. 2017; Reese et al. 2019). Consistent with this, kallikreins are well documented as androgen-regulated genes in the mouse submandibular gland (Van Leeuwen et al. 1987; Kurabuchi et al. 2001; Lawrence et al. 2010). In contrast, the motif is absent from *Klk* genes that do not show sex-biased expression. Moreover, none of the human *KLK* genes present this motif at the orthologous locus (**Table S11**). Given that the *Klk1b* paralogs arose through lineage-specific duplications of an ancestral *Klk1* gene (Lawrence et al. 2010; Pavlopoulou et al. 2010), our results support an evolutionary scenario in which the CTGATCCTGTTC motif emerged in the ancestral murine *Klk1* gene before subsequent gene expansion. Subsequent duplications may have increased gene expression through this regulatory mechanism, thereby amplifying male-biased expression in the submandibular gland.

Previous work in humans identified two distinct topologically associating domains (TADs) that structure the regulatory landscape of the *KLK* gene family, both demarcated by CTCF binding sites (Khoury et al. 2020). We therefore asked whether the lineage-specific duplications in mice resulted in the formation of a new or expanded the existing topologically associating domain (TAD). To address this, we analyzed publicly available Hi-C, ATAC-seq, and ChIP-seq datasets, including histone modification profiles, from the mouse B-cell lymphoma cell line (see **Methods**).

421 Consistent with the human architecture, we identified two TADs at the mouse *Klk* locus
422 of which the second TAD is markedly expanded, encompassing 15 genes, including
423 all those that show sex-biased expression (**Figure 4b**). Chromatin accessibility
424 (ATAC-seq), histone modification (H3K27ac), and transcriptional anchor protein
425 binding (Ctcf, Rad21) profiles delineate the borders of these TADs.

426 Although the TAD boundaries in mice largely mirror those observed in humans and
427 encompass all sex-biased *Klk* genes, it remains unclear how *Klk1* retains both the
428 regulatory motif and its position within the same TAD. To explore whether sex-specific
429 chromatin organization might contribute to this regulatory difference, we compared
430 TAD boundaries between male and female cells. Using publicly available Hi-C data
431 from male Sertoli cells and female granulosa cells, we reconstructed TAD structures
432 and compared their boundaries (**Figure 4c**). This analysis revealed a slight boundary
433 shift that coincides with the region showing male-specific expression, suggesting that
434 *Klk1* in females may be regulated within a different TAD context. These findings
435 indicate that regulatory control at this locus is more complex than previously
436 appreciated, and that the data presented here are insufficient to fully explain the
437 expression pattern of *Klk1*.

438 Together, these results suggest that lineage-specific *Klk* gene duplications have
439 reshaped the ancestral TAD structure shared with humans (**Figure 4d**), expanding it
440 in mice in a sex-specific manner.

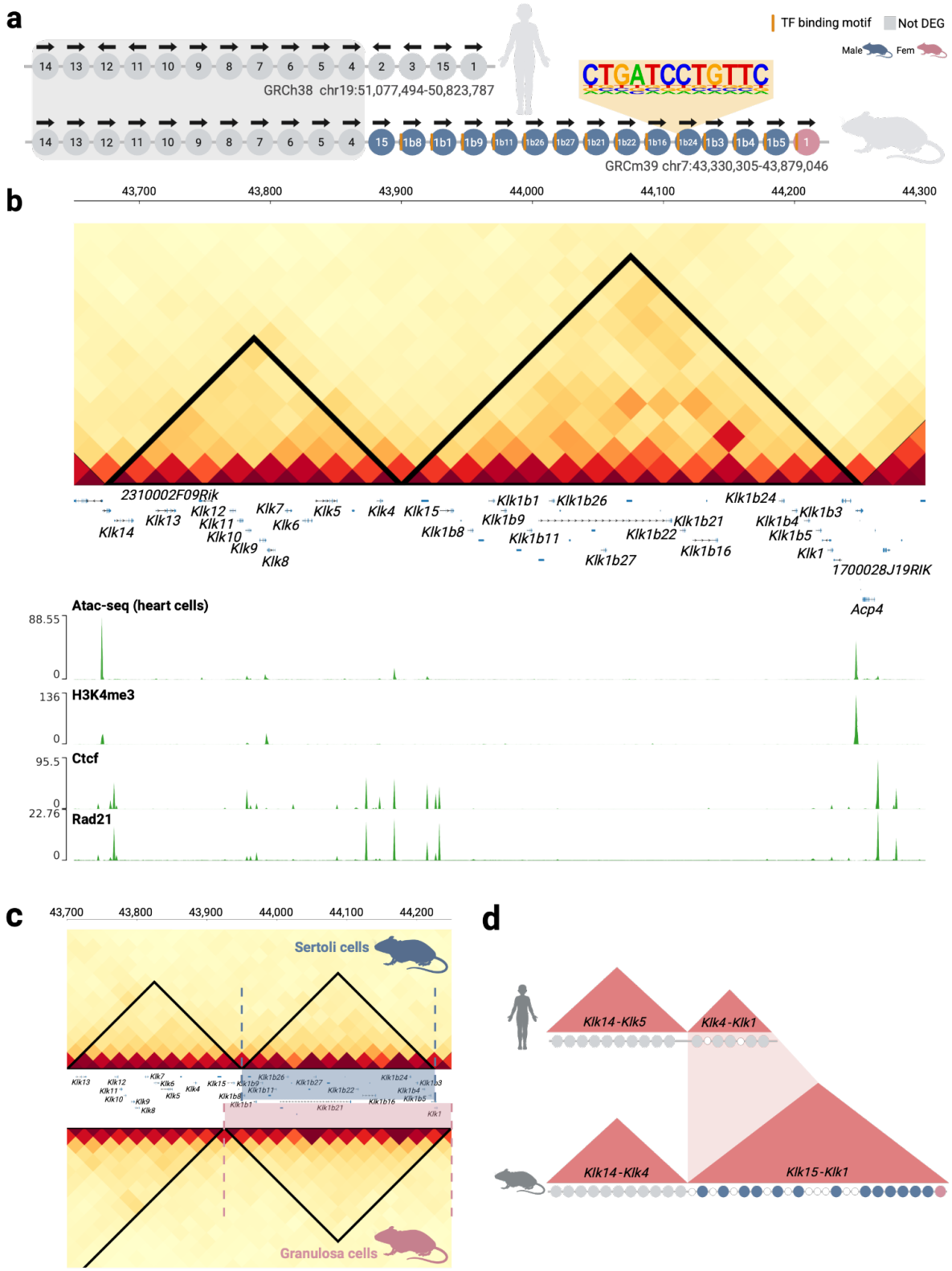

Figure 4. Expansion of the mouse *Klk* locus is associated with sex-biased expression and reorganization of local 3D genome structure. **(a)** Schematic comparison of the *KLK/Klk* locus organization in humans (top) and mice (bottom). Circles represent individual *KLK/Klk* genes ordered by genomic position and transcriptional orientation. Genes showing female- or male-biased expression are indicated in pink and blue, respectively. Orange bars denote putative transcription factor binding motifs shared across the expanded mouse cluster. The mouse lineage shows extensive tandem duplication

of *Klk* genes compared with the more compact human locus. **(b)** Hi-C contact map of the mouse *Klk* region generated from B-cell lymphoma cell line, revealing two adjacent topologically associating domains (TADs). The left TAD predominantly contains non–sex-biased genes, whereas the right TAD encompasses the expanded cluster of sex-biased *Klk* paralogs. Gene annotations are shown below the contact map together with chromatin accessibility and regulatory marks (ATAC-seq, H3K4me3, CTCF, and RAD21), highlighting active promoters and architectural features associated with TAD boundaries. **(c)** Hi-C contact maps of the mouse *Klk* locus reconstructed from male Sertoli cells (top) and female granulosa cells (bottom). Dashed vertical lines indicate shifts in TAD boundaries between the two cell types. Genes located within these alternative domain configurations are highlighted, suggesting that differences in higher-order chromatin organization may influence sex-specific regulation of subsets of *Klk* paralogs. **(d)** Conceptual model illustrating how tandem gene duplication at the *Klk* locus may have been accompanied by expansion and restructuring of the surrounding TAD landscape.

DISCUSSION AND CONCLUSION

In this study, we investigated the evolutionary forces shaping the composition of the salivary gland secretome in mice. We show that most highly expressed genes encoding secreted saliva proteins in mice are lineage-specific and lack direct human orthologs. We further identify the submandibular gland as the focal point of sex-specific expression. The kallikrein (*Klk*) gene family in mice accounts for ~16.4% of male-biased expression, and we discuss how lineage-specific gene family expansion, emergence of novel regulatory motifs, and topological clustering contribute to the pronounced sex differences in expression. Together, the findings illustrate how gene duplication and regulatory rewiring drive sex-biased gene expression and contribute to the rapid evolution of the mammalian salivary gland secretome in different lineages.

When we examined sex differences in mouse saliva, we found that the submandibular gland is the main contributor to the protein differences observed between males and females, consistent with previous results (Treister et al. 2005; Mukaibo et al. 2019). This male bias in submandibular expression is particularly notable when compared with the liver, the most studied tissue for sex differences in mice (Nikkanen et al. 2022), where the degree of dimorphism is lower. We further found that this bias can be traced to the lineage-specific gene duplication clusters identified in our analysis, with the kallikrein (*Klk*) cluster standing out as a single gene family accounting for ~16.4% of male-biased expression in the submandibular gland. The sex-biased expression of kallikrein genes is well documented in the literature, aligning with the patterns we observe here (Karn and Laukaitis 2011; Stopka et al. 2016; Moustardas et al. 2021). Specifically, our results support a model in which the expression of a hypothetical ancestral *Klk* gene in the murine lineage was subsequently amplified through a burst of tandem duplications within the same topologically associating domain (TAD), a process that could have both accompanied and facilitated developmental shifts in the male submandibular gland. This mechanistic framework suggests a broader evolutionary trend: rapid regulatory change across multiple genes can emerge through *de novo* motif evolution at a single locus, followed by its propagation via gene duplication within the same 3D genomic domain (**Figure 5**). A similar trend for motif evolution has recently been observed at the primate amylase locus (Karageorgiou et al. 2025)

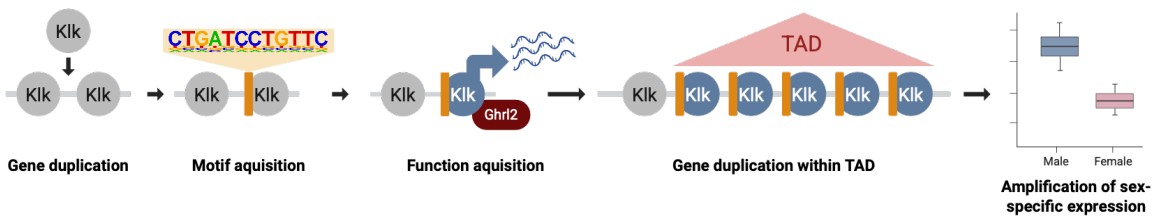

Figure 5. Proposed model for the evolutionary trajectory of mouse kallikrein (*Klk*) genes in salivary glands. Gene duplication created multiple *Klk* paralogs, which subsequently acquired new regulatory motifs. These regulatory changes enabled functional diversification, including interaction with Ghrl2.

Further duplications occurred within a conserved topologically associating domain (TAD), promoting coordinated regulation of the expanded gene set. This combination of gene family expansion and regulatory rewiring led to increased male-biased expression in the submandibular gland.

The pronounced male-biased expression of kallikreins in saliva raises questions about their phenotypic impact in male mice. A prevailing hypothesis proposes that the presence of sex-specific proteins in saliva in male mice is related to improved healing through wound licking behavior after male-male aggression (Hutson et al. 1979; Stopka et al. 2016). While this hypothesis may be valid for the higher expression of wound healing related saliva proteins in male mice, such as epidermal growth factor and fibroblast growth factor (Hirata and Orth 1979; Brown et al. 2020), it may not extend to the kallikrein gene family. An alternative hypothesis has been posited by Karn and Laukaitis where sex-biased expression of proteins in rodent saliva is explained by either signaling male dominance among rival males or mate identification through pheromone-like mechanisms (Karn and Laukaitis 2011). Our results cannot confirm or falsify either of these hypotheses, but they may provide a mechanistic base for formulating testable hypotheses approaching the questions from a molecular genetic level.

While we propose a hypothesis of co-regulation among male-biased *Klk* genes, our results do not fully explain the expression pattern of the only female-biased *Klk* gene (*Klk1*). Previous studies suggest that additional regulatory mechanisms contribute to the regulation of *Klk* transcripts, including microRNA-mediated silencing (Kurihara et al. 2011), highlighting the regulatory complexity of this locus. Moreover, *Klk1* has been reported to be regulated by estrogen in mice through mechanisms distinct from those described for other *Klk* family members (Iwasaki et al. 2023). Future studies integrating genomic, regulatory, and three-dimensional genome architecture data from salivary glands of the same individuals ideally using a single-cell resolution will be necessary to validate our proposed model of co-evolution among regulatory motifs, chromatin organization, gene duplications, and cell-type composition.

The biological functions and physiological relevance of the extensive species- and sex-specific differences in salivary gene and protein expression remain largely unknown. While addressing these functions was beyond the scope of the present study, our results provide a comprehensive catalogue of candidate genes and proteins that can now be used to test hypotheses about the roles of salivary gland-derived proteins in physiology, immunity, and behavior. One novel observation is the sex-specific glycosylation of the Muc19 protein in mouse saliva, suggesting that differences in glycosylation machinery, particularly the differential expression of sialyltransferases between males and females, contribute to sex-specific modification of secreted mucins. This stands in contrast to humans, where MUC19 expression is low in salivary glands and no clear sexual dimorphism has been reported. Notably, genetic variation in human *MUC19* has been linked to complex introgression patterns in hominins and signatures of local selection in Indigenous American populations (Villanea et al. 2025), marking this gene as a hotspot of adaptive evolution. Together,

our observations provide a foundation for future work aimed at dissecting the regulatory basis of species-specific expression, the molecular mechanisms underlying sex-specific glycosylation, and the evolutionary processes shaping Muc19 variation across mammals.

Of importance for biomedical research, our findings raise more fundamental questions about the suitability of the mouse as a standard model for studying human salivary gland biology, a concern that has also been noted previously (Maruyama et al. 2019). Focusing on salivary glands and their secreted products, we show that mice and humans differ fundamentally in both the expression patterns and composition of salivary proteins, including their post-translational modifications. With respect to disease relevance in humans, we found that several highly expressed Sjögren's-related genes (Horeth et al. 2021), including *Dcpp1*, *Scgb2b26*, and *Scgb2b27*, are mouse-specific and lack human counterparts, thus highlighting the importance of validating translational studies involving mice salivary glands and saliva in human models. In addition, our results underscore the importance of explicitly accounting for sex as a biological variable in studies of oral and salivary gland biology, particularly in mouse models.

Overall, this study, although limited to a comparison between two species, contributes to the emerging view that saliva exhibits considerable evolutionary plasticity, enabling accelerated evolution under selective pressures (Ruhl 2012; Pajic et al. 2019; Pajic et al. 2025) and offers novel additional mechanistic hypotheses to explain these trends. Furthermore, our results are consistent with the hypothesis of developmental systems drift (True and Haag 2001), whereby similar phenotypic traits across species evolve through distinct genetic or gene-regulatory mechanisms. Although saliva performs broadly conserved functions across mammals, primarily related to digestion, protection of oral tissues, microbial homeostasis, and pathogen defense, our results suggest that these functions may be achieved through different regulatory architectures across species. In this context, the murine *Klk* locus provides a useful model for investigating how gene duplication, regulatory motif evolution, and 3D genome organization can shape the evolution of sex-biased gene regulation and the emergence of novel sexually dimorphic traits. Expanding similar analyses across multiple mammalian species will be crucial for uncovering general principles governing the evolution of salivary gland gene regulation and function.

571 **METHODS**

572 **Sample collection for RNA-seq**

573 A total of 12 mice (3 months old) were used, consisting of three biological replicates
574 for each strain-sex combination: 3 female CD1, 3 male CD1, 3 female C57BL/6, and
575 3 male C57BL/6 (**Figure S8**). CD1 and C57BL/6 mice were selected as representative
576 outbred and inbred strains, respectively, as both are widely used in laboratory studies
577 of salivary glands (Nguyen et al. 2009; Gervais et al. 2015; Hsieh et al. 2017). Mouse
578 samples were provided by UCSF. Animals were euthanized following the approved
579 protocol AN201280-00I. From each mouse, we collected five glands: the three major
580 salivary glands (parotid, sublingual, and submandibular glands), as well as the
581 pancreas and liver, which were used as reference tissues.

582 **RNA-sequencing and analysis of mouse transcriptomes and differential**
583 **analysis**

584 Total RNA was extracted from each tissue, and RNA-seq libraries were prepared by
585 Azenta Life Sciences (GENEWIZ) using their Standard/Strand-Specific RNA-Seq
586 service with poly(A) selection for mRNA enrichment, followed by sequencing on an
587 Illumina® NovaSeq™ platform (2×150 bp). Raw reads were processed with Trim
588 Galore (Krueger et al. 2023), which uses cutadapt for adapter trimming and FastQC
589 for quality assessment, to remove low-quality bases and adapter sequences. Cleaned
590 reads were then aligned to the *Mus musculus* reference genome (GRCm39) using
591 HISAT2 (Kim et al. 2019), and gene-level quantification was performed with
592 featureCounts (Liao et al. 2014), which assigns aligned reads to annotated genes and
593 produces a matrix of raw read counts per gene.
594
595 Six parotid samples (two male and one female CD1, and three female C57BL/6)
596 exhibited high expression of genes associated with muscle function and contraction,
597 which is unusual in parotid gland transcriptomes. These samples were excluded from
598 downstream analyses, as their expression profiles suggested possible contamination
599 during tissue dissection. Next, we measured differential gene expression between
600 sexes for each tissue using DESeq2 (Love et al. 2014), applying a significance
601 threshold of $p_{adj} < 0.05$ and |log2 fold change| > 1. DESeq2 models raw read counts
602 for each gene and compares expression levels between sexes, estimating differential
603 expression while accounting for biological variability. Counts are internally normalized
604 by library size and sequencing depth to allow accurate comparisons across samples,
605 and significantly differentially expressed genes (DEGs) were identified for each tissue.
606 To assess sample quality and confirm proper clustering by biological variables, we
607 performed principal component analysis (PCA) across all glands and separately for
608 the three major salivary glands (**Figure S1**).

609 **Orthology and Secreted Gene Identification**

610 We first sought to determine for all genes in each tissue the human orthologs and
611 which were mouse specific. For this, we used an orthology reference table provided
612 by The Jackson Laboratory (JAX)
613 (https://www.informatics.jax.org/downloads/reports/HMD_HumanPhenotype.rpt),
614 which lists mouse-human orthologous genes. Using this table, we grouped genes
615 based on the number of orthologous relationships between species. Specifically,
616 genes were classified as one-to-one when a single mouse gene mapped to a single
617 human gene; one-to-many when a gene in one species mapped to multiple genes in
618 the other; and many-to-many when multiple genes in both species were linked through
619 orthology relationships. Genes without any annotated orthologs in the other species
620 were classified as mouse- or human-specific.

621 Next, we aimed to characterize the contribution of highly expressed genes to salivary
622 secretion by identifying secreted proteins and building an expression profile for each
623 gland. Transcript abundance was calculated in transcripts per million (TPM) using the
624 standard formula (Wagner et al. 2012). TPM values were computed for all genes in
625 the three major salivary glands (parotid, sublingual, and submandibular). All genes in
626 each gland were annotated as secreted or non-secreted based on the UniProt mouse
627 proteome (https://www.uniprot.org/proteomes/UP000000589) (UniProt Consortium
628 2025) and human from Human Protein Atlas (proteinatlas.org) (Uhlén et al. 2019). For
629 the mouse proteome, specific annotations for salivary proteins were not available;
630 therefore, we classified all genes as secreted or non-secreted based on general
631 proteome annotations from UniProt. Combined with the orthology annotations
632 described above, these data were used to generate Figure 1. We identified the 30
633 most highly expressed genes that code for secreted proteins in each salivary gland
634 and plotted their expression levels (in TPM) to visualize gland-specific expression
635 patterns using packcircle (https://mbedward.r-universe.dev/packcircles).

636 **Analysis of human transcriptomes**

637 For humans, we reanalyzed publicly available RNA-seq data from salivary glands
638 (parotid, submandibular, and sublingual) generated by Saitou et al. (2020). Raw
639 sequencing reads were processed as described above for mouse, using the human
640 reference genome (hg38). Expression levels were quantified as transcripts per million
641 (TPM), and mean TPM values were calculated for each gland. For liver and pancreas
642 tissues, transcript data in TPMs was downloaded from GTEx and filtered for the same
643 genes present in our dataset for salivary glands. TPMs were then recalculated for only
644 those genes, and mean TPMs were calculated using all individuals. For orthology
645 classification, please see the description of orthology classification above. Differential
646 expression analysis was performed using DESeq2 (Love et al. 2014) using raw gene
647 counts only for submandibular gland samples.

648 **Correlation analysis of human-mouse transcriptomes**

649 To assess the similarity of gene expression patterns between mouse and human
650 salivary glands, we performed cross-species correlations of gene expression (mean
651 TPM values) for each gland, focusing on one-to-one orthologs for genes with TPM>2
652 in each tissue. Because genes encoding secreted proteins and non-secreted proteins
653 are subject to distinct evolutionary pressures, and secreted proteins are often
654 expected to evolve and diverge more rapidly (Winter et al. 2004; Feyertag et al. 2016),
655 we focused our analysis on genes encoding secreted proteins, which constitute the
656 primary functional components of saliva. For the mouse, we used the annotated
657 transcriptomes generated in this study for all glands. For the human data, we used the
658 transcriptomic profiles from Saitou et al. (2020) for the major salivary glands (parotid,
659 submandibular, and sublingual) and from GTEx for the pancreas and liver. Spearman
660 correlation coefficients (ρ) were calculated for each comparison. Finally, we note that
661 these analyses involve data from different cohorts and species and are therefore
662 subject to potential batch effects; as such, these results should be interpreted with
663 caution. Because of these technical limitations, we avoided direct comparisons of
664 human and mouse expression data as a primary analytical approach. Nevertheless,
665 we identified several one-to-one orthologs that exhibit dozens- to hundreds-fold
666 differences in expression between human and mouse salivary glands (**Table S3 and**
667 **S4)**, differences that are unlikely to be explained solely by batch effects and therefore
668 warrant further investigation.

669 **Mouse saliva collection**

670 Saliva samples were collected from 4-month-old male (n=5) and female (n=5)
671 C57BL/6 mice. Mice were anesthetized with isoflurane and administered the
672 sialogogue pilocarpine dissolved in sterile saline via intraperitoneal (i.p.) injection.
673 Pilocarpine was delivered at a final dose of 4.53mg/kg (concentration 0.68mg/mL;
674 injection volume 200uL per 30g body weight). Saliva collection commenced 3 minutes
675 post-injection and continued for 10 minutes using a pipette placed in the oral cavity.
676 Samples were immediately snap-frozen and stored at -80 degrees Celsius for
677 subsequent protein analysis.

678 **Human saliva collection**

679 Saliva samples were collected from humans (male and female, n=5 each) using the
680 passive drooling technique as previously described (Thamadilok et al. 2020).

681 **SDS-PAGE**

682 The 8-16% gradient pre-cast Novex Tris-glycine gel (LifeTechnologies; catalog no.
683 XP08162BOX) was used to study the protein profiling of saliva samples based on the
684 molecular size of the protein. Meanwhile, the samples were prepared for gel
685 electrophoresis by mixing with a 2X sample buffer followed by boiling at 100^0C for 10
686 min. The samples (~15 µg of protein/well) were loaded into the wells, and the gel was

run on an electrophoresis system (XCell SureLock Mini-Cell , Invitrogen) at 170 V for approximately 40 minutes. Subsequently, the gel was prepared for gel staining.

Coomassie blue-PAS (Periodic acid Schiff) Staining

Protein bands were first stained using the Coomassie blue staining procedure. In brief, the procedure included three steps: a) fixation using fixing solution for 15 min twice, b) staining in Coomassie stain blue (R-250) solution for 1 hr., and c) destaining for 30 min thrice. After the destaining step, the gel was stained subsequently using the PAS staining procedure as described (Thamadilok et al. 2020). Stained gels were digitized using the GE Healthcare ImageScanner III.

Protein identification from excised gel bands by LC-MS/MS

Protein bands of interest were excised from Coomassie blue-stained SDS-PAGE gels corresponding to the ~150 kDa (male) and ~130 kDa (female) regions using a clean surgical blade and processed for protein identification by mass spectrometry. Gel slices were subjected to in-gel digestion with trypsin following standard protocols. Peptides were extracted and analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). The resulting spectra were processed using Spectronaut (Biognosys AG, version 20.0) and searched against the UniProt *Mus musculus* protein database. Protein identifications were filtered using a false discovery rate (FDR) threshold of 1% at both the peptide and protein levels.

Lectin Blotting

Proteins were separated by SDS-PAGE under reducing conditions using 8-16% gradient gels and subsequently transferred to nitrocellulose membranes. For lectin detection, membranes were blocked in Tris-buffered saline containing 0.1% Tween-20 (TBST) supplemented with 2% Polyvinyl alcohol (PVA) for 1 h at room temperature. Membranes were then incubated with biotinylated lectins diluted in blocking buffer. Peanut agglutinin (PNA) was used to detect exposed Galβ1–3GalNAc (T-antigen) O-glycan structures, while Maackia amurensis lectin II (MAL-II) was used to detect α2,3-linked sialic acid residues. Following lectin incubation, membranes were washed extensively with TBST and incubated with Alexa fluor-488-conjugated neutravidin. Lectin binding was detected by enhanced chemiluminescence and imaged with a digital imaging system.

Protein digestion

Saliva specimens were processed following the E3 protocol described recently (Martin et al. 2024). In brief, around 20 µL of crude saliva was first mixed with 100 µL SDS buffer that contained 5% (w/v) sodium dodecyl sulfate (SDS), 100 mM Tris-HCl solution (pH 8.0) and final concentration of 10 mM tris(2-carboxyethyl) phosphine (TCEP) and 40 mM 2-chloroacetamide (CAA), and then boiled at 95 ºC for 10 min. Next, the lysate was mixed with 4x volume of 80% acetonitrile (CAN) aqueous solution

727 and transferred to E3filters (CDS Analytical, Oxford, PA). Following a centrifugation
728 step, the E3filters were washed with 200 µL of 80% ACN aqueous solution three times.
729 For protein digestion, 1 mg of trypsin (Promega, WI) in a 200 µL of 50 mM
730 triethylammonium bicarbonate (TEAB) buffer was added to E3filters and incubated at
731 37 ºC for 16-18 h with gentle shaking (300 rpm). To elute peptides, the E3filters were
732 first centrifuged at 3,000 rpm and then flushed with two elution buffers, 200 ml of 0.2%
733 formic acid in water, and 200 ml of 0.2% formic acid in 50% ACN, respectively. The
734 eluents were dried in SpeedVac, and then subjected to desalting cleanup using C18
735 based StageTips (CDS Analytical, Oxford, PA). Dried peptide samples were stored
736 under -80 ºC until further analysis.

737 **LC-MS/MS analysis**

738 The LC-MS/MS analysis was performed using an UltiMate 3000 RSLCnano system
739 coupled with an Orbitrap Eclipse mass spectrometer and a FAIMS Pro Interface
740 (Thermo Scientific). The dried peptides were resuspended into 13 µl of LC buffer A
741 (0.1% formic acid in water) and 10 µl of it was then loaded onto a trap column
742 (PepMap100 C18, 300 µm × 2 mm, 5 µm; Thermo Scientific) followed by separation
743 on an analytical column (PepMap100 C18, 50 cm × 75 µm i.d., 3 µm; Thermo
744 Scientific). A linear LC gradient flowing at 250 nL/min was applied from 1% to 25%
745 mobile phase B (0.1% formic acid in ACN) over 125 min, followed by an increase to
746 32% mobile phase B over 10 min. The column was washed with 95% mobile phase B
747 for 5 min, followed by equilibration with mobile phase A for 15 min. For the ion source
748 settings, the spray voltage was set to 1.9 kV, funnel RF level at 50%, and heated
749 capillary temperature at 275°C. The MS data were acquired in Orbitrap at 120K
750 resolution, followed by MS/MS acquisition in data-independent mode following a
751 protocol described previously (Elsayyid et al. 2025). The MS scan range (m/z) was set
752 to 380-985, maximum injection time was 246 ms, and normalized AGC target was
753 100%. For MS/MS analysis, the spectra were acquired with Orbitrap at resolution of
754 30K. The isolation mode was Quadrupole, the isolation window was 8 m/z, and the
755 window overlap was 1 m/z. The collision energy was 30%, AGC Target was 400K, and
756 normalized AGC target was 800%. For FAIMS analysis, a 3-CV experiment
757 (−40|−55|−75) was applied.

758 **Proteome quantitation and data analysis**

759 The MS data were processed using Spectronaut software (version 20.0), a library-free
760 directDIA+ workflow, and the UniProt protein databases (17,141 sequences; reviewed
761 only). Briefly, the settings for Pulsar and library generation include: Trypsin/P as
762 specific enzyme; peptide length from 7 to 52 amino acids; allowing 2 missed
763 cleavages; toggle N-terminal M turned on; Carbamidomethyl on C as fixed
764 modification; Oxidation on M and Acetyl at protein N-terminus as variable
765 modifications; FDRs at PSM, peptide and protein level all set to 0.01; Quantity MS
766 level set to MS2, and cross-run normalization turned on.

767 Differential protein abundance between male and female samples was assessed
768 using the built-in statistical framework in Spectronaut (version 20.0). The protein LFQ
769 method was set to Automatic (MaxLFQ); Quantity MS level was set to MS2, and the
770 Quantity type was Area. The Candidates output from Spectronaut was used for
771 quantitation analysis. Proteins were considered significantly differentially abundant
772 between sexes if they met the thresholds of |fold change| > 1.5 and q-value < 0.05.

773 **Transcriptome-Proteome Integration and Correlation Analysis**

774 Proteomic data were filtered to retain only proteins with a single corresponding gene
775 annotation to avoid ambiguity in gene-protein assignments. For each protein, mean
776 intensity-based absolute quantification (iBAQ) values were calculated separately for
777 males, females, and across all samples. For visualization and correlation analyses,
778 expression values were transformed as $\log_{10}(1 + \text{mean iBAQ})$. To enable
779 transcriptome-proteome comparisons, we first filtered the transcriptomic dataset to
780 genes annotated as secreted proteins (see description above). Proteins whose
781 corresponding genes were present in this filtered transcriptomic dataset were then
782 retained, resulting in 467 matched gene-protein pairs. Spearman correlation analyses
783 were then performed using these matched protein-transcript pairs. It is important to
784 note that each salivary gland contributes only partially to the overall proteome
785 composition of mixed whole saliva, the oral fluid analyzed in this study. Therefore,
786 strong correlations between glandular transcriptomes and the whole-saliva proteome
787 are not necessarily expected, given the distinct spectrum of proteins secreted by each
788 gland.

789 Furthermore, several highly expressed salivary gland transcripts were not detected by
790 mass spectrometry, consistent with known technical limitations of proteomic assays
791 for certain salivary proteins (Malaker 2024). Conversely, some proteins detected in
792 saliva at high abundance, such as serum albumin, are not synthesized by salivary
793 glands but instead enter whole saliva through plasma exudation in the oral cavity, as
794 reported previously (Oppenheim and Hay 1972; Ruhl 2012).

795 **Immunohistochemistry analysis**

796 Submandibular glands were manually dissected and immediately fixed in 4%
797 paraformaldehyde (PFA) at 4°C overnight. Following fixation, tissues were
798 cryoprotected using a graded sucrose approach (10% sucrose w/v followed by 25%
799 sucrose w/v in PBS). Tissues were then embedded in optimal cutting temperature
800 (OCT) compound (Fisher Scientific, Cat # 4585) and cryosectioned at a thickness of
801 16µM. Sections were permeabilized with 0.5% Triton X-100 in PBS for 20 minutes at
802 room temperature. Following permeabilization, sections were blocked for 1 hour at
803 room temperature in a buffer containing 10% donkey serum (Jackson
804 ImmunoResearch, Cat# 017-000-121) and 5% BSA (Fisher BioReagents, Cat#
805 BP9700100) in PBS.

806 Sections were incubated with primary antibodies, Kallikrein 1 polyclonal antibody
807 (1:300; Invitrogen, Cat# PA5-115481) and Nerve Growth Factor (NGF) polyclonal
808 antibody (1:300; Sigma-Aldrich, Cat# N6655), diluted in antibody diluent overnight at
809 4°C. Following washing with 0.05% Tween-20 in PBS, antibodies were detected using
810 Cy2-conjugated secondary Fab fragment antibodies (1:300; The Jackson Laboratory)
811 incubated for 2 hours at room temperature. Nuclei were counterstained with Hoechst
812 33342 (1:1000; Anaspec Inc. Cat# AS-83218). Slides were mounted using
813 Fluoromount-G (SouthernBiotech, Cat# 01100-01) and imaged using a Zeiss LSM 900
814 confocal microscope with a 40 X oil immersion objective.

815 Quantification of KLK-1 and NGF expression was performed using Fiji software.
816 Regions of interest (ROIs) were manually delineated around salivary gland ductal
817 structures identified by distinct morphology. The Mean Fluorescence Intensity (MFI)
818 was calculated for each ductal ROI. Results are reported as MFI (a.u.).

819 **Clustering of DEGs, and motif enrichment analysis**

820 We hypothesized that co-regulated gene clusters contribute to the differences in gene
821 expression observed between sexes. To test this, we focused on the submandibular
822 gland for sex differences, as this tissue showed the highest numbers of DEGs. Using
823 bedtools, we grouped DEGs into genomic clusters defined by a 100 kb window, such
824 as the following example:

```
825 bedtools cluster -i SM_sex_DEGs.sorted.bed -d 100000 >
826 SM_sex_DEGs.clustered_100kb.bed
```

827 To evaluate whether DEGs were more clustered than expected by chance, we
828 performed a permutation analysis. Specifically, we randomly sampled 10,000 subsets
829 of 1615 genes (matching the observed DEG count) from the total set of expressed
830 genes and applied the same clustering procedure to each subset. From these
831 permutations, we built an expected distribution of the number of genes per cluster size
832 and compared it to the observed data (**Table S9**).

833 Next, we focused on the kallikrein (Klk) gene cluster in the mouse submandibular
834 gland. Promoter regions for genes within the cluster were extracted and analyzed for
835 motif enrichment using HOMER (Hypergeometric Optimization of Motif EnRichment)
836 (Heinz et al. 2010). HOMER performs *de novo* motif discovery by identifying sequence
837 motifs that are significantly enriched in a target set of sequences relative to a
838 background set using a hypergeometric framework, and compares discovered motifs
839 to known transcription factor binding motifs in curated databases (including HOMER
840 and JASPAR libraries). For each gene in the cluster, sequences spanning −400 bp
841 upstream to +100 bp downstream of the transcription start site (TSS) were analyzed.
842 De novo motif discovery was performed using motif lengths of 8, 10, and 12 bp (**Table
843 S10**), according to the following code

```
findMotifs.pl "$gene_list_in cluster" mouse "out_folder/" \
        -start -400 -end 100 \
        -len 8,10,12 \
        -p 4
```

To further confirm the motifs identified in the enrichment analysis and evaluate whether they are present across other Klk genes in mouse and in their human orthologs, we searched for these motifs across all Klk gene sequences in both species. Gene sequences for all annotated Klk genes were downloaded from the GRCm39 mouse genome assembly and the corresponding human genome assembly in FASTA format. Extracted sequences included the full gene regions encompassing coding sequences as well as 5′ and 3′ untranslated regions (UTRs). The short motif sequence identified by HOMER (CTGATCCTGTTC) was queried against these gene sequences using BLASTN with the blastn-short task, which is optimized for short query sequences, according to the following example:

```
blastn -task blastn-short -query $motif.fa -subject $genes.fa -
outfmt 6 -out out.blastn.out
```

This approach allowed us to identify exact or near-exact matches of the motifs within Klk gene sequences (**Table S11**).

Evolution of the *Klk* genes, chromatin conformation and Topological Associated Domains Visualization

To investigate the evolution of the Kallikrein genes and gene expansions-contractions, we used CAFE 5 (Mendes et al. 2021). To investigate whether the sex-biased expression of Kallikrein (*Klk*) genes could be explained by differences in genome architecture and transcriptional regulation, we analyzed publicly available epigenomic datasets from the ENCODE project. Specifically, we visualized the Klk locus (chr7:43650000-44300000, mm10) using Hi-C contact maps at 25kb resolution generated from B-cell lymphoma cell line (https://www.encodeproject.org/files/ENCFF076GYR/) using HiCExplorer (Wolff et al. 2020). We further identified topologically associating domains (TADs) from the Hi-C data and overlaid them onto the figure. The code used for TAD calling and visualization is provided below.

```
hicFindTADs \
  --matrix ENCFF651SGY.mcool::/resolutions/25000 \
  --chromosomes chr7 \
  --minDepth 75000 \
  --maxDepth 250000 \
  --step 50000 \
  --correctForMultipleTesting fdr \
  --thresholdComparisons 0.05 \
  --outPrefix TADs_heart_chr7_25kb_Local
```

```
WIN2="chr7:43650000-44300000"

hicPlotTADs --tracks 25kbTADs_selec.ini --region $WIN2 \
  --outFileName 25kbTADs_selec.svg --dpi 300
```

To explore potential regulatory features underlying the observed expression patterns, we overlaid ATAC-seq data (https://www.encodeproject.org/experiments/ENCSR451NAE/) representing chromatin accessibility, chromatin immunoprecipitation sequencing (ChIP-seq) data for the architectural protein Ctcf (https://www.encodeproject.org/experiments/ENCSR000ERM/) and for histone modification H3K27me3 (https://www.encodeproject.org/experiments/ENCSR000CGK/), and for Rad21 (https://www.encodeproject.org/experiments/ENCSR000ERK/) from the same cell line.

To investigate sex-specific differences in chromatin organization at the kallikrein (Klk) locus, we analyzed publicly available Hi-C datasets generated from granulosa (female) and Sertoli (male) cells (Lindeman et al., 2021). These datasets were obtained from the 4DN Data Portal (https://data.4dnucleome.org/) under accession 4DNEXDK2R8M5 and 4DNEX99DGLEF. Multi-resolution cooler (.mcool) files were downloaded for each dataset (Sertoli: 4DNFIB7RFFBB; granulosa: 4DNFIZ59STGB). These cell types were selected because they represent sex-specific gonadal cell lineages, providing a model to examine potential sex-dependent chromatin architecture. Hi-C contact maps were extracted at 25 kb resolution, and topologically associating domains (TADs) were independently identified for each dataset using HiCExplorer as described above. The resulting TAD boundaries were then compared between male and female datasets to assess differences in higher-order genome organization at the Klk locus.

ACKNOWLEDGMENTS

We thank Alber Aqil, Charikleia Karagerogiou, Kendra Scherr, Vincent Lynch, and Petar Pajic for helpful discussions and insights. We thank Yanbao Yu for assistance with proteomic analyses and mass spectrometry. We thank Azenta Life Sciences for assistance with RNA-seq library preparation and sequencing. This work utilized computational resources provided by the Center for Computational Research at the University at Buffalo. This work was supported by the National Science Foundation (NSF) (2049947) (to O.G.) and the National Institute of General Medical Sciences (R35-GM156519) (to O.G.).

DATA AVAILABILITY

RNA-seq data generated in this study will be deposited in the NCBI Gene Expression Omnibus (GEO) prior to publication, and accession numbers will be provided upon acceptance. Human salivary gland RNA-seq data were obtained from Saitou et al. 2020 and are publicly available under the corresponding accession numbers. Proteomic data generated in this study will be deposited in an appropriate public repository (e.g., ProteomeXchange/PRIDE) prior to publication. All scripts used for data processing and analysis will be made publicly available upon acceptance.

REFERENCES

Bie B, Zhang Z, Cai Y-Q, Zhu W, Zhang Y, Dai J, Lowenstein CJ, Weinman EJ, Pan ZZ. 2010. Nerve Growth Factor-Regulated Emergence of Functional δ-Opioid Receptors. *J. Neurosci.* 30:5617–5628.

Blanchard AA, Ezzati P, Shamshurin D, Nistor AC, Leygue E, Wilkins JA, Myal Y. 2015. Towards further defining the proteome of mouse saliva. *Proteome Sci.* 13:10.

Brown CT, Nam K, Zhang Y, Qiu Y, Dean SM, Dos Santos HT, Lei P, Andreadis ST, Baker OJ. 2020. Sex-dependent Regeneration Patterns in Mouse Submandibular Glands. *J. Histochem. Cytochem.* 68:305–318.

Caramia F. 1966. Ultrastructure of mouse submaxillary gland. *J. Ultrastruct. Res.* 16:505–523.

Chai Y, Klauser DK, Denny PA, Denny PC. 1993. Proliferative and structural differences between male and female mouse submandibular glands. *Anat. Rec.* 235:303–311.

Chibly AM, Aure MH, Patel VN, Hoffman MP. 2022. Salivary gland function, development, and regeneration. *Physiol. Rev.* 102:1495–1552.

Dawes C, Pedersen AML, Villa A, Ekström J, Proctor GB, Vissink A, Aframian D, McGowan R, Aliko A, Narayana N, et al. 2015. The functions of human saliva: A review sponsored by the World Workshop on Oral Medicine VI. *Arch. Oral Biol.* 60:863–874.

De Sousa-Pereira P, Cova M, Abrantes J, Ferreira R, Trindade F, Barros A, Gomes P, Colaço B, Amado F, Esteves PJ, et al. 2015. Cross-species comparison of mammalian saliva using an LC–MALDI based proteomic approach. *PROTEOMICS* 15:1598–1607.

Elsayyid M, Tanis JE, Yu Y. 2025. Simple In-Cell Processing Enables Deep Proteome Analysis of Low-Input Caenorhabditis elegans. *Anal. Chem.* 97:9159–9167.

Enard D, Cai L, Gwennap C, Petrov DA. 2016. Viruses are a dominant driver of protein adaptation in mammals. *eLife* 5:e12469.

Feyertag F, Berninsone PM, Alvarez-Ponce D. 2016. Secreted proteins defy the expression level–evolutionary rate anticorrelation. *Mol. Biol. Evol.*:msw268.

Gahmberg CG, Andersson LC. 1982. Role of Sialic Acid in the Mobility of Membrane Proteins Containing O-Linked Oligosaccharides on Polyacrylamide Gel Electrophoresis in Sodium Dodecyl Sulfate. *Eur. J. Biochem.* 122:581–586.

Gao X, Oei MS, Ovitt CE, Sincan M, Melvin JE. 2018. Transcriptional profiling reveals gland-specific differential expression in the three major salivary glands of the adult mouse. *Physiol. Genomics* 50:263–271.

Gervais EM, Desantis KA, Pagendarm N, Nelson DA, Enger T, Skarstein K, Liaaen Jensen J, Larsen M. 2015. Changes in the Submandibular Salivary Gland Epithelial Cell Subpopulations During Progression of S jögren's Syndrome-Like Disease in the NOD / S hi L t J Mouse Model. *Anat. Rec.* 298:1622–1634.

Hatakeyama S, Sashima M, Suzuki A. 1987. A sexual dimorphism of mucous cells in the submandibular salivary gland of rat. *Arch. Oral Biol.* 32:689–693.

975 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh
976 H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription
977 Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell
978 Identities. *Mol. Cell* 38:576–589.
979 Hirata Y, Orth DN. 1979. Concentrations of Epidermal Growth Factor, Nerve Growth
980 Factor, and Submandibular Gland Renin in Male and Female Mouse Tissue
981 and Fluids*. *Endocrinology* 105:1382–1387.
982 Hiremath IS, Wagner CE. 2026. Mechanisms for Physicochemical Interaction between
983 Mucus and Respiratory Viruses. *ACS Biomater. Sci. Eng.* 12:13–30.
984 Horeth E, Oyelakin A, Song E-AC, Che M, Bard J, Min S, Kiripolsky J, Kramer JM,
985 Sinha S, Romano R-A. 2021. Transcriptomic and Single-Cell Analysis Reveals
986 Regulatory Networks and Cellular Heterogeneity in Mouse Primary Sjögren's
987 Syndrome Salivary Glands. *Front. Immunol.* 12:729040.
988 Hsieh LS, Wen JH, Miyares L, Lombroso PJ, Bordey A. 2017. Outbred CD1 mice are
989 as suitable as inbred C57BL/6J mice in performing social tasks. *Neurosci. Lett.*
990 637:142–147.
991 Hutson JM, Niall M, Evans D, Fowler R. 1979. Effect of salivary glands on wound
992 contraction in mice. *Nature* 279:793–795.
993 Ikemoto S, Matsushima Y. 1984. Sex and strain differences of mouse variant salivary
994 proteins. *Differentiation* 26:55–58.
995 Iwasaki T, Tokumori M, Matsubara M, Ojima F, Kamigochi K, Aizawa S, Ogoshi M,
996 Kimura AP, Takeuchi S, Takahashi S. 2023. A regulatory mechanism of mouse
997 kallikrein 1 gene expression by estrogen. *Mol. Cell. Endocrinol.* 577:112044.
998 Karageorgiou C, Ruhl S, Gokcumen O. 2025. Convergent evolution through
999 independent rearrangements in the primate amylase locus. Available from:
1000 http://biorxiv.org/lookup/doi/10.1101/2025.08.14.670395
1001 Karn R, Chung A, Laukaitis C. 2013. Shared and Unique Proteins in Human, Mouse
1002 and Rat Saliva Proteomes: Footprints of Functional Adaptation. *Proteomes*
1003 1:275–289.
1004 Karn RC, Laukaitis CM. 2011. Positive Selection Shaped the Convergent Evolution of
1005 Independently Expanded Kallikrein Subfamilies Expressed in Mouse and Rat
1006 Saliva Proteomes.Welch JJ, editor. *PLoS ONE* 6:e20979.
1007 Khoury A, Achinger-Kawecka J, Bert SA, Smith GC, French HJ, Luu P-L, Peters TJ,
1008 Du Q, Parry AJ, Valdes-Mora F, et al. 2020. Constitutively bound CTCF sites
1009 maintain 3D chromatin architecture and long-range epigenetically regulated
1010 domains. *Nat. Commun.* 11:54.
1011 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome
1012 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*
1013 37:907–915.
1014 Konttinen Y, Stegaev V, Mackiewicz Z, Porola P, Hänninen A, Szodoray P. 2010.
1015 Salivary glands - `an unisex organ'?: Salivary glands -'an unisex organ'? *Oral*
1016 *Dis.* 16:577–585.
1017 Krueger F, James F, Ewels P, Afyounian E, Weinstein M, Schuster-Boeckler B,
1018 Hulselmans G, Sclamons. 2023. FelixKrueger/TrimGalore: v0.6.10 - add

1019 default decompression path. Available from:
1020 https://zenodo.org/record/5127898
1021 Kryuchkova-Mostacci N, Robinson-Rechavi M. 2016. Tissue-Specificity of Gene
1022 Expression Diverges Slowly between Orthologs, and Rapidly between
1023 Paralogs.Ouzounis CA, editor. *PLOS Comput. Biol.* 12:e1005274.
1024 Kurabuchi S, Hosoi K, Gresik EW. 2001. Androgen Regulation of the Cellular
1025 Distribution of the True Tissue Kallikrein mK1 in the Submandibular Gland of
1026 the Mouse. *J. Histochem. Cytochem.* 49:801–802.
1027 Kurihara K, Nakanishi N, Tomomura A. 2011. Interference of kallikrein 1b26 (klk1b26)
1028 translation by microRNA specifically expressed in female mouse
1029 submandibular glands: an additional mechanism for sexual dimorphism of
1030 klk1b26 protein in the glands. *Biol. Sex Differ.* 2:13.
1031 Lai C-W, Chen H-L, Tsai T-C, Chu T-W, Yang S-H, Chong K-Y, Chen C-M. 2016.
1032 Sexually Dimorphic Expression of eGFP Transgene in the Akr1A1 Locus of
1033 Mouse Liver Regulated by Sex Hormone-Related Epigenetic Remodeling. *Sci.*
1034 *Rep.* 6:24023.
1035 Lawrence MG, Lai J, Clements JA. 2010. Kallikreins on Steroids: Structure, Function,
1036 and Hormonal Regulation of Prostate-Specific Antigen and the Extended
1037 Kallikrein Locus. *Endocr. Rev.* 31:407–446.
1038 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program
1039 for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930.
1040 Lindeman RE, Murphy MW, Agrimson KS, Gewiss RL, Bardwell VJ, Gearhart MD,
1041 Zarkower D. 2021. The conserved sex regulator DMRT1 recruits SOX9 in
1042 sexual cell fate reprogramming. *Nucleic Acids Res.* 49:6144–6164.
1043 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and
1044 dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.
1045 Malaker SA. 2024. Glycoproteomics: Charting new territory in mass spectrometry and
1046 glycobiology. *J. Mass Spectrom.* 59:e5034.
1047 Martin KR, Le HT, Abdelgawad A, Yang C, Lu G, Keffer JL, Zhang X, Zhuang Z, Asare-
1048 Okai PN, Chan CS, et al. 2024. Development of an efficient, effective, and
1049 economical technology for proteome analysis. *Cell Rep. Methods* 4:100796.
1050 Maruyama C, Monroe M, Hunt J, Buchmann L, Baker O. 2019. Comparing human and
1051 mouse salivary glands: A practice guide for salivary researchers. *Oral Dis.*
1052 25:403–415.
1053 Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2021. CAFE 5 models variation in
1054 evolutionary rates among gene families.Robinson P, editor. *Bioinformatics*
1055 36:5516–5518.
1056 Moustardas P, Yamada-Fowler N, Apostolou E, Tzioufas AG, Turkina MV, Spyrou G.
1057 2021. Deregulation of the Kallikrein Protease Family in the Salivary Glands of
1058 the Sjögren's Syndrome ERdj5 Knockout Mouse Model. *Front. Immunol.*
1059 12:693911.
1060 Mukaibo T, Gao X, Yang N, Oei MS, Nakamoto T, Melvin JE. 2019. Sexual
1061 dimorphisms in the transcriptomes of murine salivary glands. *FEBS Open Bio*
1062 9:947–958.

1063 Naqvi S, Godfrey AK, Hughes JF, Goodheart ML, Mitchell RN, Page DC. 2019.
1064 Conservation, acquisition, and functional impact of sex-biased gene expression
1065 in mammals. *Science* 365:eaaw7317.
1066 Nguyen CQ, Sharma A, Lee BH, She J-X, McIndoe RA, Peck AB. 2009. Differential
1067 gene expression in the salivary gland during development and onset of
1068 xerostomia in Sjögren's syndrome-like disease of the C57BL/6.NOD-
1069 Aec1Aec2mouse. *Arthritis Res. Ther.* 11:R56.
1070 Nikkanen J, Leong YA, Krause WC, Dermadi D, Maschek JA, Van Ry T, Cox JE,
1071 Weiss EJ, Gokcumen O, Chawla A, et al. 2022. An evolutionary trade-off
1072 between host immunity and metabolism drives fatty liver in male mice. *Science*
1073 378:290–295.
1074 Oppenheim FG, Hay DI. 1972. Further studies of human serum albumin in oral fluid.
1075 *Helv. Odontol. Acta* 16:22–26.
1076 Pajic P, Landau L, Gokcumen O, Ruhl S. 2025. Saliva Protein Genes in Humans were
1077 Shaped During Primate Evolution.Sohail M, editor. *Genome Biol. Evol.*
1078 17:evaf165.
1079 Pajic P, Pavlidis P, Dean K, Neznanova L, Romano R-A, Garneau D, Daugherity E,
1080 Globig A, Ruhl S, Gokcumen O. 2019. Independent amylase gene copy number
1081 bursts correlate with dietary preferences in mammals. *eLife* 8:e44628.
1082 Pajic P, Shen S, Qu J, May AJ, Knox S, Ruhl S, Gokcumen O. 2022. A mechanism of
1083 gene evolution generating mucin function. *Sci. Adv.* 8:eabm8757.
1084 Paltoglou S, Das R, Townley SL, Hickey TE, Tarulli GA, Coutinho I, Fernandes R,
1085 Hanson AR, Denis I, Carroll JS, et al. 2017. Novel Androgen Receptor
1086 Coregulator GRHL2 Exerts Both Oncogenic and Antimetastatic Functions in
1087 Prostate Cancer. *Cancer Res.* 77:3417–3430.
1088 Pardini LC, Taga R. 1996. Stereological Study of the Sexual Dimorphism in Mouse
1089 Submandibular Glands. *Okajimas Folia Anat. Jpn.* 73:119–124.
1090 Pavlopoulou A, Pampalakis G, Michalopoulos I, Sotiropoulou G. 2010. Evolutionary
1091 History of Tissue Kallikreins.DeSalle R, editor. *PLoS ONE* 5:e13781.
1092 Penschow JD, Drinkwater CC, Haralambidis J, Coghlan JP. 1991. Sites of expression
1093 and induction of glandular kallikrein gene expression in mice. *Mol. Cell.*
1094 *Endocrinol.* 81:135–146.
1095 Price SA, Hopkins SSB, Smith KK, Roth VL. 2012. Tempo of trophic evolution and its
1096 impact on mammalian diversification. *Proc. Natl. Acad. Sci.* 109:7008–7012.
1097 Reese RM, Harrison MM, Alarid ET. 2019. Grainyhead-like Protein 2: The Emerging
1098 Role in Hormone-Dependent Cancers and Epigenetics. *Endocrinology*
1099 160:1275–1288.
1100 Ruhl S. 2012. The scientific exploration of saliva in the post-proteomic era: from
1101 database back to basic function. *Expert Rev. Proteomics* 9:85–96.
1102 Saitou M, Gaylord EA, Xu E, May AJ, Neznanova L, Nathan S, Grawe A, Chang J,
1103 Ryan W, Ruhl S, et al. 2020. Functional Specialization of Human Salivary
1104 Glands and Origins of Proteins Intrinsic to Human Saliva. *Cell Rep.* 33:108402.
1105 Stopka P, Kuntová B, Klempt P, Havrdová L, Černá M, Stopková R. 2016. On the
1106 saliva proteome of the Eastern European house mouse (Mus musculus

1107 musculus) focusing on sexual signalling and immunity. *Sci. Rep.* 6:32481.
1108 Thamadilok S, Choi K-S, Ruhl L, Schulte F, Kazim AL, Hardt M, Gokcumen O, Ruhl
1109 S. 2020. Human and Nonhuman Primate Lineage-Specific Footprints in the
1110 Salivary Proteome.Nowick K, editor. *Mol. Biol. Evol.* 37:395–405.
1111 Treister NS, Richards SM, Lombardi MJ, Rowley P, Jensen RV, Sullivan DA. 2005.
1112 Sex-related Differences in Gene Expression in Salivary Glands of BALB/c Mice.
1113 *J. Dent. Res.* 84:160–165.
1114 True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary
1115 trajectories. *Evol. Dev.* 3:109–119.
1116 Uhlén M, Karlsson MJ, Hober A, Svensson A-S, Scheffel J, Kotol D, Zhong W, Tebani
1117 A, Strandberg L, Edfors F, et al. 2019. The human secretome. *Sci. Signal.*
1118 12:eaaz0274.
1119 UniProt Consortium. 2025. UniProt: the Universal Protein Knowledgebase in 2025.
1120 *Nucleic Acids Res.* 53:D609–D617.
1121 Van Leeuwen BH, Penschow JD, Coghlan JP, Richards RI. 1987. Cellular basis for
1122 the differential response of mouse kallikrein genes to hormonal induction.
1123 *EMBO J.* 6:1705–1713.
1124 Varki A, Schauer R. 2009. Sialic Acids. In: Varki A, Cummings RD, Esko JD, Freeze
1125 HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME, editors. Essentials of
1126 Glycobiology. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory
1127 Press. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1920/
1128 Villanea FA, Peede D, Kaufman EJ, Añorve-Garibay V, Chevy ET, Villa-Islas V, Witt
1129 KE, Zeloni R, Marnetto D, Moorjani P, et al. 2025. The *MUC19* gene: An
1130 evolutionary history of recurrent introgression and natural selection. *Science*
1131 389:eadl0882.
1132 Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-
1133 seq data: RPKM measure is inconsistent among samples. *Theory Biosci.*
1134 131:281–285.
1135 Winter EE, Goodstadt L, Ponting CP. 2004. Elevated Rates of Protein Secretion,
1136 Evolution, and Disease Among Tissue-Specific Genes. *Genome Res.* 14:54–
1137 61.
1138 Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, Grüning BA. 2020.
1139 Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and
1140 single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids*
1141 *Res.* 48:W177–W184.
1142 Wong HLX, Qin H, Tsang SW, Zuo X, Che S, Chow CFW, Li X, Xiao H, Zhao L, Huang
1143 T, et al. 2019. Early life stress disrupts intestinal homeostasis via NGF-TrkA
1144 signaling. *Nat. Commun.* 10:1745.
1145 Xu D, Pavlidis P, Thamadilok S, Redwood E, Fox S, Blekhman R, Ruhl S, Gokcumen
1146 O. 2016. Recent evolution of the salivary mucin MUC7. *Sci. Rep.* 6:31791.