Comparative transcriptomic analysis defines shared and mammary-specific gene expression programs across glandular tissues

Author Name(s)

Marie Saitou, Guro Katrine Sandvik

marie.saitou@nmbu.no

guro.katrine.sandvik@nmbu.no

Affiliation(s)

Faculty of Biosciences, Norwegian University of Life Sciences (NMBU)

Author Contributions

M.S. designed the study and conducted the bioinformatics analysis. G.K.S. interpreted glandular gene programs. Both authors contributed to writing the manuscript and approved the final version.

Abstract

The mammary gland is an epithelial secretory organ whose development and differentiation are regulated by hormonal and transcriptional programs. While mammary differentiation has been extensively studied, how mammary gene expression relates to that of other epithelial secretory organs remains incompletely defined.

Here, we performed a comparative transcriptomic analysis of human mammary tissue in the context of multiple epithelial secretory and non-glandular tissues using GTEx bulk RNA-seq data. By integrating cross-tissue differential expression analysis with single-cell expression references, co-expression network analysis, and evolutionary comparisons, we aimed to distinguish transcriptional programs that are shared across glandular epithelia from those that are specific to the mammary gland.

Our analyses identify three major classes of gene expression programs: glandular-enriched programs shared among epithelial secretory organs, mammary-specific programs associated with metabolic and hormone-responsive functions, and female-biased programs largely linked to immune-related processes. Comparative evolutionary analyses further suggest that mammary specialization predominantly reflects regulatory redeployment of conserved gene modules rather than the emergence of mammalian-specific genes.

Together, these findings place mammary gene expression within a broader framework of epithelial secretory biology and support a model in which mammary-specific functions arise through selective reweighting and organization of conserved transcriptional programs. This comparative perspective provides insight into the molecular basis of mammary differentiation and the evolutionary logic underlying specialization of epithelial secretory organs.

Keywords

mammary gland differentiation; secretory epithelia; comparative transcriptomics; gene co-expression; single-cell RNA-seq; evolutionary conservation; lipid metabolism

Introduction

The mammary gland is a secretory organ that undergoes coordinated changes in growth, differentiation, and function from embryogenesis through adulthood and lactation[1]. Its development involves interactions among epithelial, stromal, and immune compartments, regulated by hormonal signaling and transcriptional control mechanisms [2–5]. Previous studies have described cellular hierarchies, stem and progenitor populations, and endocrine regulation associated with mammary morphogenesis and milk production [5–7]. However, how mammary transcriptional programs relate to those of other epithelial secretory organs remains incompletely defined.

Anatomical and functional comparisons indicate that the mammary gland shares features with other exocrine tissues, including the salivary gland, pancreas, stomach, and skin [8]. These organs are all composed of epithelial-derived secretory structures, but differ in physiological roles and regulatory organization [9]. Distinguishing transcriptional programs that are common across glandular tissues from those that are specific to the mammary gland is necessary for understanding mammary differentiation and its evolutionary context [10]. This distinction is also relevant for disease studies, as pathways involved in normal mammary differentiation are frequently altered in breast cancer and other gland-associated disorders [11].

Although hormonal regulation and cell-type–specific differentiation within the mammary gland have been extensively studied [12], broader transcriptional relationships between the mammary gland and other epithelial secretory organs have received less attention. Many transcriptomic and single-cell studies have focused on individual tissues, providing detailed descriptions of cellular composition and lineage organization within specific glands. In the mammary gland, single-cell RNA sequencing of tissue and milk has characterized epithelial and immune cell populations across developmental and physiological states [13,14]. Comparable transcriptomic analyses have been performed for other secretory organs, including the salivary gland[15–18], pancreas [18–20], stomach and intestinal glands [21,22][23], and skin tissues [24]. These studies have primarily addressed tissue-specific questions. As a result, a systematic cross-tissue comparison that defines shared transcriptional features of glandular epithelia and identifies regulatory programs specific to the mammary gland remains limited. Without such comparative analyses, it is difficult to place mammary differentiation within a broader framework of epithelial secretory biology or to assess how mammary-specific functions relate to conserved glandular gene expression patterns.

To address this, we sought to characterize transcriptional features that distinguish the mammary gland from other epithelial secretory tissues in humans. By integrating comparative bulk transcriptomic data across multiple organs with cell-type–resolved expression references, we aimed to identify both shared glandular expression programs and regulatory features specific to mammary tissue. This framework is intended to provide a basis for comparative analyses of epithelial differentiation and secretory function across glandular organs.

Materials and Methods

Dataset Selection and Tissue Group Definition

Bulk RNA-seq expression data were obtained from the GTEx v10 dataset. Tissues were categorized into glandular and non-glandular groups for differential expression analyses.

The glandular group comprised six tissues characterized by epithelial origin and secretory function: Breast – Mammary Tissue, Minor Salivary Gland, Skin – Not Sun Exposed (Suprapubic), Pancreas, Stomach, and Small Intestine – Terminal Ileum. Breast, salivary gland, and skin were included due to their abundant exocrine acinar or ductal structures [25,26], despite differences in secreted products. Pancreas, stomach, and small intestine represent classical exocrine organs involved in digestive enzyme and fluid secretion. GTEx breast samples are derived from adult donors, and pregnancy or lactation status is not publicly available. Based on donor age distribution and sampling context, most breast tissues are interpreted as representing non-lactating adult mammary tissue.

The non-glandular group consisted of six tissues lacking glandular epithelium: Muscle – Skeletal, Heart – Left Ventricle, Artery – Tibial, Esophagus – Muscularis, Adipose – Subcutaneous, and Adipose – Visceral (Omentum).

Data Filtering

Samples were filtered using GTEx-provided quality metrics. Inclusion required RNA Integrity Number (SMRIN) ≥ 6.0, mapping rate (SMMAPRT) ≥ 0.8, sequencing efficiency (SMEXPEFF) ≥ 0.6, at least 50 million total reads (SMRDTTL), and 50 million uniquely mapped non-duplicate reads (SMMPPDUN). Samples were required to express at least 10,000 genes (SMGNSDTC).

Samples with high duplicate rates (SMDPMPRT > 0.6), elevated base mismatch rates (SMBSMMRT > 0.01), excessive chimeric reads (SMCHMRT > 0.015), UniVec contamination (SMUVCRT > 0.1), or extreme 3′ bias (SM3PBMN outside 0.4–0.75) were excluded. A minimum estimated library complexity of 30 million unique fragments (SMESTLBS) was required. Samples missing key metadata required for modeling were excluded. After filtering, samples from twelve tissue types were retained.

Covariates and Sample Metadata

Sample-level biological covariates included sex, age category, and Hardy death classification. The variable *SEX* is coded as 1 for male and 2 for female, while *AGE* corresponds to decade intervals (e.g., 20–29, 30–39, etc.). The *DTHHRDY* score describes the mode of death on a five-level Hardy scale, which serves as a proxy for postmortem condition.

Technical covariates included collection center (SMCENTER), nucleic acid batch (SMNABTCH), gene expression batch (SMGEBTCH), analyte type (ANALYTE_TYPE), and storage condition (SMAFRZE). Sequencing and RNA quality metrics used either for filtering or modeling included SMRIN, SMTSISCH, SMTSPAX, SMRDTTL, SMMAPRT, SMMPPDUN, SMEXPEFF, SMEXNCRT, SMRRNART, SMCHMRT, and SM3PBMN.

For differential expression modeling, SEX, AGE, SMCENTER, and SMEXPEFF were included as fixed effects. Samples missing any of these covariates were excluded prior to normalization and statistical analysis.

Differential Expression Analysis

Raw RNA-seq read count data from thirteen human tissues were obtained from the GTEx v10 dataset. Count matrices were imported directly from the compressed GTEx `.gct.gz` files, with sample columns matched to the curated metadata. The combined matrix was then used to create *DGEList* objects in *edgeR* *(Robinson et al. 2010)*, and lowly expressed genes were excluded using the `filterByExpr()` function with the relevant grouping factor. Counts were normalized by the trimmed mean of M-values (TMM) method to correct for differences in library size and sequencing depth.

Differential expression analyses were performed under several contrasts. In the first analysis, glandular tissues were compared to non-glandular tissues to identify genes generally enriched in epithelial secretory organs. In the second analysis, mammary gland samples were contrasted specifically against non-glandular tissues to reveal mammary-specific transcriptional programs. A third comparison was carried out within mammary tissues to detect sex-dependent expression differences between male and female samples. For each analysis, a linear modeling framework was applied using the *limma–voom* pipeline. Coefficients corresponding to the main factor of interest (glandular versus non-glandular tissues, mammary versus non-glandular tissues, or female versus male in mammary glands) were automatically identified from the fitted model. Genes were considered significantly differentially expressed when the absolute log□ fold change exceeded one and the false discovery rate (Benjamini–Hochberg adjusted p-value) was below 0.001.

Differentially expressed genes were grouped into three functional expression categories based on their tissue and sex specificity. Genes upregulated in both the mammary gland and other glandular tissues were classified as (1) “Glandular-enriched genes”. Genes upregulated exclusively in the mammary gland, but not in other glandular tissues, were defined as (2) “Mammary-specific genes”. Finally, genes

exhibiting higher expression in females relative to males were categorized as (3) “Female-biased genes”.

Integration of Bulk RNA-seq data and Single-cell RNA-seq Data

Single-cell RNA-seq datasets of breast tissue and subcutaneous adipose tissue [27,28] were downloaded at CELLxGENE Discover portal [29]. We used publicly available single-cell RNA-seq data from the Human Breast Cell Atlas (HBCA, global dataset, ~803,000 cells), generated from normal adult human breast tissue and snRNA-seq of human subcutaneous adipose tissue - all cells (37,879 cells, GSE281356) [27,28]. In this study, we used subcutaneous adipose tissue as the comparator for reference breast adipose data because the adipose component of the human breast is anatomically and histologically classified as subcutaneous adipose tissue. The human breast is composed of glandular (mammary) and adipose tissue embedded in a stromal framework beneath the skin, with adipose tissue forming the major volume of the breast outside the glandular component[30]. Adipose tissue located beneath the skin is defined as subcutaneous adipose tissue, in contrast to visceral adipose tissue, which is localized around internal organs within the abdominal cavity [31].

Cell-type composition of bulk mammary gland RNA-seq data was estimated using a reference-based deconvolution approach implemented in the DeconRNA-seq [32] package (Bioconductor). Reference expression matrices were derived from single-cell RNA-seq datasets of human mammary and subcutaneous adipose tissues [27,28]. For each dataset, mean expression profiles were calculated for every annotated cell type, with genes as rows and cell types as columns.

From the resulting deconvolution output, adipocyte-related fractions were obtained by summing columns annotated as “adipocyte,” “mesenchymal stem cell of adipose tissue,” or similar lipid-associated terms. These estimated fractions were then integrated with GTEx sample metadata containing sex and age information. Age was converted to numeric values where possible and categorized into five intervals (≤30, 31–40, 41–50, 51–60, and ≥61 years). While pregnancy/lactation status was not available in the sample metadata, among female donors, samples from individuals aged 39 years or younger accounted for approximately 17% of the dataset, with the majority derived from donors aged 40 years and above.

Single-cell Expression Profiling and Gene Clustering

Single-cell reference data were processed in R using the *zellkonverter* and *DelayedArray* packages to enable efficient handling of large HDF5-backed matrices. Annotated expression data were imported from the `.h5ad` file, and cells were grouped according to their annotated cell type. To focus on biologically relevant features, only genes previously identified as differentially expressed were retained.

For each cell type, we calculated the detected fraction, defined as the proportion of cells in which a given gene was expressed (nonzero counts). This metric was used as the sole quantitative representation of gene activity across cell populations.

Detected-fraction matrices were subjected to hierarchical clustering (Euclidean distance, complete linkage) to identify groups of genes with similar cell-type specificity. Each dendrogram was divided into three clusters, and functional enrichment analyses were performed for each cluster using *g:Profiler* (*gprofiler2* package) with Gene Ontology (BP, MF, CC), KEGG, and Reactome databases. Enrichment significance was assessed using the false discovery rate (FDR) correction.

Evolutionary Conservation Analysis

To assess the evolutionary conservation of mammary gland–related gene expression programs, we compiled a list of human genes differentially expressed in mammary and other glandular tissues and examined their orthologous relationships across representative vertebrate lineages using Ensembl BioMart (release ≥111). The ortholog presence matrix included 13 species covering major vertebrate clades: Eutheria (*mouse, Mus musculus; rat, Rattus norvegicus; rabbit, Oryctolagus cuniculus; pig, Sus scrofa; cattle, Bos taurus; goat, Capra hircus; sheep, Ovis aries*), Marsupialia (koala, *Phascolarctos cinereus*), Monotremata (platypus, *Ornithorhynchus anatinus*), Aves (chicken, *Gallus gallus*), Amphibia (*Xenopus tropicalis*), and Teleostei (zebrafish, *Danio rerio*; medaka, *Oryzias latipes*).

For each human Ensembl gene ID, ortholog presence (1) or absence (0) was determined in each species, resulting in a binary matrix that retained the original annotation of mammary, glandular, and sex-related differential expression. Genes were grouped into three expression categories based on the direction of differential expression in human tissues. The proportion of genes possessing identifiable orthologs in each species was calculated for each category, using the number of human genes in the category as the denominator.

GWAS-based Functional Enrichment Analysis

To assess the relevance of the differentially expressed gene (DEG) categories to complex human traits, we integrated our DEG dataset with the GWAS Atlas (version 20191115) using gene-level association statistics generated by MAGMA.

First, we imported three datasets: (i) the DEG table, (ii) GWAS Atlas trait metadata (`gwasATLAS_v20191115.txt.gz`), and (iii) MAGMA gene-level association results (`gwasATLAS_v20191115_magma_P.txt.gz`).

MAGMA p-values were converted to false discovery rates (FDR) using the Benjamini–Hochberg method, and traits with FDR < 0.0001 were retained. For each gene, only the most significant trait association (smallest FDR) was retained as its *primary trait*. This resulted in a mapping of each significantly associated gene to a single SubchapterLevel trait category defined in the GWAS Atlas. The resulting gene–trait associations were then intersected with the three DEG categories derived from the transcriptomic analyses. To establish a null background, 1,000 random genes were sampled from the MAGMA dataset and processed in the same way to obtain background counts of trait associations.

For each combination of DEG category and GWAS trait, Fisher's exact test was applied to test for enrichment relative to the background set. The log□ enrichment ratio was computed as:

$$\log_2 \text{enrich} = \log_2 \left(\frac{(a + 0.5)/(M + 1)}{(c + 0.5)/(N + 1)} \right)$$

where *a* is the number of DEG-associated genes for a given trait, *M* is the total number of genes in that DEG category, *c* is the number of background genes associated with the trait, and *N* = 1000 is the

background total. Trait categories with FDR < 0.05 were retained for downstream interpretation. All analyses were performed in R (version 4.5.2) using *tidyverse* and *stats* packages, with random seed fixed at 20250817 for reproducibility.

Weighted gene co-expression network analysis (WGCNA) and cross-tissue module comparison

Normalized gene expression counts from the GTEx dataset were used to construct tissue-specific co-expression networks using the WGCNA package (v1.72). For the mammary gland, we first regressed out cell-type composition effects estimated by DeconRNA-seq, specifically the fractions of endothelial and macrophage cells, using `removeBatchEffect` in the limma package. Log-transformed CPM values were filtered to retain genes expressed in at least 20% of samples, and a signed network was built with bicor correlation and dynamic tree cutting (minimum module size = 120, merge cut height = 0.35). Modules were labeled in descending order of size (Module 1, Module 2, …), and module eigengenes (MEs) were extracted for downstream analyses.

To evaluate cross-tissue conservation of co-expression structure, module assignments from the mammary gland were compared with those from other GTEx tissues. For each tissue, Jaccard similarity was computed between all pairs of mammary and tissue modules based on shared gene memberships. For each mammary module, the maximum Jaccard index across modules in a given tissue was retained (“best-per-tissue” score). Permutation tests (n = 2000) were performed by shuffling gene labels in the comparison tissue while preserving module sizes, yielding empirical p-values and FDR-corrected significance estimates.

Results

Comparative transcriptomic analyses reveal glandular-enriched, mammary-specific, and female-biased gene expression programs

Principal component analysis (PCA) was performed using the top 2000 most variable genes across 12 tissue types. The first two components explained 37.6% (PC1) and 19.7% (PC2) of the variance, respectively. Samples from glandular tissues, including breast, salivary gland, pancreas, stomach, small intestine, and skin, formed a coherent cluster that was clearly separated from non-glandular tissues such as muscle, heart, esophagus, and adipose. Breast samples were positioned between the glandular and adipose clusters, while remaining closer to other glandular tissues than to non-glandular tissues. Notably, breast, salivary gland, and pancreas localized in relative proximity within the PCA space (**Figure 1**).

We next compared gene expression profiles across three functional categories defined by tissue-specific differential expression patterns. Differentially expressed genes were grouped into three functional expression categories based on their tissue- and sex- biased enrichment. Genes upregulated in both the mammary gland and other glandular tissues were classified as (1) “Glandular-enriched genes” (711 genes). Genes upregulated exclusively in the mammary gland, but not in other glandular tissues, were defined as (2) “Mammary-specific genes” (189 genes). Finally, genes exhibiting higher expression in females relative to males were categorized as (3) “Female-biased genes” (617 genes) (**Figure 2A**; **Table S1**). We observed substantial sharing between glandular-enriched and female-biased genes, whereas mammary-specific genes formed a largely distinct group with minimal overlap with either category (**Figure 2B**). Functional enrichment analysis further showed that these gene sets were associated with different biological processes (**Figure 2C; Table S2**). Glandular-enriched genes were predominantly enriched for immune-related and extracellular functions, mammary-specific genes

were enriched for metabolic and endocrine pathways, and female-biased genes showed strong enrichment for immune and complement-associated processes.

In the “Glandular-enriched genes” group, we identified a cluster of genes associated with epithelial structure, secretion, and ion transport. This set included *DSP* (desmoplakin), *CDH1* (E-cadherin), *GRHL2*, and *BARX2*, which regulate epithelial integrity and branching morphogenesis, as well as several solute carriers *(SLC13A2, SLC5A1, SLC15A1*) and secretory enzymes (*PRSS8, PRSS22, SULT2B1*). In contrast, the “Mammary-specific genes” group was characterized by genes associated with hormonal signaling, metabolic specialization, and epithelial differentiation. This set included hormone- and metabolite-responsive receptors such as *OXTR* (oxytocin receptor) and *HCAR1* (lactate receptor), together with the developmental regulator *TBX3* and the mammary-enriched long non-coding RNA *LINC00993*. Additional genes in this category comprised regulators of lipid metabolism, immune-associated processes, and cellular stress responses (e.g. *CREB3L4, STC2, CMTM7, SAA2,* and *SAA4*). The group further contained multiple transcription factors and structural or metabolic genes linked to epithelial organization and function, including *TWIST1, HOXD9, TRPS1, CLPSL1, PEMT,* and *SLC5A6*. Finally, the “Female-biased genes” category was dominated by immunoglobulin genes, including IGHV, IGKJ, IGLJ, and IGKC family members, together with POU2AF1 and CD27.

Single-cell mapping and bulk deconvolution resolve the cellular architecture underlying mammary gene expression programs

Bulk RNA-seq data from GTEx breast tissue were deconvoluted using single-cell mammary and adipose reference profiles from CELLxGENE Discover portal to estimate cell-type composition across

sex and age groups. Across all age groups and in both sexes, breast tissue samples consistently contained a substantial adipose-associated fraction, accounting for approximately half of the total cellular composition (**Figure 3**). While the relative proportion of adipose-associated cells showed some variability across groups, the overall composition was largely consistent between males and females across age groups.

The remaining fraction was primarily composed of mammary epithelial populations, including basal–myoepithelial and perivascular cells, with smaller contributions from macrophage and endothelial compartments. Given the age distribution of GTEx breast sample donors, which is skewed toward middle-aged and older individuals, the observed cellular composition is consistent with a predominantly non-lactational mammary tissue state. This cellular profile is also concordant with the intermediate positioning of breast tissue between adipose and glandular tissues observed in the global PCA of tissue transcriptomes (**Figure 1A**).

To determine how each differential expression category is distributed across mammary cell populations, we examined the cellular expression patterns of genes belonging to the three defined groups: “Glandular-enriched genes”, “Mammary-specific genes”, and “Female-biased genes”. For each gene, we calculated the detected fraction, defined as the proportion of cells within each annotated mammary cell type in which the gene was expressed (**Figure 4**).

Within the “Glandular-enriched genes” category, hierarchical clustering separated genes into two major clusters reflecting distinct single-cell detection patterns. One cluster comprised genes broadly detected across multiple mammary epithelial populations, including genes associated with vesicular transport and protein secretion, such as *PRSS22*. A second cluster showed preferential detection in epithelial cells, secretory precursor and hormone-sensing cells. These genes are associated with extracellular matrix and immune-related components, including *COL1A1*, *CXCL12*, and *ICAM1*.

For “Mammary-specific genes”, cluster 1 showed a broad and heterogeneous expression pattern and included genes linked to mammary differentiation and metabolic specialization such as *BTN1A1*, and

LALBA. Cluster 2 genes were expressed in mammary epithelial and luminal secretory precursor cells, and encompassed genes associated with hormone-responsive differentiation and developmental control, such as *ESR1, GATA3*, and *TBX3*.

Genes in the “Female-biased genes” category also segregated into two major cellular expression patterns. Cluster 1 showed broad expression, and included estrogen-driven epithelial proliferation genes and stromal signaling genes, including *ESR1, PGR* and *IGF1*. Cluster 2 genes were expressed in mammary epithelial and luminal secretory precursor cells, and associated with hormone-linked inflammatory and oxidative stress regulation, including *SAA1, HMOX1* and *IL6*.

Comparative genomics and GWAS integration place mammary transcriptional layers in evolutionary and functional context

To examine whether genes defining mammary transcriptional programs represent evolutionarily recent innovations in mammals or derive from more ancient gene repertoires in broader vertebrates, we compared the evolutionary conservation of the three gene categories identified in this study. Specifically, for genes classified as “Glandular-enriched genes”, “Mammary-specific genes”, or “Female-biased genes”, we assessed whether their orthologous counterparts could be identified across vertebrate species, including mammalian and non-mammalian lineages, spanning monotremes, birds, amphibians, and teleost fish (**Table S3**).

Quantitative comparison of ortholog retention ratios revealed a significant difference among the three gene categories (Kruskal–Wallis test, $\chi^2 = 32.25$, df = 2, $p = 9.9 \times 10^{\square\square}$; **Figure 5**). Pairwise comparisons showed that “Mammary-specific genes” exhibited significantly higher evolutionary conservation than both “Glandular-enriched genes” (Wilcoxon test, $p = 3.5 \times 10^{\square\square}$) and “Female-biased genes” (Wilcoxon test, $p = 6.1 \times 10^{\square\square}$).

To further assess whether these gene categories are differentially associated with phenotypic variation in humans, we examined their enrichment across trait domains reported in genome-wide association studies (GWAS) using gene-level association statistics from the GWAS Atlas [33] using gene-level association statistics from MAGMA [34]. For each gene, the most strongly associated trait (smallest FDR) was retained, and enrichment within each expression-defined gene category was evaluated against a background of 1,000 randomly selected genes using Fisher's exact tests. Trait associations passing an FDR threshold of 0.05 were retained, and log enrichment ratios were calculated to summarize the strength and direction of associations (**Table 1**). "Glandular-enriched genes" were significantly associated with the *Immunological System Functions* and *Protein Level* trait domains. "Mammary-specific genes" showed associations with developmental and ectodermal phenotypes, particularly the *Functions of Hair* and *Height* categories.In contrast, "Female-biased genes" were associated with *Immunological System Functions*, *Protein Level*, and *Functions of Hair*.

Mammary co-expression modules and their overlap with other glandular tissues

To assess how these transcriptional signatures are organized into gene networks within the mammary gland and other tissues, we performed weighted gene co-expression network analysis (WGCNA). WGCNA of mammary gland transcriptomes identified 14 co-expression modules, ranging in size from 232 to 2,429 genes (median ≈ 540 genes). The two largest modules (Modules 1 and 2) contained more than 1,000 genes each, whereas the remaining modules comprised between 200 and 900 genes. To assess cross-tissue similarity of mammary co-expression structure, we calculated pairwise Jaccard indices between mammary modules and modules derived from 12 other GTEx tissues **(Figure S1**).

Overall overlap between mammary modules and those from other tissues was low, with median Jaccard indices below 0.1 and no significant difference between glandular and non-glandular tissue groups (Wilcoxon test, $p > 0.05$). Among all modules, Mammary Module 2 showed the highest and most recurrent overlap with modules from several glandular tissues, including salivary gland, pancreas, stomach, and skin (**Figure 6A**).

To characterize the gene composition underlying the overlaps between Mammary Module 2 and other tissues, we examined gene subsets shared between Mammary Module 2 and one or more comparison tissues. For each comparison tissue, only the module showing the highest overlap with Mammary Module 2 was considered (**Figure 6B**). Genes unique to the mammary module were enriched for pathways related to lipid metabolism and mitochondrial function, including fatty acid biosynthesis, β-oxidation, and PPAR signaling, as well as hormone-associated signaling pathways. Gene sets shared between the mammary gland and salivary gland or pancreas were enriched for pathways related to vesicle trafficking, Golgi–ER processing, glycosylation, and exocytosis. Gene sets shared across the mammary gland, salivary gland and stomach, were enriched for extracellular matrix organization (**Table S4**).

Discussion

Our analyses indicate that the mammary gland shares a substantial transcriptional component with other epithelial secretory organs in humans. Genes classified as glandular-enriched were expressed in the mammary gland as well as in salivary gland, pancreas, stomach, and skin, and were primarily associated with epithelial structure, secretory processes, extracellular matrix organization, and immune-related functions. The glandular-enriched genes identified here are consistent with prior transcriptomic analyses in mouse, which showed that exocrine organs such as the salivary gland, pancreas, and skin

exhibit broadly similar epithelial gene expression profiles at the global level [35]. Single-cell mapping further showed that these shared transcriptional programs were predominantly detected in luminal epithelial populations, suggesting that common glandular features are largely carried by epithelial cell types with secretory or barrier-related functions.

Genes classified as glandular-enriched included regulators of epithelial adhesion and associated with breast cancer, such as *CDH1* [36,37], transcription factors implicated in epithelial differentiation including *GRHL2[38,39]* and *BARX2[40]*. These genes have also been extensively studied in the context of breast cancer, where they are linked to epithelial identity and disease-associated transcriptional programs. Glandular-enriched genes also included solute carriers involved in epithelial transport, such as *SLC15A1*, which is well characterized for its duodenal-specific expression and role in peptide transport in the human upper gastrointestinal tract (Thwaites and Anderson, 2007; Hediger et al., 2013). Its presence among glandular-enriched genes is consistent with the reuse of established epithelial transport mechanisms across secretory and absorptive tissues.

In contrast, mammary-specific genes were enriched for regulators of hormonal responsiveness, metabolism, and epithelial differentiation. This group included hormone-related receptors such as *OXTR* [41–44] and *HCAR1[45–47]*, transcription factors involved in mammary development and lineage specification including *TBX3 [48]*, *TRPS1[49]*, and *TWIST1[50,51]*, and metabolic regulators such as *PEMT* [52] and *SLC5A6[53]*. The functional composition of this gene set is consistent with the metabolic and endocrine requirements of the mammary gland, particularly in relation to milk production. Some of these mammary-specific genes are known to respond to hormonal cues involved in mammary differentiation and secretory activation, and their expression patterns align with established roles of

endocrine and metabolic regulation in mammary epithelial function [44]. Together, these findings suggest that mammary specialization is associated with selective enrichment of metabolic and hormone-responsive programs layered onto a shared glandular transcriptional framework.

Genes exhibiting female-biased expression were dominated by B cell-related immunoglobulin genes, including *IGKC*, *IGHM*, *IGHV*, and *IGKJ* family members, which are well characterized as components of B cell and antibody-mediated immune responses [54][55]. In addition, immune-associated regulators such as *POU2AF1* [56] and *CD27* [57], which are primarily involved in T cell function and immune modulation, were also enriched. The differential expression of these genes in female mammary tissue likely reflects both the representation of B and T cells and their activation states within the gland.

Taken together, these results support an interpretation in which mammary gland specialization arises through the reuse of genes and transcriptional modules that are shared with epithelial secretory tissues. Rather than representing a wholly distinct transcriptional program or mammalian specific novel genes, the mammary gland appears to combine conserved glandular modules with a more limited set of regulatory features that are enriched in mammary tissue. This layered organization is supported by the co-expression analyses presented in this study, which reveal partial overlap between mammary co-expression modules and those identified in other secretory tissues. Within this framework, mammary-associated genes can be interpreted as components of conserved regulatory modules that have been redeployed in a mammary-specific context. One example is *STC2 (Stanniocalcin 2),* a mammary-specific gene in our analysis. Although stanniocalcins originated as calcium-regulatory hormones in non-mammalian vertebrates, STC2 is expressed in adult, non-lactational mammary tissue and likely

reflects redeployment of an ancestral endocrine regulator within a mammary-specific regulatory context [58].

From an evolutionary perspective, this pattern is compatible with models of the emergence of new traits that emphasize regulatory divergence within conserved genetic frameworks. Studies in both plants and animals have shown that novel organ morphologies often emerge through changes in the deployment, connectivity, or regulation of pre-existing gene modules rather than through the invention of new genes (e.g. “old genes, new functions”) [59–62]. In this context, the mammary gland can be viewed as a mosaic organ, assembled from ancestral epithelial and secretory modules that are shared with other glands, together with regulatory modifications that support mammary-specific metabolic, endocrine, and reproductive functions. This interpretation is further supported by the evolutionary conservation of many mammary-associated genes across vertebrates, indicating that mammary specialization likely reflects regulatory reorganization of conserved gene repertoires rather than lineage-specific gene innovation.

Integration of gene-level GWAS associations provided functional context for the transcriptional programs identified in this study. Using MAGMA-based gene-level statistics, enrichment was observed for trait domains related to immune function, growth, and ectodermal biology, including height- and hair-related traits. These trait categories align with core functional properties of epithelial secretory tissues, which integrate barrier immunity [63,64], epithelial turnover[65], and growth-related processes [66], and reflect the shared developmental and physiological roles of glandular epithelia.

By contrast, no significant enrichment of gene numbers in breast cancer–specific trait categories was detected among glandular-enriched, mammary-specific, or female-biased genes in this category-wide analysis. This absence does not preclude roles for individual genes in cancer-related processes, but indicates that the transcriptional programs identified here are more strongly associated with broadly polygenic traits linked to epithelial and glandular function than with disease categories driven by a more

limited number of high-impact loci, which may be less readily captured by MAGMA-based enrichment approaches [67].

Together, these findings place the mammary gland within a continuum of epithelial secretory tissues that share conserved transcriptional and functional foundations. Mammary specialization appears to arise not through the acquisition of many mammary-specific unique gene sets, but through the selective reuse, reweighting, and regulatory redeployment of ancestral glandular modules. Viewing mammary gene expression through this modular and comparative framework provides a basis for integrating developmental, evolutionary, and functional perspectives on mammary biology. More broadly, this approach may be applicable to other epithelial organs that exhibit both shared ancestry and organ-specific physiological specialization.

Data Availability

All data and analysis scripts supporting this study are available at: https://github.com/mariesaitou/paper_2025-/tree/main/mammary_gland

The study utilized publicly accessible human transcriptomic datasets from the GTEx Consortium (v10) and the CZ CELLxGENE Discover portal, including the Human Breast Cell Atlas and human subcutaneous adipose tissue datasets (See **Materials and Methods**). No additional restrictions apply to data sharing.

Funding

This research did not receive any specific grant from funding agencies.

Acknowledgments

Computational analyses were performed using the High Performance Computing (HPC) cluster *Orion* at the Norwegian University of Life Sciences (NMBU). We thank Sabine Anne-Lie Ferneborg (NMBU, BIOVIT) for helpful comments and constructive feedback on the manuscript.

Author Contributions

M.S. designed the study and conducted the bioinformatics analysis. G.K.S. interpreted glandular gene programs. Both authors contributed to writing the manuscript and approved the final version.

Competing Interests

The authors declare no competing financial or non-financial interests.

Ethical Approval

This study utilized publicly available human transcriptomic datasets (GTEx and CellxGene). No new human or animal samples were collected for this work.

Figures

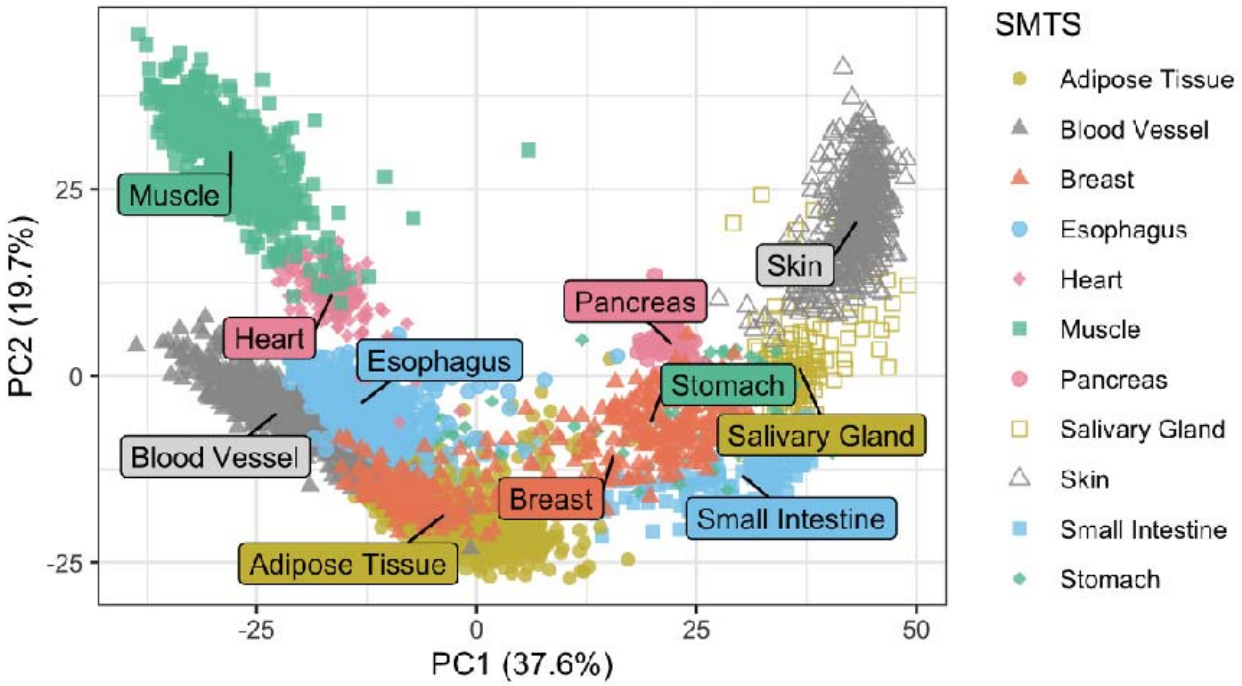

Figure 1. Principal component analysis (PCA) of tissue transcriptomes

Principal component analysis (PCA) was performed using the 2,000 genes with the highest expression variance across tissues, to summarize the overall structure of transcriptome variation. Each point represents a tissue sample, and tissues are colored according to their anatomical source (legend, right), Glandular tissues (breast, salivary gland, pancreas, stomach, small intestine, skin) cluster separately from non-glandular tissues (muscle, heart, blood vessel, esophagus, adipose). The x-axis (PC1) and y-axis (PC2) show the first and second principal components, which together explain the largest proportions of total expression variance (37.6% and 19.7%, respectively). Samples that are closer together in this plot have more similar global expression profiles, whereas those farther apart differ more strongly in their transcriptome composition. Colored clusters correspond to distinct tissue types. Labels mark the centroids of major tissue groups to aid visual interpretation.

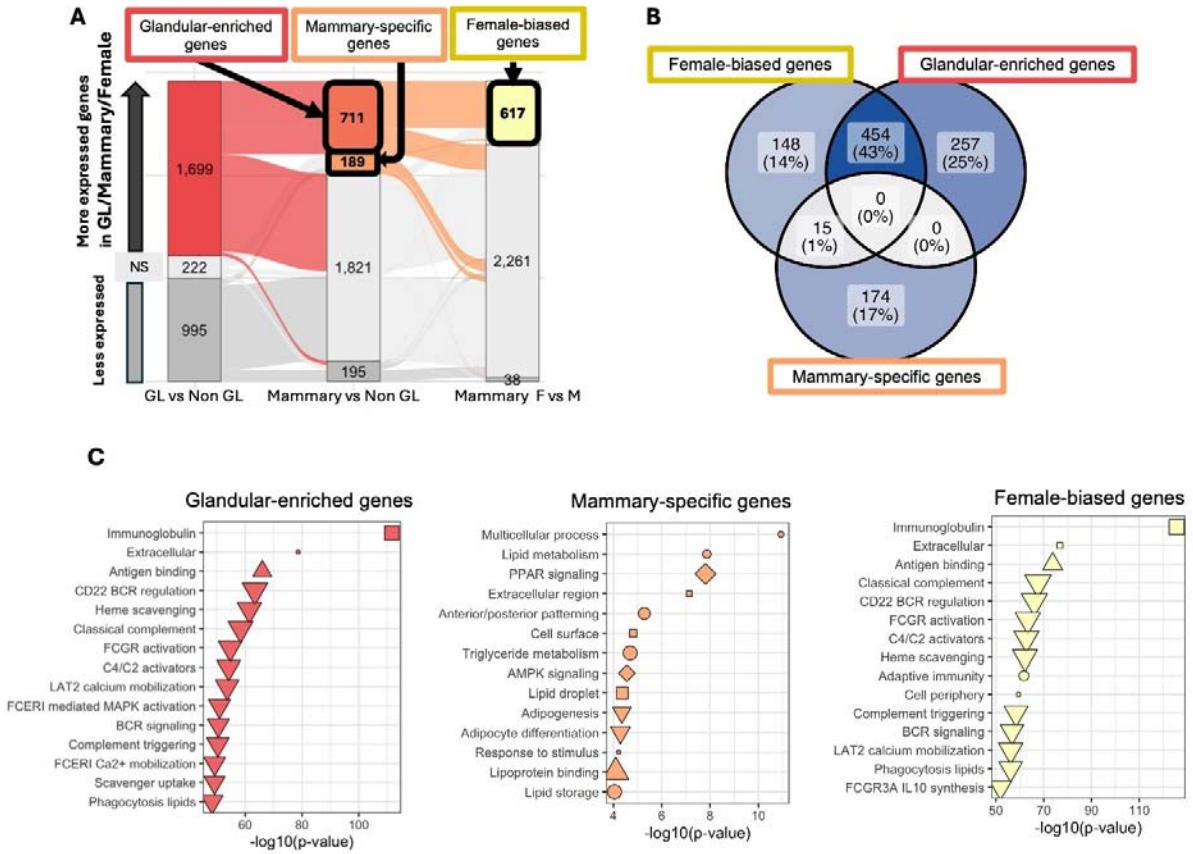

Figure 2. Differentially expressed gene categories across comparisons

(A) Sankey diagram summarizing the number of differentially expressed genes (DEGs) identified in three pairwise comparisons: (i) all the glandular tissues versus non-glandular tissues, (ii) mammary gland versus non-glandular tissues, and (iii) female versus male mammary gland. Bars indicate the number of upregulated (colored), downregulated (grey), and non-significant (NS; white) genes. Flows illustrate overlaps of gene categories across comparisons.

(B) Venn diagram illustrating the overlap among three categories of upregulated genes: glandular-enriched genes, mammary-specific genes, and female-biased genes. Numbers and percentages indicate the proportion of genes within each intersection.

(C) Gene Ontology (GO) and pathway enrichment analyses were performed using g:Profiler for upregulated genes in three DEG categories: (i) glandular tissues versus non-glandular tissues, (ii) mammary gland versus non-glandular tissues, and (iii) female versus male mammary gland.

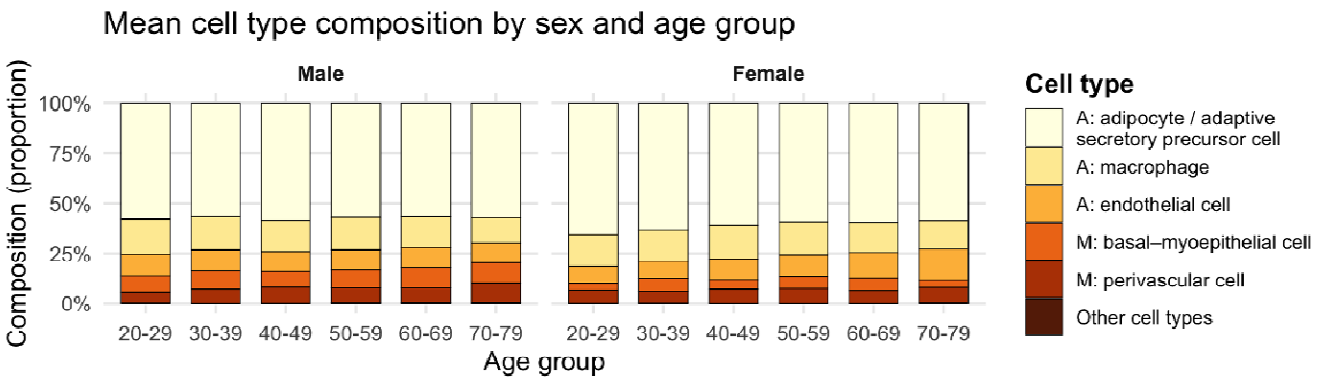

Figure 3. Mean cell type composition in breast tissue by sex and age group

Bulk RNA-seq data from GTEx breast tissue were deconvoluted using single-cell reference expression profiles to estimate relative cell-type composition. The deconvolution revealed the relative proportions of major mammary cell types, averaged within each sex (male, female) and age group (10-year bins). Stacked bar plots display mean fractional compositions for each group, with the y-axis normalized to 100%. Colors denote inferred cell populations: adipocyte / adaptive secretory precursor cell (A: Adipose), macrophages (A: Adipose), endothelial cells (A: Adipose), basal–myoepithelial cells (M: Mammary),perivascular cells (M: Mammary), and minor fractions classified as “Other cell types.” Single-cell RNA-seq datasets of breast tissue and subcutaneous adipose tissue [27,28] were downloaded at CELLxGENE Discover portal [29].

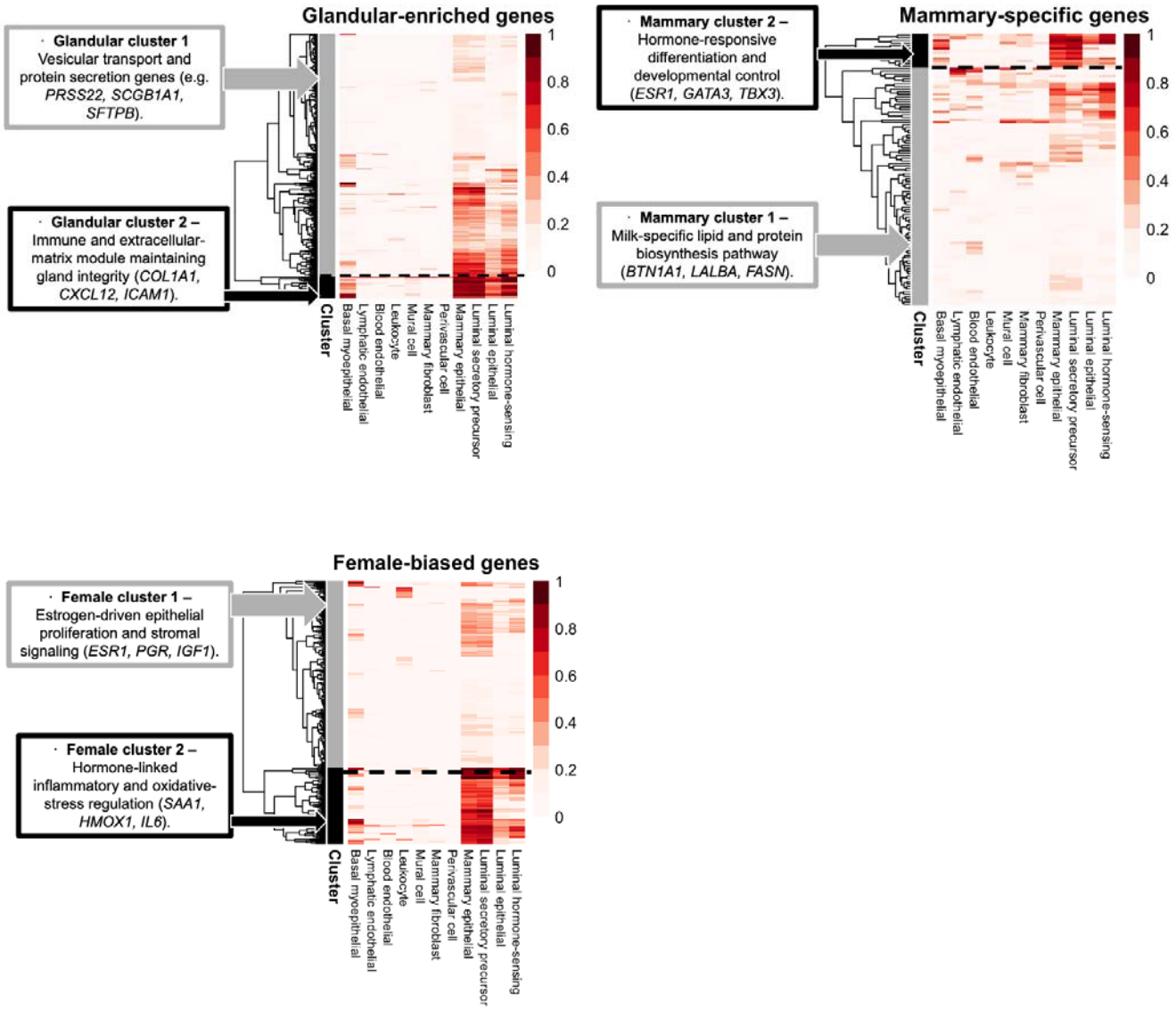

Figure 4. Single-cell expression patterns of upregulated gene sets

Heatmaps showing the proportion of single-cell populations expressing upregulated genes in each DEG category: (i) genes upregulated across glandular tissues, (ii) genes upregulated only in mammary gland, and (iii) genes upregulated in female versus male mammary gland. Cell types include basal myoepithelial cells, luminal epithelial subtypes (hormone-sensing, secretory precursor, differentiated luminal), fibroblasts, endothelial cells (blood and lymphatic), mural cells, perivascular cells, and leukocytes. Expression levels are represented by scaled proportions of expressing cells (0–1), with rows corresponding to genes and columns to cell types.

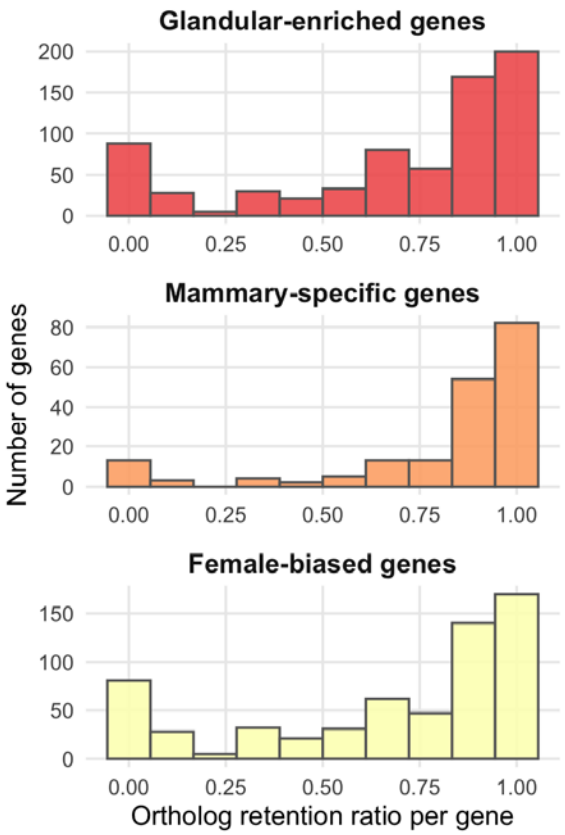

Figure 5. Evolutionary conservation of mammary gene categories across vertebrate species. For each gene category, the presence or absence of orthologs of human genes, (genes in other species that descend from a common ancestral gene), was examined across vertebrate species: *Mouse*, *Rat*, *Cattle*, *Goat*, *Sheep*, *Pig*, *Rabbit*, *Koala*, *Platypus*, *Chicken*, *Xenopus tropicalis*, *Zebrafish*, and *Medaka*. A gene's ortholog retention ratio represents the fraction of these species in which a corresponding ortholog is detected (0 = no orthologs retained, 1 = orthologs present in all species). Histograms display the distribution of retention ratios for three categories: Glandular-enriched genes (red), Mammary-specific genes (orange), and Female-biased genes (yellow). The x-axis shows the ortholog retention ratio per gene, and the y-axis indicates the number of genes in each bin.

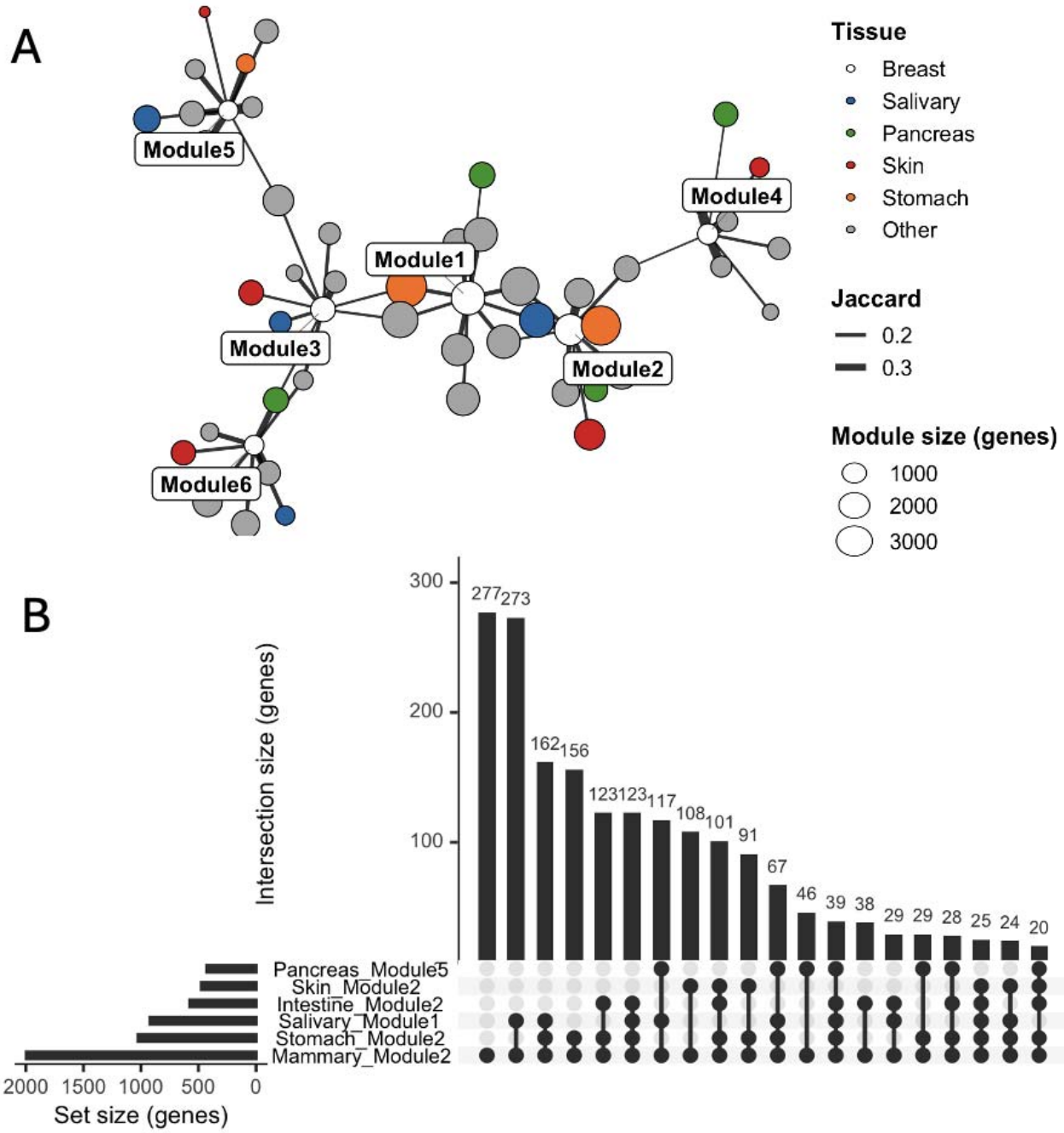

Figure 6. Cross-tissue network comparison of mammary gene co-expression modules and their gene set overlaps.

(A) Network visualization of breast-specific gene co-expression modules (Modules 1–6) and their best-matching modules identified across other human tissues. Each node represents a gene co-expression module, defined as a cluster of genes that show highly correlated expression patterns within a given

tissue, derived from tissue-resolved network analysis. Node size corresponds to the number of genes included in each module. Node color indicates the tissue of origin, with white nodes representing breast modules and colored nodes representing modules from other glandular tissues. Edge thickness is proportional to the Jaccard index value, such that thicker lines indicate stronger gene overlap between modules.

(B) UpSet plot summarizing gene set intersections between breast modules and their top-matching modules from other tissues. Each filled dot in the bottom matrix represents the inclusion of a particular module in a given intersection set, and connecting lines indicate combinations of modules that share overlapping genes. The vertical bars above the matrix show the number of genes (intersection size) for each specific module combination, while the horizontal bars on the left indicate the total number of genes (set size) contained in each module.

Tables

Category	Subchapter Level	k	M	bg k	a	b	c	d	pval	$\log_2$(enrich)	FDR
Glandular enriched genes	Immunological System Functions	73	325	141	73	252	141	959	6.7×10^{-4}	0.87	0.023
Glandular enriched genes	Protein Level	12	325	10	12	313	10	990	0.0022	1.87	0.048
Mammary specific genes	Functions of Hair	10	121	11	10	111	11	989	1.9×10^{-5}	2.91	0.0029
Mammary specific genes	Height	32	121	140	32	89	140	960	7.5×10^{-4}	0.92	0.023
Female biased genes	Functions of Hair	14	267	11	14	253	11	989	1.2×10^{-4}	2.24	0.0089
Female biased genes	Immunological System Functions	60	267	141	60	207	141	959	0.0013	0.68	0.033
Female biased genes	Protein Level	13	267	10	13	254	10	990	1.8×10^{-4}	2.26	0.0093

Table 1. Enrichment of GWAS trait categories among gene groups

This table summarizes enrichment analyses linking gene categories to human traits based on GWAS Atlas annotations (SubchapterLevel) using MAGMA gene-level association statistics. For each combination of gene category and GWAS trait, Fisher's exact test was applied to evaluate enrichment relative to a background set of 1,000 randomly sampled genes from the MAGMA dataset. The table reports, for each pair, the number of trait-associated genes within the gene category (*k*), the total number of genes in that category (*M*), the number of background genes associated with the same trait (*bg*□), and the corresponding Fisher's exact test P value, log□ enrichment ratio, and false discovery rate (FDR) adjusted by the Benjamini–Hochberg method. Only traits with FDR < 0.05 are shown. Positive log□ enrichment values indicate overrepresentation of trait-associated genes within the corresponding gene group.

Supplemental Material

Table S1. Differentially expressed genes defining glandular-enriched, mammary-specific, and female-biased expression programs

List of differentially expressed genes identified from GTEx RNA-seq analyses and classified into three expression categories: glandular-enriched genes, mammary-specific genes, and female-biased genes.

Table S2. Functional enrichment of single-cell gene clusters based on detected fraction patterns

Functional enrichment results from g:Profiler for gene clusters identified by hierarchical clustering of detected fraction profiles across mammary cell types. Enrichment is reported for Gene Ontology (BP, MF, CC), KEGG, and Reactome categories, with FDR-adjusted p-values.

Table S3. Evolutionary conservation of glandular-enriched, mammary-specific, and female-biased genes across vertebrates

Presence or absence of orthologs for genes in each expression category across representative vertebrate species. Orthology was assessed using Ensembl BioMart, and conservation is reported as binary presence across species.

Table S4. Functional enrichment of gene subsets from Mammary Module 2 shared with other glandular tissues

g:Profiler enrichment results for gene subsets defined by overlap between Mammary Module 2 and modules from other epithelial secretory tissues.

Supplementary Figures

Figure S1. Cross-tissue overlap of mammary co-expression modules.
Heatmap showing Jaccard indices between mammary gland co-expression modules and modules derived from other GTEx tissues. For each non-mammary tissue, the Jaccard index represents the maximum overlap observed between a given mammary module and any module from that tissue (“best-per-tissue” overlap). Rows correspond to comparison tissues, and columns correspond to mammary co-expression modules. Tissues are annotated as glandular or non-glandular based on tissue classification. Warmer colors indicate higher Jaccard similarity.

References

1. Medina D. The mammary gland: a unique organ for the study of development and tumorigenesis. J Mammary Gland Biol Neoplasia. 1996;1:5–19.

2. Kaur TP, Verma R, Choudhary RK. Introduction to mammary gland and its cell types. Stem Cells in Veterinary Science. Singapore: Springer Nature Singapore; 2021. p. 25–37.

3. Rauner G, Kuperwasser C. Microenvironmental control of cell fate decisions in mammary gland development and cancer. Dev Cell. 2021;56:1875–83.

4. Dawson CA, Visvader JE. The cellular organization of the mammary gland: Insights from microscopy. J Mammary Gland Biol Neoplasia. 2021;26:71–85.

5. Slepicka PF, Somasundara AVH, Dos Santos CO. The molecular basis of mammary gland development and epithelial differentiation. Semin Cell Dev Biol. 2021;114:93–112.

6. Chen X, Liu Q, Song E. Mammary stem cells: angels or demons in mammary gland? Signal Transduct Target Ther. 2017;2:16038.

7. Tharmapalan P, Mahendralingam M, Berman HK, Khokha R. Mammary stem cells and progenitors: targeting the roots of breast cancer for prevention. EMBO J. 2019;38:e100852.

8. McManaman JL, Reyland ME, Thrower EC. Secretion and fluid transport mechanisms in the mammary gland: comparisons with the exocrine pancreas and the salivary gland. J Mammary Gland Biol Neoplasia. 2006;11:249–68.

9. Khan S, Fitch S, Knox S, Arora R. Exocrine gland structure-function relationships. Development [Internet]. 2022;149. Available from: http://dx.doi.org/10.1242/dev.197657

10. Brückner A, Parker J. Molecular evolution of gland cell types and chemical interactions in animals. J Exp Biol. 2020;223:jeb211938.

11. Xiong X, Zheng L-W, Ding Y, Chen Y-F, Cai Y-W, Wang L-P, et al. Breast cancer: pathogenesis and treatments. Signal Transduct Target Ther. 2025;10:49.

12. Majood M, Rao R. Human milk: insights on cell composition, organoids and emerging applications. Pediatr Res. 2025;1–12.

13. Li CM-C, Shapiro H, Tsiobikas C, Selfors LM, Chen H, Rosenbluth J, et al. Aging-associated alterations in mammary epithelia and stroma revealed by single-cell RNA sequencing. Cell Rep. 2020;33:108566.

14. Twigger A-J, Engelbrecht LK, Bach K, Schultz-Pernice I, Pensa S, Stenning J, et al. Transcriptional changes in the mammary gland during lactation revealed by single cell sequencing of cells from human milk. Nat Commun. 2022;13:562.

15. Saitou M, Gaylord EA, Xu E, May AJ, Neznanova L, Nathan S, et al. Functional Specialization of Human Salivary Glands and Origins of Proteins Intrinsic to Human Saliva. Cell Rep. 2020;33:108402.

16. Gao X, Oei MS, Ovitt CE, Sincan M, Melvin JE. Transcriptional profiling reveals gland-specific differential expression in the three major salivary glands of the adult mouse. Physiol Genomics.

2018;50:263–71.

17. Hauser BR, Aure MH, Kelly MC, Genomics and Computational Biology Core, Hoffman MP, Chibly AM. Generation of a single-cell RNAseq atlas of Murine salivary gland development. iScience. 2020;23:101838.

18. Altrieth AL, Kenney J, Nelson DA, Suarez EG, Gellatly V, Gabunia S, et al. Single-cell transcriptomic analysis of salivary gland endothelial cells. J Dent Res. 2024;103:269–78.

19. Ku GM, Kim H, Vaughn IW, Hangauer MJ, Myung Oh C, German MS, et al. Research resource: RNA-Seq reveals unique features of the pancreatic β-cell transcriptome. Mol Endocrinol. 2012;26:1783–92.

20. Bailey P, Chang DK, Nones K, Johns AL, Patch A-M, Gingras M-C, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016;531:47–52.

21. Huang RJ, Wichmann IA, Su A, Sathe A, Shum MV, Grimes SM, et al. A spatial transcriptomic signature of 26 genes resolved at single-cell resolution characterizes high-risk gastric cancer precursors. NPJ Precis Oncol. 2025;9:52.

22. Öling S, Struck E, Noreen-Thorsen M, Zwahlen M, von Feilitzen K, Odeberg J, et al. A human stomach cell type transcriptome atlas. BMC Biol. 2024;22:36.

23. Elmentaite R, Kumasaka N, Roberts K, Fleming A, Dann E, King HW, et al. Cells of the human intestinal tract mapped across space and time. Nature. 2021;597:250–5.

24. Inoue T, Kuwano T, Uehara Y, Yano M, Oya N, Takada N, et al. Non-invasive human skin transcriptome analysis using mRNA in skin surface lipids. Commun Biol. 2022;5:215.

25. Pia-Foschini M, Reis-Filho JS, Eusebi V, Lakhani SR. Salivary gland-like tumours of the breast: surgical and molecular pathology. J Clin Pathol. 2003;56:497–506.

26. Wang S, Sekiguchi R, Daley WP, Yamada KM. Patterned cell and matrix dynamics in branching morphogenesis. J Cell Biol. 2017;216:559–70.

27. Lazarescu O, Ziv-Agam M, Haim Y, Hekselman I, Jubran J, Shneyour A, et al. Human subcutaneous and visceral adipocyte atlases uncover classical and nonclassical adipocytes and depot-specific patterns. Nat Genet. 2025;57:413–26.

28. Reed AD, Pensa S, Steif A, Stenning J, Kunz DJ, Porter LJ, et al. A single-cell atlas enables mapping of homeostatic cellular shifts in the adult human breast. Nat Genet. 2024;56:652–62.

29. CZI Cell Science Program, Abdulla S, Aevermann B, Assis P, Badajoz S, Bell SM, et al. CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. Nucleic Acids Res. 2025;53:D886–900.

30. Nickell WB, Skelton J. Breast fat and fallacies: more than 100 years of anatomical fantasy. J Hum Lact. 2005;21:126–30.

31. Ibrahim MM. Subcutaneous and visceral adipose tissue: structural and functional differences. Obes Metab. 2010;7:64–5.

32. Gong T, Szustakowski JD. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. Bioinformatics. 2013;29:1083–5.

33. Watanabe K, Stringer S, Frei O, Umićević Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet. 2019;51:1339–48.

34. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:e1004219.

35. Gluck C, Min S, Oyelakin A, Smalley K, Sinha S, Romano R-A. RNA-seq based transcriptomic map reveals new insights into mouse salivary gland development and maturation. BMC Genomics. 2016;17:923.

36. Zhang H, Gao M, Zhao W, Yu L. The chromatin architectural regulator SND1 mediates metastasis in triple-negative breast cancer by promoting CDH1 gene methylation. Breast Cancer Res. 2023;25:129.

37. Dopeso H, Gazzo AM, Derakhshan F, Brown DN, Selenica P, Jalali S, et al. Genomic and epigenomic basis of breast invasive lobular carcinomas lacking CDH1 genetic alterations. NPJ Precis Oncol. 2024;8:33.

38. Wang Z, Coban B, Wu H, Chouaref J, Daxinger L, Paulsen MT, et al. GRHL2-controlled gene expression networks in luminal breast cancer. Cell Commun Signal. 2023;21:15.

39. Kumegawa K, Takahashi Y, Saeki S, Yang L, Nakadai T, Osako T, et al. GRHL2 motif is associated with intratumor heterogeneity of cis-regulatory elements in luminal breast cancer. NPJ Breast Cancer. 2022;8:70.

40. Sun Y, Pan J, Li Y, Hu Y, Ma J, Chen F, et al. Restoring BARX2 in OSCC reverses partial EMT and suppresses metastasis through miR-186-5p/miR-378a-3p-dependent SERPINE2 inhibition. Oncogene. 2024;43:1941–54.

41. Reversi A, Cassoni P, Chini B. Oxytocin receptor signaling in myoepithelial and cancer cells. J Mammary Gland Biol Neoplasia. 2005;10:221–9.

42. Liu H, Gruber CW, Alewood PF, Möller A, Muttenthaler M. The oxytocin receptor signalling system and breast cancer: a critical review. Oncogene. 2020;39:5917–32.

43. Li D, San M, Zhang J, Yang A, Xie W, Chen Y, et al. Oxytocin receptor induces mammary tumorigenesis through prolactin/p-STAT5 pathway. Cell Death Dis. 2021;12:588.

44. Li D, Ji Y, Zhao C, Yao Y, Yang A, Jin H, et al. OXTR overexpression leads to abnormal mammary gland development in mice. J Endocrinol. 2018;239:121–36.

45. Mohammad Nezhady MA, Modaresinejad M, Zia A, Chemtob S. Versatile lactate signaling via HCAR1: a multifaceted GPCR involved in many biological processes. Am J Physiol Cell Physiol. 2023;325:C1502–15.

46. Jin L, Guo Y, Chen J, Wen Z, Jiang Y, Qian J. Lactate receptor HCAR1 regulates cell growth, metastasis and maintenance of cancer□specific energy metabolism in breast cancer cells. Mol Med Rep [Internet]. 2022;26. Available from: http://dx.doi.org/10.3892/mmr.2022.12784

47. Wagner W, Kania KD, Ciszewski WM. Stimulation of lactate receptor (HCAR1) affects cellular DNA repair capacity. DNA Repair (Amst). 2017;52:49–58.

48. Douglas NC, Papaioannou VE. The T-box transcription factors TBX2 and TBX3 in mammary gland development and breast cancer. J Mammary Gland Biol Neoplasia. 2013;18:143–7.

49. Cornelissen LM, Drenth AP, van der Burg E, de Bruijn R, Pritchard CEJ, Huijbers IJ, et al. TRPS1 acts as a context-dependent regulator of mammary epithelial cell growth/differentiation and breast cancer development. Genes Dev. 2020;34:179–93.

50. Xu Y, Qin L, Sun T, Wu H, He T, Yang Z, et al. Twist1 promotes breast cancer invasion and metastasis by silencing Foxa1 expression. Oncogene. 2017;36:1157–66.

51. Foubert E, De Craene B, Berx G. Key signalling nodes in mammary gland development and cancer. The Snail1-Twist1 conspiracy in malignant breast cancer progression. Breast Cancer Res. 2010;12:206.

52. Resseguie M, Song J, Niculescu MD, da Costa K-A, Randall TA, Zeisel SH. Phosphatidylethanolamine N-methyltransferase (PEMT) gene expression is induced by estrogen in human and mouse primary hepatocytes. FASEB J. 2007;21:2622–32.

53. Montalbetti N, Dalghi MG, Albrecht C, Hediger MA. Nutrient transport in the mammary gland: calcium, trace minerals and water soluble vitamins. J Mammary Gland Biol Neoplasia. 2014;19:73–90.

54. Hill L, Wutz G, Jaritz M, Tagoh H, Calderón L, Peters J-M, et al. Igh and Igk loci use different folding principles for V gene recombination due to distinct chromosomal architectures of pro-B and pre-B cells. Nat Commun. 2023;14:2316.

55. Lycke N, Bemark M, Spencer J. Mucosal B Cell Differentiation and Regulation. Mucosal Immunology. Elsevier; 2015. p. 701–19.

56. Sun W, Hughes EP, Kim H, Perovanovic J, Charley KR, Perkins B, et al. OCA-B/Pou2af1 is sufficient to promote CD4+ T cell memory and prospectively identifies memory precursors. Proc Natl Acad Sci U S A. 2024;121:e2309153121.

57. Jaeger-Ruckstuhl CA, Lo Y, Fulton E, Waltner OG, Shabaneh TB, Simon S, et al. Signaling via a CD27-TRAF2-SHP-1 axis during naive T cell activation promotes memory-associated gene regulatory networks. Immunity. 2024;57:287–302.e12.

58. Hasilo CP, McCudden CR, Gillespie JRJ, James KA, Hirvi ER, Zaidi D, et al. Nuclear targeting of stanniocalcin to mammary gland alveolar cells during pregnancy and lactation. Am J Physiol Endocrinol Metab. 2005;289:E634–42.

59. Rosin FM, Kramer EM. Old dogs, new tricks: regulatory evolution in conserved genetic modules leads to novel morphologies in plants. Dev Biol. 2009;332:25–35.

60. Defoort J, Van de Peer Y, Vermeirssen V. Function, dynamics and evolution of network motif modules in integrated gene regulatory networks of worm and plant. Nucleic Acids Res. 2018;46:6480–503.

61. Pires ND, Dolan L. Morphological evolution in land plants: new designs with old genes. Philos Trans R Soc Lond B Biol Sci. 2012;367:508–18.

62. Uebbing S, Kocher AA, Baumgartner M, Ji Y, Bai S, Xing X, et al. Evolutionary innovations in conserved regulatory elements associate with developmental genes in mammals. Mol Biol Evol [Internet]. 2024;41. Available from: http://dx.doi.org/10.1093/molbev/msae199

63. Hammad H, Lambrecht BN. Barrier epithelial cells and the control of type 2 immunity. Immunity. 2015;43:29–40.

64. Yang J, Yan H. Mucosal epithelial cells: the initial sentinels and responders controlling and

regulating immune responses to viral infections. Cell Mol Immunol. 2021;18:1628–30.

65. Ojeh N, Akgül B, Tomic-Canic M, Philpott M, Navsaria H. In vitro skin models to study epithelial regeneration from the hair follicle. PLoS One. 2017;12:e0174389.

66. Macias H, Hinck L. Mammary gland development. Wiley Interdiscip Rev Dev Biol. 2012;1:533–57.

67. Watanabe K, Umićević Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. Nat Commun. 2019;10:3222.