A Multi-Omics Bioinformatic Characterization of Human BTN1A1

Samantha Lane – M.S Bioinformatics, Johns Hopkins University

Abstract

BTN1A1 (Butyrophilin Subfamily 1 Member A1) is a mammary-specific, immunoglobulin-superfamily glycoprotein essential for lipid droplet secretion during lactation. Throughout a semester-long bioinformatics investigation, I performed integrated genomic, structural, evolutionary, regulatory, and transcriptomic analyses using NCBI, Ensembl, UCSC Genome Browser, BLAST/FASTA, CDD, InterPro, TMHMM, SignalP, PSIPRED, Swiss-Model, GTEx, GEO, and ENCODE. BTN1A1 displays strong mammary-specific expression, deep evolutionary conservation, multiple immunoglobulin-like extracellular domains, and a characteristic butyrophilin-family transmembrane architecture. Promoter analyses identified a TATA-less promoter with mammary-associated transcription factor motifs and ENCODE-validated enhancer/promoter chromatin marks. Comparative genomics confirmed conservation across mammals, particularly within coding regions. This work provides an integrated multi-omics profile of BTN1A1, illustrating how modern bioinformatics tools illuminate gene structure, regulation, and tissue-specific biology.

Introduction

BTN1A1 is the primary structural glycoprotein coating the milk-fat globule membrane (MFGM), essential for lipid secretion in mammary epithelial cells and additionally implicated in immune signaling (Heid et al., 1983; Mather & Jack, 1993; Kim et al., 2024). BTN1A1 belongs to the immunoglobulin superfamily and resides within the MHC class I region of human chromosome 6, suggesting evolutionary relationships with immune regulators (Vernet et al., 1993).

The objective of this study was to synthesize all BTN1A1-focused analyses from a semester-long bioinformatics course (Morrow, 2025) into a cohesive multi-omics characterization. The analyses included genomic organization and annotation; transcript isoforms, SNPs, paralogs, and orthologs; evolutionary conservation using BLAST/FASTA and multiple sequence alignment; promoter prediction and regulatory architecture; protein domain, motif, signal peptide, and transmembrane prediction; secondary and tertiary structure modeling; tissue specificity (GTEx) and disease-associated expression changes (GEO); and regulatory chromatin context via ENCODE.

Methods

Genomic and Transcript Annotation – Genomic coordinates, exon structures, and transcript models for BTN1A1 were obtained from NCBI Gene, Ensembl, and UCSC Genome Browser (NCBI Gene, 2025; Ensembl, 2025; Kent et al., 2002). Only BTN1A1's validated transcript NM_001732.3 and protein NP_001723.2 were analyzed.

Variant and Gene Family Analysis – Variant data (SNPs), alternative transcripts, paralogs, and orthologs were obtained via Entrez Gene, dbSNP, and Ensembl comparative genomics.

Sequence Conservation and Homology – BLASTN, BLASTP, and FASTA were used to assess conservation across species (Pearson, 2013; Zhang et al., 2000). Multiple sequence alignments were generated using Clustal Omega, T-Coffee, and MUSCLE hosted at EMBL-EBI.

Promoter and Regulatory Prediction – Promoter features, transcription start sites, and regulatory motifs were identified using Softberry FPROM, Promoter 2.0, and MEME Suite. Chromatin-level regulatory features were validated using ENCODE histone modifications and DNase-seq accessibility (ENCODE Project Consortium, 2012).

Protein Structure and Domain Prediction – Protein motifs, domains, transmembrane regions, and signal peptides were predicted using InterPro, CDD, TMHMM, SignalP, PSIPRED, and Swiss-Model (Uniprot Q13410).

Expression Analysis – Tissue specificity was examined using GTEx RNA-seq. Differential expression in breast tumors vs. normal tissue was analyzed using GEO dataset GSE15852 (Oh et al., 2012). ENCODE HMEC RNA-seq (ENCFF647QLX) was viewed in IGV for promoter activity.

Results

1. Genomic Organization of BTN1A1

BTN1A1 is located on chromosome 6p22.1 at genomic coordinates:
hg38: chr6:26,500,303–26,510,425

BTN1A1 contains two major validated transcript variants (e.g., NM_001732.3 and NM_013483.4), which differ in UTR composition and minor exon-edge boundaries but encode the same protein product (Ensembl, 2025; NCBI Gene, 2025). Importantly:

The exact exon boundaries differ slightly between RefSeq and GENCODE.

The gene lies within the MHC class I region, sharing synteny with immunoglobulin-superfamily genes (Vernet et al., 1993).

2. Sequence Variation and Gene Family Context

BTN1A1 displays ~3,479 SNPs (Ensembl, 2025), ~763 SNPs within the coding sequence, two mRNA isoforms, 98 orthologs (primarily mammalian), 15 paralogs, consistent with the butyrophilin gene family's expansion (Ensembl, 2025).

Coding-region SNPs included both synonymous and missense variants, while noncoding variants were distributed across promoter and intronic regions (dbSNP, 2025).

3. Evolutionary Conservation

BLAST and FASTA analyses revealed 99-100% identity to human BTN1A1 genomic sequences, 97-99% identity to murine, bovine, and ovine BTN1A1 cDNA sequences, high conservation across all mammalian coding exons, and lower conservation in UTRs and introns. Multiple sequence alignment (Clustal Omega, T-Coffee, MUSCLE) demonstrated strong global conservation of BTN1A1's coding sequence, especially within Ig-like and transmembrane domains. These results emphasize BTN1A1 as a conserved structural protein essential for lactation across mammals (Mather & Jack, 1993).

4. Promoter Architecture and Regulatory Landscape

All promoter prediction tools agreed: BTN1A1 lacks a canonical TATA box (Softberry FPROM; Promoter 2.0).

5. Predicted transcription start sites (TSS)

Promoter tools converged on multiple TSS clusters, consistent with genes regulated by hormonal or developmental cues common in the mammary gland (Promoter 2.0).

MEME and FPROM identified:

- GATA-binding motifs, associated with mammary epithelial differentiation,
- CAAT boxes in mid-promoter regions
- GC-rich regions, consistent with immune-related gene promoters
- Poly-A/T tracts likely contributing to nucleosome depletion

ENCODE mammary epithelial cell tracks showed:

- H3K27ac and H3K4me1 peaks upstream (enhancer activity)
- H3K4me3 and DNase hypersensitivity at the promoter/TSS
- RNA-seq peaks aligned precisely with BTN1A1 exons (ENCODE Project Consortium, 2012; ENCODE, 2020).

This confirms active, open chromatin and strong transcriptional initiation in mammary cells.

6. Protein Structural and Functional Features

BTN1A1 encodes a ~527 aa glycoprotein (isoform-dependent) with Ig-like Domains. InterPro and CDD identified an N-terminal IgV-like domain, and a following IgC-like domain. These domains mediate protein–protein interactions typical of immunoglobulin superfamily members (Mather & Jack, 1993).

TMHMM predicted a single C-terminal transmembrane helix, placing the N-terminal Ig-like region extracellularly and the C-terminal tail cytoplasmically. This matches the known BTN/MHC-family topology.

SignalP identified a strong N-terminal signal sequence, targeting BTN1A1 to the secretory pathway.

PSIPRED predicted:

- β-sheet–rich Ig-like extracellular domains
- α-helical transmembrane region

Swiss-Model produced a homology model consistent with established BTN-family crystal structures, including paired Ig domains and a membrane-proximal stalk.

7. Expression Profiles

GTEx – BTN1A1 is among the most breast-specific genes in the human genome, highly expressed in female breast tissue (>400 TPM), and nearly silent in all other tissues (GTEx Consortium, 2020).

GEO (GSE15852: Breast Tumor vs. Normal) – BTN1A1 was significantly downregulated in tumors (Oh et al., 2012), consistent with loss of epithelial/lactation identity and tumor dedifferentiation.

ENCODE Mammary Epithelial RNA-seq – IGV visualization (ENCFF647QLX) confirmed strong RNA-seq signal across BTN1A1 exons, clean intron depletion (polyA-mRNA), active promoter marks in mammary epithelial cells (ENCODE, 2020).

Discussion

BTN1A1 is a deeply conserved, mammary-specific glycoprotein essential for milk secretion and implicated in epithelial–immune interactions. The multi-omics evidence collected throughout the project demonstrates:

1. Strong coding-region conservation – BTN1A1 exhibits evolutionary constraints typical of essential structural proteins.

2. A tightly regulated promoter – TATA-less configuration, mammary-associated motifs, and enhancer marks support hormonally regulated transcription.

3. A domain architecture consistent with immune and epithelial function – IgV and IgC domains pair with a single membrane anchor, matching butyrophilin family biology.

4. Mammary-specific expression and tumor suppression – GTEx and GEO analyses confirm BTN1A1 as a hallmark of normal mammary epithelial differentiation.

5. The chromatin environment supports active transcription – ENCODE marks validate promoter predictions and expression data.

Together, these findings portray BTN1A1 as a canonical mammary epithelial gene, integrating secretory, immunological, and developmental roles.

References

Afrache, H., Harly, C., Cherai, M., et al. (2012). The butyrophilin family: Emerging roles in immune regulation. *Trends in Immunology.*

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, 28–36. (Foundational reference for MEME Suite)

dbSNP. (2025). *BTN1A1 variant database.* National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/snp/

DTU Health Tech. (2025). *SignalP 6.0* and *TMHMM 2.0*. https://services.healthtech.dtu.dk/

ENCODE Project. (2020). *ENCFF647QLX: HMEC RNA-seq.* https://www.encodeproject.org/

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57–74.

Ensembl. (2025). *BTN1A1 Gene Summary (ENSG00000124557)*. https://www.ensembl.org/

GTEx Consortium. (2020). The GTEx Portal: Tissue-specific gene expression. https://gtexportal.org/

Heid, H. W., Winter, S., Bruder, G., Keenan, T. W., & Jarasch, E. D. (1983). Butyrophilin, an apical plasma membrane–associated glycoprotein characteristic of lactating mammary glands. *Biochimica et Biophysica Acta*, 728(2), 228–238.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, 12(6), 996–1006.

Kim, Y. S., Lee, S. H., Park, A. H., et al. (2024). BTN1A1 is a novel immune checkpoint mutually exclusive to PD-L1. *Journal for Immunotherapy of Cancer*, 12(3), e008303.

Mather, I. H., & Jack, L. J. (1993). A review of the molecular and cellular biology of butyrophilin. *Journal of Dairy Science*, 76(12), 3832–3850.

Morrow, J. (2025). *Intro to Bioinformatics Course Lectures and Materials*. Johns Hopkins University.

NCBI CDD. (2025). *Conserved Domain Database: BTN1A1*. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/Structure/cdd/

NCBI Gene. (2025). *BTN1A1 (Gene ID: 696)*. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/gene/

Oh, D.-S., et al. (2012). Expression data from human breast tumors and matched normal tissues (GSE15852). *Gene Expression Omnibus*. https://www.ncbi.nlm.nih.gov/geo/

Pearson, W. R. (2013). Selecting the right similarity-scoring matrix for sequence comparisons. In *FASTA sequence comparison* (NCBI Book Chapter). National Center for Biotechnology Information.

Promoter 2.0 Prediction Server. (n.d.). Technical University of Denmark. http://www.cbs.dtu.dk/services/Promoter/

PSIPRED. (2025). *Protein structure prediction server.* http://bioinf.cs.ucl.ac.uk/psipred/

Softberry FPROM. (n.d.). *Promoter prediction tool.* http://www.softberry.com/

Swiss-Model. (2025). *BTN1A1 homology model (Q13410)*. https://swissmodel.expasy.org/

Taylor, M. R., Peterson, J. A., Ceriani, R. L., & Couto, J. R. (1996). Cloning and sequence analysis of human butyrophilin. *Biochimica et Biophysica Acta*, 1306(1), 1–4.

UniProt Consortium. (2025). *BTN1A1 (Q13410)*. https://www.uniprot.org/

Vernet, C., et al. (1993). A cluster of genes on human chromosome 6 encodes a putative new family of proteins related to the Ig superfamily. *Journal of Molecular Evolution*, 37(6), 600–612.

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, 7(1–2), 203–214.