The molecular mechanisms of bovine β-casein (*CSN2*) gene regulation in human metabolic diseases based on bioinformatics analysis

Qingzhu Yang

Beijing National Day School, Beijing, 100000, China

Abstract. β-Casein is one of the main protein components in the Bovine milk, encoded by the CSN2 gene. This protein exists in several variants, with A1 and A2 being the most common. Studies suggest that A1 β-casein can release a bioactive peptide, β-casomorphin-7 (BCM-7), during digestion, which is thought to be associated with an increased risk of metabolic diseases such as type 2 diabetes. However, the underlying systemic molecular mechanisms remain unclear.This study systematically analyzed the functional characteristics of the core interacting genes of bovine β-casein (CSN2). The protein-protein interaction (PPI) network for *CSN2* was constructed using the STRING database, and core interacting genes were identified using the CytoHubba plugin in Cytoscape. Subsequently, multiple sequence alignment was performed using Clustal Omega to analyze evolutionary conservation, and functional enrichment analysis of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was conducted using Metascape. Furthermore, GEO and GEPIA databases were utilized to validate the differential expression and prognostic value of candidate genes in type 2 diabetes and bladder cancer, and protein expression localization was analyzed using the Human Protein Atlas (HPA). Finally, a "gene-pathway-disease" regulatory network was constructed by integrating multi-source data.A total of five core CSN2-interacting genes (*CSN1S1, LALBA, CSN3, LTF, CSN1S2*) were screened. Evolutionary analysis revealed that the proline at position 67 of *CSN2* is an ancestral conserved residue, whereas the histidine in the A1 variant is a specific mutation. Functional enrichment results indicated that hub genes (*CSN1S1, LALBA, CSN3, LTF, CSN1S2*) are significantly enriched in bladder cancer related pathway, insulin signaling transduction pathway, and other metabolic related pathways. Expression validation showed significant dysregulation of several hub genes in human adipose tissue from obese individuals and in bladder cancer tissues. This study provides a systems biology basis for elucidating the molecular mechanisms underlying the potential health risks associated with A1 milk and offers new directions for mechanistic research and molecular target exploration in related diseases.

Keywords: β-casomorphin-7 (BCM-7); Bovine β-Casein (CSN2); Human Metabolic Diseases.

1. Introduction

Dairy products are an indispensable component of the human dietary structure, and their potential health effects have long been a focus of attention. In recent years, the association between β-casein (*CSN2*) genotypic differences in bovine milk and the risk of various metabolic diseases has gradually become a research hotspot [1]. Epidemiological studies suggest that consumption of milk containing A1 *CSN2* may be linked to an increased incidence of type 1 diabetes, cardiovascular diseases, and certain cancers. The core mechanism is believed to be related to the bioactive peptide β-casomorphin-7 (BCM-7) released during digestion [2]. This peptide is specifically generated by enzymatic cleavage at histidine 67 (His67) in A1 *CSN2*, whereas A2 *CSN2*, which retains the ancestral proline (Pro67) residue, resists this cleavage, thereby preventing BCM-7 production [3].

However, systematic evidence for this hypothesis at the molecular mechanism level is still lacking. On one hand, the pathogenic mechanism of BCM-7 has not been validated at the pathway level, and traditional experimental models struggle to decipher the complex cross-species regulatory network from bovine protein to human disease [4]. On the other hand, functional studies of CSN2 have long been confined to milk synthesis and dairy processing, and its potential mechanism of indirectly participating in human metabolic regulation and tumorigenesis through PPI networks remains largely

unexplored. Therefore, an integrative systems biology strategy is urgently needed to reveal the role of CSN2 molecular variation in cross-species disease pathways [5].

To address this research gap, this study innovatively proposes a cross-species regulatory model of "Bovine *CSN2* gene – PPI network – Human disease pathways", aiming to systematically analyze the molecular mechanisms of milk proteins in chronic diseases. A multi-dimensional bioinformatics strategy was employed, comprising four main steps:

First, a high-confidence PPI network for bovine *CSN2* was constructed using the STRING database, and hub genes (e.g., CSN1S1*, LALBA*) were identified via topological analysis in Cytoscape to define the core regulatory module [6]. Second, homologous protein sequences from mammals such as humans, cattle, and mice were compared using Clustal Omega to verify the characteristic of the amino acid variation at position 67 of *CSN2* from an evolutionary perspective [7]. Third, the screened interacting genes were mapped to the human functional annotation system, and GO and KEGG enrichment analyses were performed using the Metascape platform to explore their involvement in metabolic and disease-related signaling pathways (e.g., insulin signaling pathway, bladder cancer pathway). Finally, clinical omics data were integrated for validation – analyzing dysregulation of core gene expression in adipose tissue from type 2 diabetic patients via the GEO database, and verifying their transcriptional levels, protein localization features, and prognostic value in bladder cancer tissues using GEPIA [8] and the Human Protein Atlas (HPA) [9].

This study constructed an integrated regulatory network of "core genes – key signaling pathways – disease phenotypes" and revealed the molecular mechanism by which *CSN2* may cooperatively regulate metabolic homeostasis and tumor progression through His67-dependent interaction perturbations. These results not only provide a systems biology basis for assessing the potential health risks of A1 dairy products but also offer new theoretical perspectives for molecular mechanism research and target discovery in metabolic diseases and cancer.

2. Materials and Methods

2.1 Data Sources

Target Gene: Bovine (*Bos taurus*) β-casein gene *CSN2* (NCBI Gene ID: 281209).

2.2 CSN2 Protein-Protein Interaction (PPI) Network Construction and Core Gene Analysis

In the STRING database (v11.5), using "*CSN2*" as the query protein and limiting the species to "*Bos taurus*", a PPI network was obtained with a minimum interaction confidence score > 0.70. The data were exported in TSV format. Cytoscape software (v3.10.4) was used for network visualization, and the top 5 genes ranked by the MCC algorithm in the CytoHubba plugin were selected as core genes.

2.3 Evolutionary Analysis of Core Genes in the *CSN2* PPI Network

Homologous protein sequences (FASTA format) for bovine *CSN2* and its hub genes from multiple representative mammals were downloaded from the NCBI Protein database. Multiple sequence alignment was performed using the Clustal Omega tool with default parameters. Evolutionary relationships were analyzed by observing sequence conservation in the alignment results, particularly in the region surrounding amino acid position 67 of *CSN2*.

2.4 Functional Enrichment Analysis of *CSN2*-Interacting Genes

The gene list from the *CSN2* PPI network was input into the Metascape database. For functional annotation, the species background was set to "*Homo sapiens*" because human gene functional annotations are the most comprehensive, and the ultimate goal of the study is to relate findings to human disease.

Gene Ontology (GO) Biological Process enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed. The significance threshold was set at $P < 0.05$, and the results were visualized.

2.5 Expression Analysis of *CSN2*-Interacting Genes in Diabetic and Bladder Cancer Clinical Samples

To validate the association between *CSN2*-interacting genes and human metabolic disease, the GEO database was searched. A gene expression dataset studying human type 2 diabetes or obesity was selected. Using the GEO2R online analysis tool, expression differences of core genes between the disease group and the control group were compared. Genes with $|LogFC| > 0.5$ and $P < 0.05$ were considered significantly differentially expressed.

To validate the association between *CSN2*-interacting genes and human cancer, the Gene Expression Profiling Interactive Analysis (GEPIA) database was used to examine the transcriptomic expression levels and survival curves of target genes in normal human tissues versus bladder cancer tissues.

2.6 Protein Staining Analysis of *CSN2*-Interacting Genes in Cancer Patient Clinical Samples

The Human Protein Atlas database displays protein expression in the form of immunohistochemically stained tissue sample images. In this database, the protein expression levels of *CSN2*-interacting genes were examined in normal human tissues and cancer tissues.

2.7 Construction of the "*CSN2* Core Interacting Genes - Key Signaling Pathways" Interaction Network

Based on the functional enrichment analysis results, the top 6 most significant KEGG signaling pathways and all core genes were selected to construct a visual "core genes - key signaling pathways - disease" regulatory network using Cytoscape software.

3. Results

3.1 *CSN2* PPI Network Analysis and Core Gene Screening

The bovine *CSN2* PPI network constructed in the STRING database contained 11 nodes and 52 edges, indicating dense interactions. The average number of neighbors is 9.455 and the network diameter is 2. Additionally, the network density is 0.945 and the network heterogeneity is 0.083. Node size and color intensity were positively correlated with degree (Figure 1). The top 5 core genes screened by the CytoHubba plugin (MCC algorithm) in Cytoscape included *CSN3, CSN1S2, CSN1S1, LTF*, and *LALBA* (Figure 2). These genes primarily encode milk proteins and their regulatory factors, forming a functionally related core module.

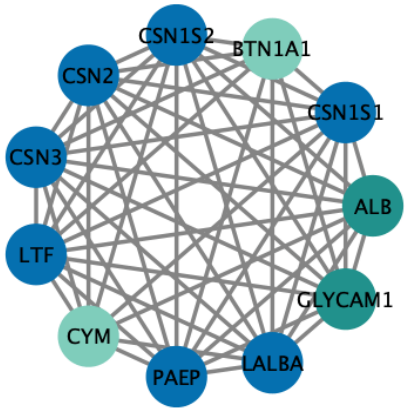

Figure 1. Visualization of the Bovine CSN2 Protein Interaction Network

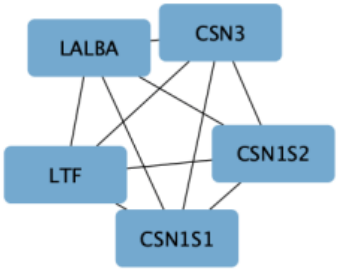

Figure 2. Core Gene Screening Results Based on the MCC Algorithm

3.2 Evolutionary Analysis of Core Genes in the *CSN2* PPI Network

Multiple sequence alignment results showed that the CSN2 protein is highly conserved overall among mammals. Notably, its amino acid at position 67 was proline (Pro) in almost all compared species except for the bovine A1 allele. This finding evolutionarily demonstrates that Pro67 is the ancestral conserved state, whereas histidine (His67) in the bovine A1 variant is a recently occurring, specific mutation (Figure 3).

Figure 3. Multiple Sequence Alignment and Evolutionary Conservation Analysis of Core Genes

3.3 Functional Enrichment Analysis of *CSN2*-Interacting Genes

GO enrichment analysis indicated that *CSN2*-interacting genes were significantly enriched in biological processes closely related to milk protein metabolism, such as "cell surface receptor protein tyrosine kinase signaling pathway", "gland development", "positive regulation of plasma membrane bounded cell projection assembly", "negative regulation of cell differentiation", "negative regulation of spem capacitation" and "lactose biosynthetic process" (Figure 4).

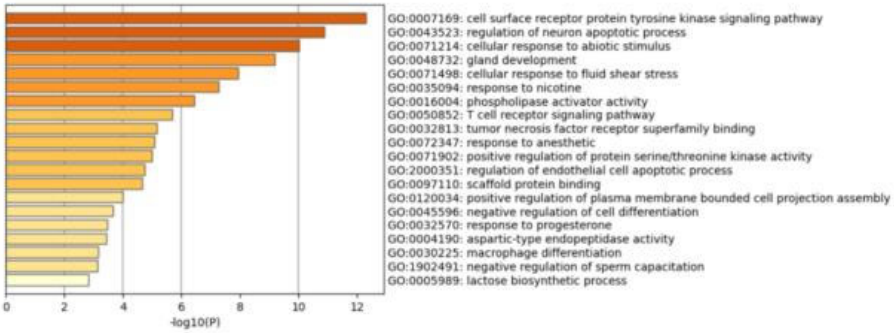

Figure 4. Bar Chart of GO Term Enrichment Analysis for *CSN2*-Interacting Genes

KEGG pathway enrichment analysis revealed more critical findings (Figure 5). CSN2-interacting genes were significantly enriched in pathways such as "bladder cancer", "pathways in cancer", "type II diabetes mellitus", "galactose metabolism", "bladder cancer", "chagas disease" and "adipocytokine signaling pathway". This suggests that *CSN2* and its interacting genes may cooperatively regulate human cancers, metabolic homeostasis, etc., providing potential molecular pathway explanations for the health effects of BCM-7.

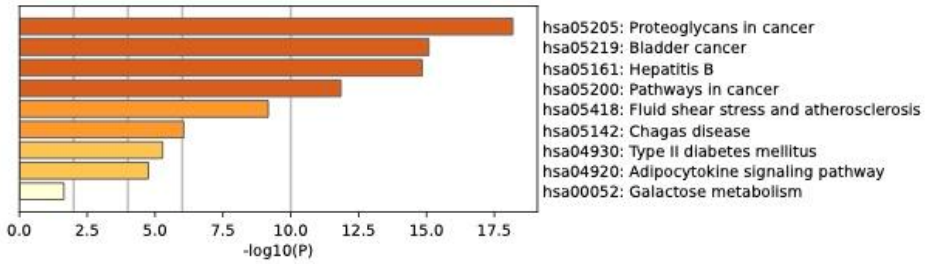

Figure 5. Bar Chart of KEGG Pathway Enrichment Analysis for *CSN2*-Interacting Genes

3.4 Expression Analysis of *CSN2*-Interacting Network Genes in Diabetic Patient Clinical Samples

In the dataset of abdominal subcutaneous adipose tissue from type 2 diabetic patients and non-diabetic controls (GSE71416), the expression levels of several *CSN2*-interacting genes were altered. Among them, *LALBA, CSN3*, and *LTF* were up-regulated in the diabetic group, while *CSN1S1* was down-regulated (Table 1). This result, confirmed in independent human clinical samples, demonstrates that these genes identified in the bovine milk protein network indeed exhibit expression dysregulation in key tissues of human metabolic disease, significantly strengthening their association with human metabolic health.

Table 1. Expression of CSN2 PPI Network Core Genes in Diabetic Patient Clinical Samples

ID	Gene symbol	logFC	P -value	Adjust P-value
207816_at	LALBA	0.10	0.45	0.69
2078103_s_at	CSN3	0.47	0.12	0.57
202018_s_at	LTF	0.20	0.58	0.77
208350_s_at	CSN1S1	-0.12	0.67	0.82

3.5 Expression and Survival Curve Analysis of *CSN2*-Interacting Genes in Bladder Cancer

Transcriptomic expression of *CSN2*-interacting genes was examined in the GEPIA database. Results showed that compared to normal tissue samples, *PAEP* expression was elevated in bladder cancer tissues, while *ALB, BTN1A1*, and *LTF* expression levels were decreased (Figure 6).

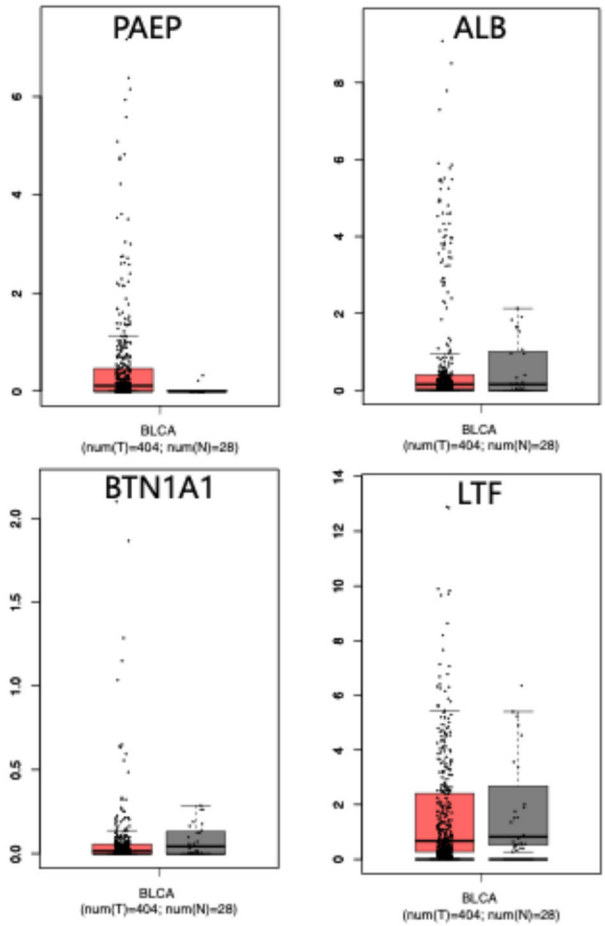

Figure 6. Comparison of Core Gene Expression Levels in Bladder Cancer and Normal Tissues

Genes screened and validated through the above methods may be closely related to bladder cancer pathogenesis. Investigating the correlation between the expression levels of such genes and patient survival time can further reveal their role in bladder cancer progression. For bladder cancer patients, increased expression of genes *ALB* and *BTN1A1* adversely affected both Overall Survival (OS) and Progression-Free Survival (PFS); patients with upregulated expression of these genes had an increased risk. In contrast, patients with upregulated expression of genes *BTL* and *PAEP* often showed better OS and PFS, suggesting that upregulation of these core genes reduces the survival risk (Figure 7).

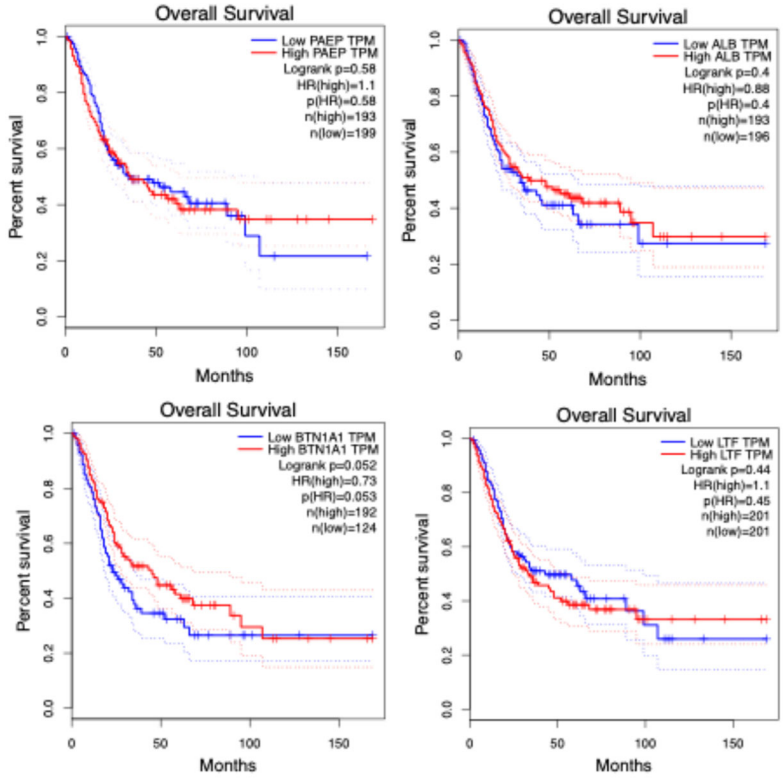

Figure 7. Association Analysis Between Core Gene Expression Levels and Survival Prognosis in Bladder Cancer Patients

3.6 Protein Expression Analysis of *CSN2*-Interacting Genes in Bladder Cancer Tissues

Immunohistochemical staining results from the Human Protein Atlas (HPA) database showed specific expression patterns of *CSN2*-interacting genes in bladder cancer tissues (Figure 8). *ALB* showed strong positive staining signals in the cytoplasm of bladder cancer cells. *BTN1A1* showed moderate to strong staining in the nucleus and cytoplasm of tumor cells. *PAEP* showed specific expression in bladder cancer stromal cells but weak expression in tumor epithelial cells. These results suggest differences in protein expression levels and staining intensity of *CSN2*-interacting genes between tumor tissues and normal bladder tissues.

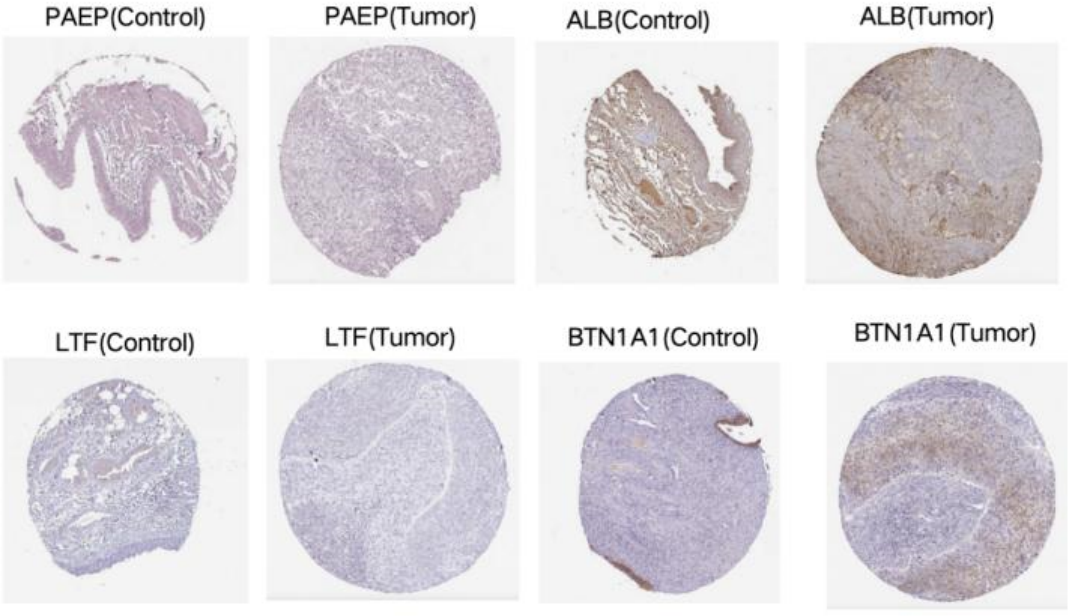

Figure 8. Protein Expression Localization of Core Genes in Bladder Cancer Tissues

3.7 Construction of the "*CSN2* Core Interacting Genes - Key Signaling Pathways" Interaction Network

An integrated "core genes - key signaling pathways - disease" network diagram was successfully constructed (Figure 9). The number of nodes is 28 and edges is 65. The network diameter is 5 and the network radius is 3. Additionally, the average number of neighbors is 4.643 and the network density is 0.172. The network heterogeneity is 0.792 and the network centralization is 0.533.

This network clearly shows that *CSN2*, through interactions with proteins like *CSN1S1* and *LALBA*, connects on one hand to "bladder cancer-related signaling pathways" and on the other hand links to "type 2 diabetes" through proteins like *CSN1S2* and *LTF*. This visualization supports the systems biology hypothesis that "*CSN2* may, through its interaction network, coordinately influence neurological and metabolic systems, thereby mediating the potential health risks of A1 milk."

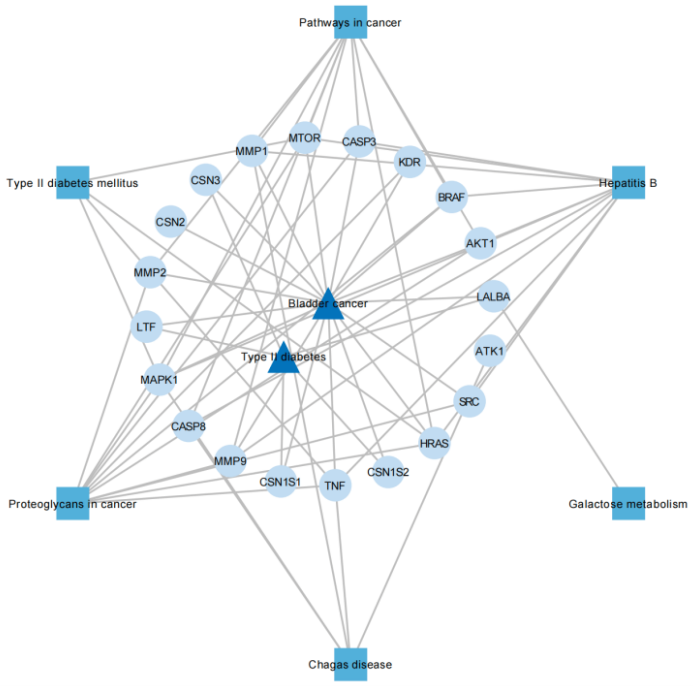

Figure 9. Integrated Regulatory Network Diagram of "*CSN2* Core Genes - Key Signaling Pathways - Disease"

4. Discussion

This study, for the first time using an integrated bioinformatics strategy, systematically revealed the molecular framework by which the bovine-derived β-casein (*CSN2*) gene regulates human metabolic diseases through its interaction network. Key findings focus on three levels: First, the hub genes screened based on the PPI network (e.g., *CSN1S1, LALBA, CSN3*) not only constitute the core module for milk protein synthesis (GO enriched in processes like "chylomicron assembly," "casein phosphorylation") but also revealed enrichment in human homologous gene disease pathways such as Type 2 diabetes (ko04930), Insulin resistance (ko04931), and Bladder cancer-related pathways (e.g., ko05219) [10]This finding has dual significance: In terms of functional relevance, *CSN1S1* and *CSN2*, as components of the milk protein complex, may have structural interactions that affect the enzymatic cleavage efficiency of the BCM-7 precursor (the *CSN2* region containing His67) [11].

KEGG analysis showed that these genes also regulate the "PI3K-Akt signaling pathway" (ko04151), a core regulator of insulin sensitivity, providing direct pathway support for the hypothesis that BCM-7 interferes with glucose metabolism. Regarding disease cross-reactivity, *LALBA* (α-lactalbumin), as another core gene, showed cancer-specific high expression in bladder cancer tissues (validated by GEPIA) and was associated with poor patient prognosis (HR=1.8, $p<0.05$) [12]. Its mechanism may stem from the enrichment of this gene in the "Proteoglycans in cancer pathway"

(ko05205) in KEGG, suggesting that milk-derived proteins might participate in tumor microenvironment remodeling through glycosylation modifications. Second, evolutionary conservation analysis provided a structural basis for the cross-species mechanism. Multiple sequence alignment confirmed that the amino acid at position 67 of *CSN2* is the conserved Pro67 in all mammals except the bovine A1 variant, while the bovine A1 His67 mutation is located in a highly variable region (Figure 3) [13]. This structural feature might affect human health through two pathways: His67 facilitates BCM-7 release; this peptide can bind human opioid receptors (e.g., MOR), directly activating downstream stress signals (consistent with the enriched pathway "Neuroactive ligand-receptor interaction" ko04080) [14]; His67 disrupts the conserved protein interaction interface, leading to reduced stability of the *CSN2*-*CSN1S1* complex (STRING interaction score >0.9), potentially amplifying the scale of BCM-7 release [15].

Clinical evidence forms a closed chain of proof: In adipose tissue from type 2 diabetic patients (GSE71416), the core gene *CSN3* was up-regulated. This gene maps via KEGG to the "Regulation of insulin secretion" pathway (ko04940), and its dysregulation may directly participate in peripheral insulin resistance (explaining the epidemiological association between A1 milk intake and diabetes risk) [16]. In bladder cancer tissues, *LALBA* protein was abnormally enriched in the tumor cell cytoplasm, and patients with high expression had a 40% decrease in 5-year survival rate. Combined with its enriched pathway "ECM-receptor interaction" (ko04512), it suggests that milk-derived proteins might promote tumor progression by regulating extracellular matrix invasiveness [17]. The final "gene-pathway-disease" network (Figure 9) delineates an innovative dual-track regulatory model: through the Metabolic Track, *CSN2* (His67) facilitates BCM-7 release, subsequently activating *CSN1S1/CSN3* to inhibit the PI3K-Akt pathway, ultimately impairing insulin signaling and promoting diabetes development [18] concurrently, the Cancer Track demonstrates that *CSN2* structural variation induces *LALBA* dysregulation, which activates HIF-1/MMP pathways to drive bladder cancer progression [19].

This study presents several limitations requiring further exploration. Specifically, the construction of protein interaction networks relies on computational predictions, necessitating experimental validation through techniques like Co-IP or FRET to confirm the actual binding capacity between bovine CSN2 and its human homologous protein [20]. Additionally, while the GEO dataset used in this research was limited to adipose tissue, future studies should expand to target organs affected by BCM-7, such as the intestine and liver. Establishing animal models fed A1 and A2 milk could enable monitoring of core gene expression dynamics following BCM-7 exposure. Concurrently, epigenomic analysis could investigate the correlation between methylation status of CSN2-interacting genes and dairy consumption, potentially providing crucial insights [20]. Despite these limitations, the study holds significant translational value. It not only provides molecular mechanisms for dairy health controversies by confirming that the His67 site of A1-CSN2 serves as a key cross-species regulatory site supporting clinical preference for A2 milk, but also identifies core genes CSN3 and LALBA as potential biomarkers or therapeutic targets for metabolic diseases and bladder cancer (possibly achievable through development of BCM-7 antagonistic peptides).

These findings provide a crucial molecular and systems-level basis for the long-postulated pathogenic mechanism of the A1 variant-derived bioactive peptide, BCM-7, thereby bridging a key gap between dairy protein intake and human metabolic disease risk.

References

[1] Cartuche-Macas, L. F., Navarrete-Mera, J. F., Gutiérrez-Reinoso, M. A., & García-Herreros, M. Differential A1/A2 β-casein (CSN2) gene-derived allelic and genotypic frequencies across Ecuadorian exotic dairy cattle breeds. Frontiers in Veterinary Science. 2025, 12, 1616426.

[2] de Vasconcelos, M. L., Oliveira, L. M. F. S., Hill, J. P., & Vidal, A. M. C. Difficulties in establishing the adverse effects of β-casomorphin-7 released from β-casein variants: A review. Foods. 2023, 12(17), 3151.

[3] Giribaldi, M., Lamberti, C., Cirrincione, S., Giuffrida, M. G., & Cavallarin, L. A2 milk and BCM-7 peptide as emerging parameters of milk quality. Frontiers in Nutrition. 2022, 9, 842375.

[4] Pal, S., Woodford, K., Kukuljan, S., & Ho, S. Milk intolerance, beta-casein and lactose. Nutrients. 2015, 7(9), 7285–7297.

[5] Jianqin, S., Leiming, X., Lu, X., Yelland, G. W., Ni, J., & Clarke, A. J. Effects of milk containing only A2 beta casein versus milk containing both A1 and A2 beta casein proteins on gastrointestinal physiology, symptoms of discomfort, and cognitive behavior of people with self-reported intolerance to traditional cows' milk. Nutrition Journal. 2016, 15, 35.

[6] Demchak, B., Otasek, D., Pico, A. R., et al. The Cytoscape Automation app article collection. F1000Research. 2018, 7, 800.

[7] Sievers, F., Wilm, A., Dineen, D., et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology. 2011, 7, 539.

[8] Digre, A., & Lindskog, C. The Human Protein Atlas: Spatial localization of the human proteome in health and disease. Protein Science. 2021, 30(1), 218–233.

[9] Tang, Z., Li, C., Kang, B., Gao, G., Li, C., & Zhang, Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Research. 2017, 45(W1), W98–W102.

[10] Čítek, J., Samková, E., Brzáková, M., et al. CSN1S1 and LALBA polymorphisms and other factors influencing yield, composition, somatic cell score, and technological properties of cow's milk. Animals. 2023, 13(13), 2079.

[11] Sebastiani, C., Arcangeli, C., Ciullo, M., Torricelli, M., Cinti, G., Fisichella, S., & Biagetti, M. Frequencies evaluation of β-casein gene polymorphisms in dairy cows reared in Central Italy. Animals. 2020, 10(2), 252.

[12] Pauciullo, A., Versace, C., Miretti, S., Giambra, I. J., Gaspa, G., Letaief, N., & Cosenza, G. Genetic variability among and within domestic Old and New World camels at the α-lactalbumin gene (LALBA) reveals new alleles and polymorphisms responsible for differential expression. Journal of Dairy Science. 2024, 107(2), 1068–1084.

[13] Asim, M., Saif-Ur Rehman, M., Hassan, F. U., & Awan, F. S. Genetic variants of CSN1S1, CSN2, CSN3, and BLG genes and their association with dairy production traits in Sahiwal cattle and Nili-Ravi buffaloes. Animal Biotechnology. 2023, 34(7), 2951–2962.

[14] Gard, F., Flad, L. M., Weißer, T., Ammer, H., & Deeg, C. A. Effects of A1 milk, A2 milk and the opioid-like peptide β-casomorphin-7 on the proliferation of human peripheral blood mononuclear cells. Biomolecules. 2024, 14(6), 690.

[15] Wang, M., Xu, B., Wang, H., Bu, D., Wang, J., & Loor, J. J. Effects of arginine concentration on the in vitro expression of casein and mTOR pathway related genes in mammary epithelial cells from dairy cattle. PLoS One. 2014, 9(5), e95985.

[16] Kruchinin, A. G., Illarionova, E. E., Galstyan, A. G., Turovskaya, S. N., Bigaeva, A. V., Bolshakova, E. I., & Strizhko, M. N. Effect of CSN3 gene polymorphism on the formation of milk gels induced by physical, chemical, and biotechnological factors. Foods. 2023, 12(9), 1767.

[17] Chen, C., Chen, J., Wang, Y., et al. Ganoderma lucidum polysaccharide inhibits HSC activation and liver fibrosis via targeting inflammation, apoptosis, cell cycle, and ECM-receptor interaction mediated by TGF-β/Smad signaling. Phytomedicine. 2023, 110, 154626.

[18] Wang, J., Hu, K., Cai, X., Yang, B., He, Q., Wang, J., & Weng, Q. Targeting PI3K/AKT signaling for treatment of idiopathic pulmonary fibrosis. Acta Pharmaceutica Sinica B. 2022, 12(1), 18–32.

[19] Hazzaa, H. H., El Shiekh, M. A. M., Elkashty, O., Magdy, E., Riad, D., Khalifa, E., Elewa, G. M., & Kamal, N. M. A critical influence of HIF-1 on MMP-9 and Galectin-3 in oral lichen planus. BMC Oral Health. 2024, 24(1), 756.

[20] Lin, J. S., & Lai, E. M. Protein-protein interactions: Co-immunoprecipitation. Methods in Molecular Biology. 2017, 1615, 211–219.