IFSCC 35TH
CONGRESS
FRANCE_CANNES 15-18 SEPT 25
THE FUTURE IS SCIENCE
ifscc2025.com
IFSCC
SOCIÉTÉ FRANÇAISE DE COSMÉTOLOGIE

IFSCC 2025 full paper (IFSCC2025-349)

Comprehensive Mechanistic Understanding and Ingredient Screening for Skin Collagen via Biological Knowledge Graph and Artificial Intelligence

Shanshan Zang [1], Shaohua Wang [2], Sonia Bouayadi [3], Jie Qiu [1], Yao Gu [1], Nan Li [1], Liang Zhang [1], Wanze Wang [1], Xinxin Chen [1], Yuan Ma [1], Amiao Pan [2], Lun Yu [2], Meijie Wang [2], Heng Luo [2,]*

[1] L'Oreal (China) Research and Innovation Center, Shanghai, China;
[2] MetaNovas Biotech Inc., San Francisco, CA, USA;
[3] L'Oréal Advanced Research, Aulnay-sous-bois, France

Corresponding author: Heng Luo (hengluo@metanovas.com)

Abstract

The rapid growth of biomedical data has made it increasingly challenging to integrate and extract useful information for therapeutic development. Type III collagen, a key component of the extracellular matrix in the dermis, plays a crucial role in skin elasticity. As we age, the content of type III collagen decreases, leading to wrinkles and signs of skin aging. Therefore, identifying substances that can boost type III collagen levels holds significant practical value in the cosmetic industry. This study leverages artificial intelligence, specifically natural language processing (NLP) and knowledge graph (KG), to discover potential ingredients that may regulate type III collagen levels from a vast body of biomedical literature. We began by systematically analyzing genes associated with type III collagen regulation, and then identified ingredients that may influence its levels. These ingredients were cross-referenced with the Inventory of Existing Cosmetic Ingredients in China (IECIC) to select those approved for cosmetic use, and were then ranked according to specific criteria. Finally, we conducted wet-lab experiments to validate the selected ingredients. The results demonstrated that these ingredients significantly increased type III collagen content in human dermal fibroblast cells. This study not only identified promising ingredients for enhancing type III collagen but also introduced an innovative approach using artificial intelligence to assist in the development of skincare products.

1. Introduction

The skin is the largest organ in the human body and acts as the first line of defense. Collagen fibers, the most abundant component of the extracellular matrix (ECM), make up about 75% of the skin's dry weight and are extremely important for skin structure and function [1]. Collagen, a member of the fibrous protein family composed of three polypeptide chains forming a triple helix, provides the skin with appropriate density and elasticity [2]. The collagen family consists of 28 members, with type I collagen (COL1) and type III collagen (COL3) being the most prevalent in the skin, making up approximately 80–85% and 8–11% of the total collagen, respectively [1].

Although type III collagen is less abundant than type I collagen in the skin, it plays a crucial role. Structurally, type I collagen forms coarse fibers and is a rigid protein that creates a framework with high tensile strength for the skin. In contrast, type III collagen forms finer fibers with greater flexibility and elasticity, creating an elastic network that is critical for the skin's stretch and recoil abilities [3]. Functionally, type III collagen regulates the fibrillogenesis of type I collagen and co-assembles with it. This interaction between different collagen types helps form the skin's overall structure and determines its biomechanical properties [4].

As skin ages, collagen synthesis decreases and degradation increases. This effect is especially pronounced for type III collagen, which is broken down more rapidly than type I collagen. This imbalance contributes to the loss of skin elasticity and the formation of wrinkles [5]. Many skincare products on the market contain type III collagen or its hydrolyzed peptides to improve skin condition [6]. Therefore, promoting the skin's own type III collagen synthesis or preventing its degradation has become an important research focus, with many products targeting these mechanisms [7,8].

The rapid growth of biomedical data has made it increasingly challenging to integrate and extract meaningful information. Natural Language Processing (NLP), an important text processing tool designed to analyze language using computational algorithms, has various capabilities, including syntax and semantic analysis, translation, summarization, entity recognition, sentiment analysis, and question answering [9]. Knowledge Graph (KG) is an effective tool for analyzing, integrating, and extracting meaningful information from large datasets. By providing heterogeneous information about various entities (such as proteins, targets, and drugs) and their inter-relationships (e.g., drug-target interactions), KG helps to deepen our understanding of complex biological systems [10]. In a knowledge graph, entities are represented as nodes, and relationships between entities are represented as edges connecting these nodes [11]. NLP and KG have broad applications in biomedicine research, including studies on skin-related conditions such as atopic dermatitis, acne, and skin aging [12].

In this study, we combined NLP and KG technologies to comprehensively review and integrate genes that regulate type III collagen. We then identified ingredients that could promote type III collagen synthesis or inhibit its degradation. Six out of seven ingredients were selected and successfully validated through cellular experiments, highlighting the potential of our artificial intelligence (AI) model in identifying cosmetic ingredients. This study provides an innovative approach using AI (NLP and KG) to assist in the development of skincare products.

2. Materials and Methods

2.1. AI-based Literature Mining and Analysis

We collected literature from 32 million PubMed abstracts and 5 million open-access full-text articles from PubMed Central (PMC). Then we identified entities as “Chemical”, “Herb”, “Gene”, “Biological Process”, “Molecular Function” and “Cellular Component” using named entity recognition (NER) model. The next step involved extracting the relationships between entities such as “down-regulate”, “affect” and “part of”. For NLP-based analysis, sentences from PubMed abstracts and PMC full-text articles were parsed, and entities co-occurring in the same sentence were identified; thus, any targets or compounds appearing in the same sentence as “type III collagen” or “COL3A1” were collected for analysis. We assigned NLP-based association weights based on two factors: (1) sentence type, with higher weights for titles and lower weights for abstracts, and (2) the contextual relevance as assessed by the language model.

2.2. Knowledge Graph Construction and Analysis

As reported at the 2023 IFSCC Congress, we constructed a biomedical knowledge graph (KG) for skin care product development [13]. Briefly, we integrated data from literature mining and multiple biomedical databases (e.g., STRING, STITCH, PharmGKB, and others). Entities such as “Chemical”, “Herb”, “Gene”, “Biological Process”, “Molecular Function” and “Cellular Component” mentioned above were identified and uniformly mapped through entity normalization models to reconcile terminology variations across different databases. We incorporated the identified entities and their relationships into a Neo4j database as nodes and edges, respectively, forming a directed knowledge graph. The knowledge graph comprised over 3 million nodes (entities) and more than 116 million edges (relationships). After the construction of Knowledge Graph, we queried it for targets and ingredients associated with “type III collagen” or “COL3A1” for analysis.

2.3. Cell Culture and Detection of Type III Collagen

BJ human dermal fibroblasts (from the National Collection of Authenticated Cell Cultures) were maintained in DMEM (Gibco) with 10% fetal bovine serum (FBS, Gibco) and penicillin/streptomycin. For assays, 10,000 cells per well were seeded in 96-well plates in serum-free DMEM and incubated overnight at 37°C, 5% CO_2. After 24 hours, test compounds (obtained from MedChemExpress) were added to the wells; 20 ng/mL TGF-β1 (GenScript) served as a positive control and the corresponding vehicle (DMSO or water) served as the negative control. After 48 hours of treatment, the supernatants were collected and type III collagen levels were measured using an ELISA kit (JonlnBio). Following supernatant collection, cell viability in each well was assessed using a CCK-8 assay (Tansoole).

3. Results

3.1. Identification of Type III Collagen-Regulating Genes

To identify genes associated with the regulation of type III collagen, we employed NLP analysis of articles from PubMed. This approach identifies genes that frequently associate with type III collagen in the same context, generating a broad list of potential candidate genes. Through this analysis, we identified 947 genes that associate with type III collagen in the literature; 249 of these genes appeared in at least three different articles and were considered more reliably linked with type III collagen (Group A genes, Figure 1A).

We further utilized our knowledge graph, an integration of multiple databases and sources [13], to identify potential gene interactions related to type III collagen. Given the extensive data available in the STRING database, we analyzed it separately from other databases. Using the knowledge graph, we identified 116 genes interacting with type III collagen from data sources other than STRING (Group B genes, Figure 1A).

For the STRING database, we applied two selection criteria to the STRING-derived interactions: first, a confidence score >500, which yielded 139 genes (Group C genes); second, inclusion of only experimentally supported interactions, yielding 428 genes (Group D genes, Figure 1A).

By merging the results from these four groups (Group A, B, C and D), we compiled a list of type III collagen-regulating genes and scored each gene based on the number of supporting pieces of evidence from each data source.

$$Score = \sqrt{N_A} + 2N_B + 0.5N_C + N_D$$

Here, N_A is the number of NLP articles, N_B is the number of KG sources other than STRING, N_C is the number of high-confidence STRING sources, and N_D is the number of STRING experimental sources.

We identified 38 core genes (*Score* > 10 and present in ≥ 3 groups) and 102 important genes (*Score* > 5). Together, these comprised 140 unique type III collagen-regulating genes (Figure 1A).

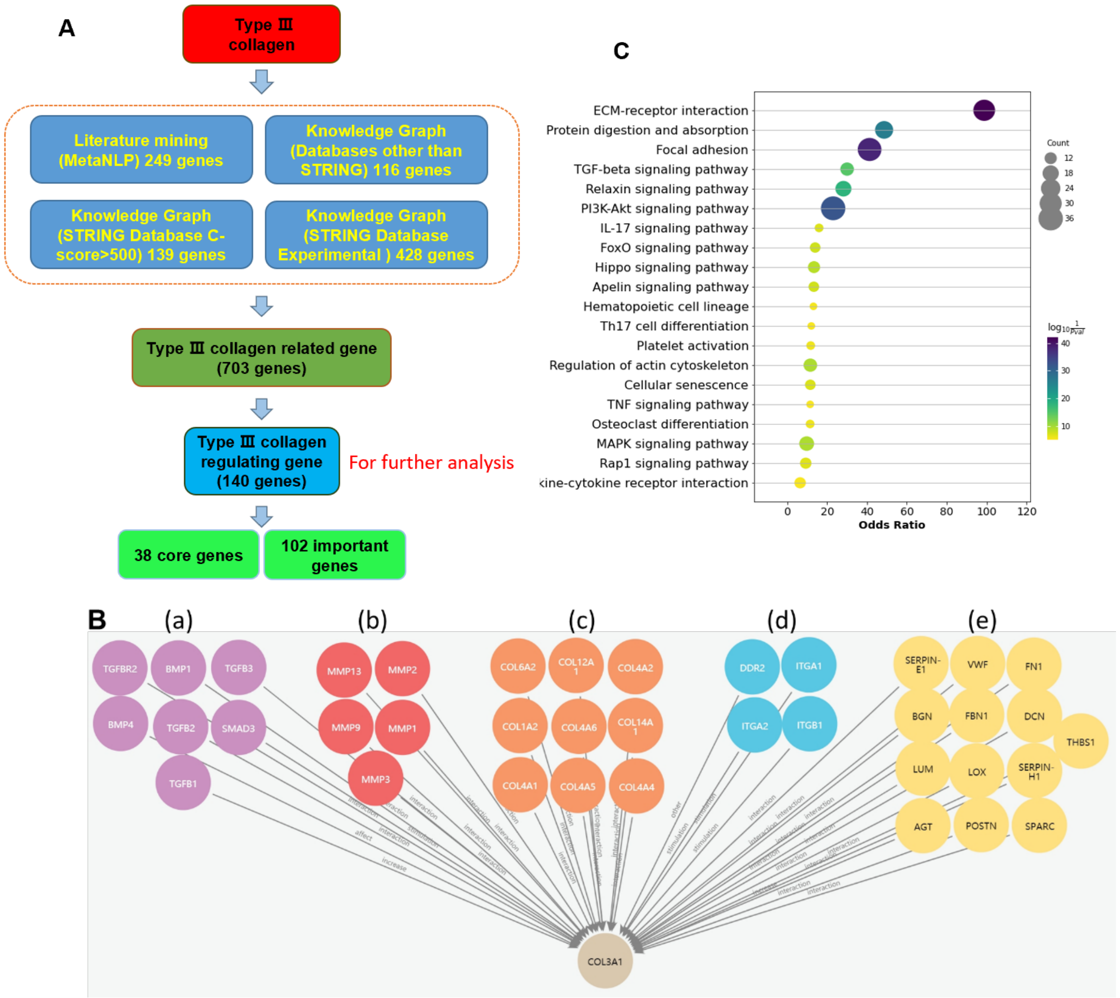

Figure 1. Identification and clustering analysis of type III collagen-regulating genes. (A) Illustration of the process of identifying and clustering genes related to type III collagen regulation. (B) Illustration of the 38 core genes and their interactions towards type III collagen. (C) The KEGG enrichment analysis of type III collagen regulating genes.

We further analyzed the pathways and functions of these 38 core genes. These core genes were involved in several key pathways: (a) TGF-β-related pathways (7 genes); (b) matrix metalloproteinases (MMPs, 5 genes); (c) collagen types (9 genes); (d) downstream collagen signaling (4 genes); and (e) genes that interact with or modify collagen (13 genes), affecting collagen assembly (Figure 1B). The core gene analysis revealed that we identified key regulatory genes for type III collagen expression. This comprehensive coverage of type III collagen's lifecycle, from expression to degradation, demonstrates the broad regulatory scope of the genes identified by our AI methods.

Further clustering analysis of the 140 type III collagen-regulating genes revealed that multiple pathways associated with collagen synthesis, degradation, and skin anti-aging were significantly enriched, such as TGF-beta signaling pathway, MAPK signaling pathway and cellular senescence (Figure 1C), which have been implicated in skin aging [14][15]. This suggests that these identified genes play an important role in collagen regulation and could serve as potential targets for developing anti-wrinkle and anti-aging products.

3.2. Identification of Ingredients Regulating Type III Collagen

Next, we identified ingredients that might regulate type III collagen levels. We first employed the NLP method to find ingredients associated with any of the 140 type III collagen-regulating genes in PubMed articles. For each ingredient, we summed the number of distinct articles that linked it to any of these genes. We also queried the knowledge graph for ingredients connected to these genes, and aggregated the number of supporting sources for each ingredient across all genes. Based on the number of articles and sources, we scored each ingredient.

In total, we identified 39,359 ingredients or candidate compounds, though most of these were not approved for use in cosmetics. By comparing these ingredients with China's Inventory of Existing Cosmetic Ingredients (IECIC, 2021), we identified 1,039 entries that were either directly approved or had approved analogues, derivatives, or botanical sources. Since the effects of these ingredients on type III collagen could be positive, negative or unclear, we manually reviewed the top-ranked ingredients to select those that positively regulate type III collagen. This led to a final reviewed list of 91 candidate ingredients potentially capable of increasing type III collagen content to combat skin aging (Figure 2A).

Among the 91 ingredients, we found several commonly used cosmetic ingredients, such as the retinoid alitretinoin and ascorbic acid (vitamin C), as well as less-common cosmetic ingredients like sulforaphane and chrysin (Table 1). Figure 2B shows that these ingredients have strong correlations with the 38 core genes for type III collagen regulation. One ingredient of particular interest is alpha-ketoglutarate (AKG). It has attracted attention in the market in recent years, and we previously discussed its potential at the 2023 IFSCC Congress [13]. AKG has reported roles in collagen synthesis, antioxidation, skin hydration, inflammation reduction, and cellular energy production [16]. As shown in Figure 2C and 2D, we identified the subset of type III collagen-regulating genes associated with AKG. KEGG pathway analysis of these genes revealed enrichment related to skin anti-aging. The AKG case study suggests that the ingredients identified by our approach can indeed influence collagen regulation and potentially alleviate skin aging.

Table 1. Representative ingredients that promote type III collagen. This table presents the scores and rankings of the 5 representative compounds with promoting type III collagen function. The listed literature PubMed IDs (PMIDs) provide evidence of these compounds' collagen-promoting and anti-aging functions.

Rank	**Ingredient**	**In China's IECIC or approved alternative**	**Score**	**Promote collagen?**	**Anti-aging?**	**Literature evidence (PMID)**
4	Alitretinoin	Retinal	47.9	Yes	Yes	16144296; 15733037
6	Ascorbic acid	Yes	39.2	Yes	Yes	37128827; 29763052

20	Sul-foraphane	Brassica Oleracea Italica	23.4	Yes	Yes	35700067;
						34271100;
						32659677

52	Chrysin	Yes	14.4	Yes	Yes	27226145;
84	α-Ketoglu-taric acid	Yes	8.2	Yes	Yes	29019707 ; 33340716

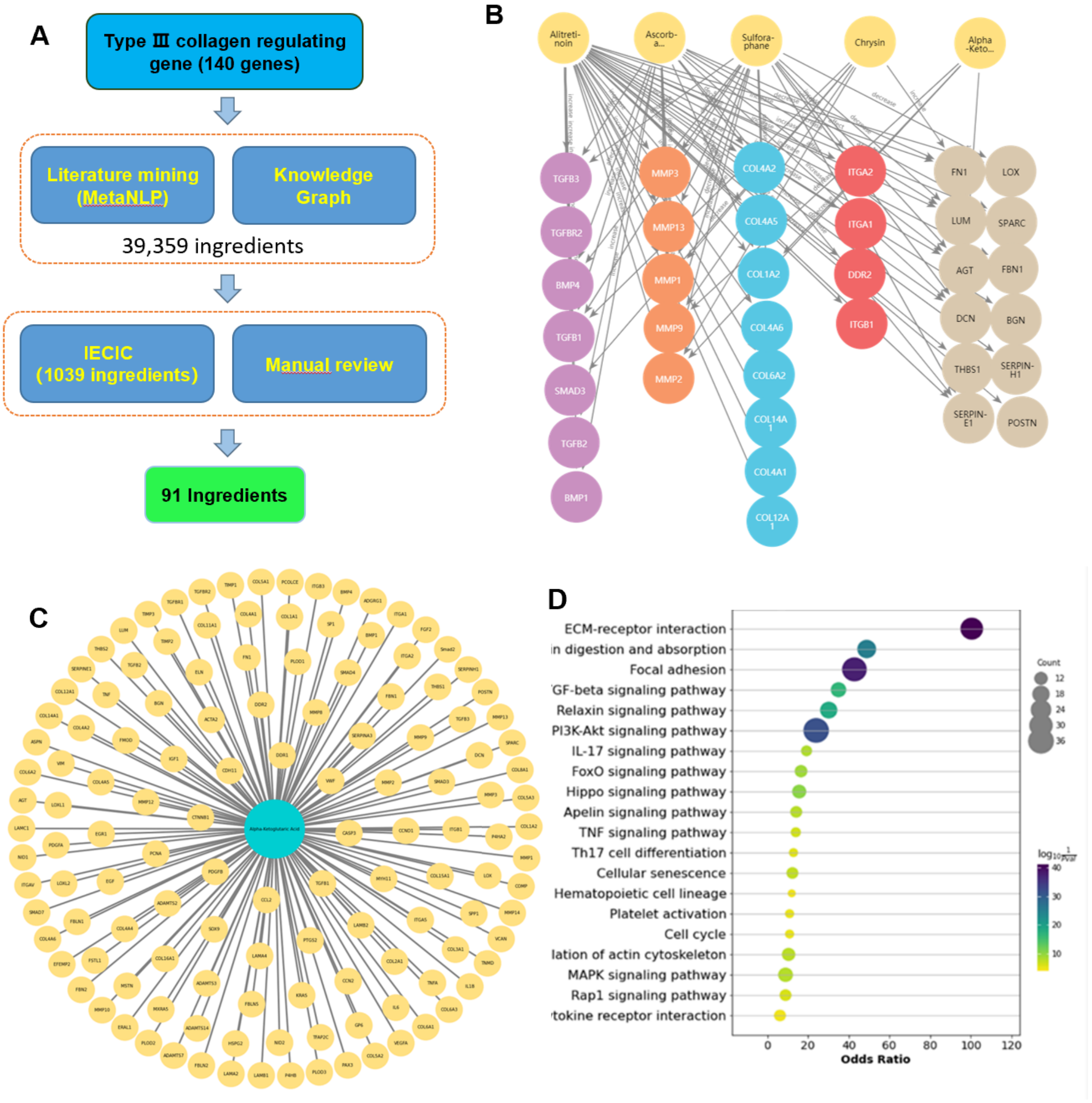

Figure 2. Representative compounds associated with type III collagen regulating genes. (A) Illustration of the process of identifying ingredients related to type III collagen regulation. (B) Representative five compounds association with type III collagen regulating core genes. (C) Alpha-ketoglutarate (AKG) association with type III collagen regulating genes. (D) The KEGG enrichment analysis of AKG associated type III collagen regulating genes.

3.3. Experimental Validation of the Ingredient Effects on Type III Collagen

To validate the effectiveness of the selected ingredients, we performed experimental validation using the BJ cells, a type of human dermal fibroblasts (HDFs). From the 91 ingredients, we selected seven ingredients with mid- to low-range rankings for further experimental testing (to test less obvious candidates rather than only the top hits). The scores and rankings of these seven ingredients are shown in Table 2. Among these, four ingredients – rutin, ellagic acid, αketoglutarate, and madecassoside – have been reported to increase collagen content in human dermal fibroblasts (HDFs). Two others, baicalein and honokiol, have shown collagen-promoting effects in other biological systems. The remaining ingredient, hesperidin methylchalcone, has no prior reports regarding its effect on collagen (Table 2). Our knowledge graph shows these ingredients modulate type III collagen regulating genes and alleviate skin aging through multiple pathways (Figure 3A).

Experimental results showed that six of the seven ingredients (all except ellagic acid) significantly increased type III collagen levels in BJ fibroblast cells, without causing significant cytotoxicity (Figure 3B and 3C). These findings demonstrate the effectiveness of our combined NLP-KG approach in identifying true positive collagen enhancers. Interestingly, hesperidin methylchalcone (HMC) is a flavonoid that has known anti-inflammatory and antioxidant properties, but it has never been reported to influence collagen synthesis. HMC, a derivative of hesperidin (a citrus bioflavonoid), can activate Nrf2 signaling – protecting cells from oxidative stress – and can inhibit cytokine production and NF-κB activation, thereby reducing inflammation [17]. Our results underscore the ability of our combined NLP-KG approach to discover novel molecules for type III collagen regulation, offering new potential ingredients for anti-wrinkle skincare products.

Table 2. Scores and rankings of the 7 selected ingredients for experimental validation

Rank	Ingredient	Score	Relationship with collagen	Literature evidence (PMID)
21	Rutin	23.1	Rutin increases the expression of collagen in human fibroblasts.	27220601
28	Honokiol	20.2	Honokiol promotes collagen in other systems, such as osteoblast MC3T3-E1 cells	21621646
35	Ellagic acid	18.4	Ellagic acid prevents collagen degradation by blocking MMP in UVB exposed HDFs	20113347
46	Baicalein	16.8	Baicalein promotes collagen in other systems, such as tendon-derived stem cells	29693006
54	Madecassoside	13.7	Madecassoside can induce collagen synthesis in HDFs	16557473
76	Hesperidin methylchalcone	10.4	No literature reported	Not available
84	α-Ketoglutaric acid	8.2	Alpha-ketoglutarate stimulates procollagen production in HDFs	17666792

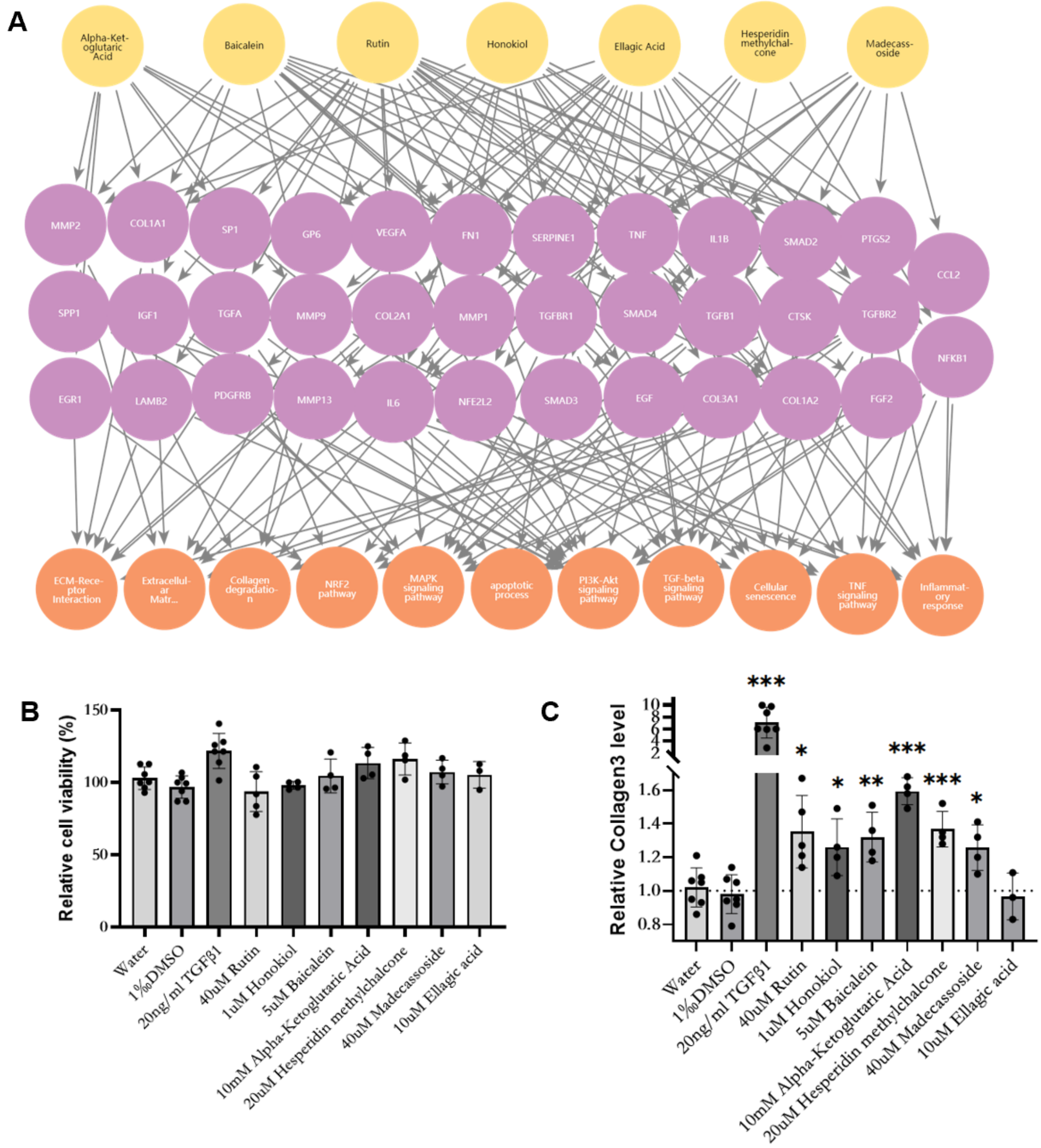

Figure 3. The selected ingredients increased the type III collagen level in human dermal fibroblasts (HDFs). (A) The knowledge graph shows selected ingredients modulate type III collagen regulating genes and alleviate skin aging through multiple pathways. (B) None of the ingredients significantly affected cell viability (CCK-8 assay). (C) Six out of seven ingredients significantly increased levels of type III collagen in cell medium after 48-hour treatment in HDFs, except for ellagic acid.

4. Discussion

In this study, we employed natural language processing (NLP) and knowledge graph (KG) technologies to identify potential ingredients for enhancing type III collagen content in the skin. After mapping 140 genes involved in type III collagen synthesis, modification, and degradation (including key pathways like TGF-β signaling and collagen assembly), we identified 91 ingredients approved for cosmetic use that could influence type III collagen levels. These included known ingredients (e.g., ascorbic acid and alitretinoin) as well as novel candidates (e.g., sulforaphane). Experimental validation in human dermal fibroblasts (HDFs) showed that six out of seven tested ingredients significantly increased type III collagen levels, with hesperidin methyl chalcone (HMC) emerging as a new potential collagen booster.

The use of AI and data-driven methods in identifying bioactive ingredients is a promising development in this field, but it also poses some challenges. AI models in data mining can be prone to overfitting, meaning they might perform well on known data but less so on truly novel data. In this study, the reliance on existing data sources might have skewed predictions toward ingredients that were already well-known in the literature. Future research should address this by integrating larger and more diverse datasets, including under-reported ingredients or proprietary datasets. Moreover, AI tools should be continuously refined to improve their performance in identifying novel bioactive ingredients. Future studies should also explore the molecular mechanisms by which these identified ingredients promote collagen synthesis and contribute to skin rejuvenation. Additionally, this work serves as a preliminary demonstration of how NLP-KG integration can retrieve known knowledge and uncover novel insights. While our current implementation employs AI models primarily for partial NLP and knowledge integration tasks, more sophisticated frameworks, such as advanced knowledge inference models and scoring/ranking algorithms, can be leveraged for deeper and more innovative analysis, depending on the specific research needs.

5. Conclusion

This study systematically identified genes and ingredients that regulate type III collagen levels, leveraging natural language processing (NLP) and knowledge graph (KG) techniques. Our findings highlighted several compounds with skin anti-aging potential and suggested new directions for modulating type III collagen. Experimental validation further supports the efficacy of these ingredients in promoting type III collagen synthesis, providing valuable insights for the development of anti-aging skincare products. In summary, our research identifies new cosmetic ingredients capable of enhancing type III collagen content. It also offers a strategy for accelerating cosmetic product development using artificial intelligence.

6. Acknowledgments

The authors would like to acknowledge the diligent proofreading of the manuscript provided by Dr. Zaihua Hou at MetaNovas Biotech Inc.

7. Conflict of Interest Statement

The authors declare that they have no conflict of interest.

8. References

[1] C.J. Zhou, Y. Guo, Mini review on collagens in normal skin and pathological scars: current understanding and future perspective, Front. Med. 11 (2024).

[2] L. Alcaide-Ruggiero, V. Molina-Hernández, M.M. Granados, J.M. Domínguez, Main and Minor Types of Collagens in the Articular Cartilage: The Role of Collagens in Repair Tissue Evaluation in Chondral Defects, Int J Mol Sci 22 (2021) 13329.

[3] H. Kuivaniemi, G. Tromp, Type III collagen (COL3A1): Gene and protein structure, tissue distribution, and associated diseases, Gene 707 (2019) 151–171.

[4] C. Wang, B.K. Brisson, M. Terajima, Q. Li, K. Hoxha, B. Han, A.M. Goldberg, X. Sherry Liu, M.S. Marcolongo, M. Enomoto-Iwamoto, M. Yamauchi, S.W. Volk, L. Han, Type III collagen is a key regulator of the collagen fibrillar structure and biomechanics of articular cartilage and meniscus, Matrix Biol 85–86 (2020) 47–67.

[5] P. Panwar, G. Lamour, N.C.W. Mackenzie, H. Yang, F. Ko, H. Li, D. Brömme, Changes in Structural-Mechanical Properties and Degradability of Collagen during Aging-associated Modifications, J Biol Chem 290 (2015) 23291–23306.

[6] B. Jadach, Z. Mielcarek, T. Osmałek, Use of Collagen in Cosmetic Products, Curr Issues Mol Biol 46 (2024) 2043–2070.

[7] R. Sriram, V. Gopal, Aging Skin and Natural Bioactives that Impede Cutaneous Aging: A Narrative Review, Indian J Dermatol 68 (2023) 414–424.

[8] Z.A.M. Yasin, F. Ibrahim, N.N. Rashid, M.F.M. Razif, R. Yusof, The Importance of Some Plant Extracts as Skin Anti-aging Resources: A Review, Curr Pharm Biotechnol 18 (2017) 864–876.

[9] P. Ernst, A. Siu, G. Weikum, KnowLife: a versatile approach for constructing a large knowledge graph for biomedical sciences, BMC Bioinformatics 16 (2015) 157.

[10] S. Ji, S. Pan, E. Cambria, P. Marttinen, P.S. Yu, A Survey on Knowledge Graphs: Representation, Acquisition, and Applications, IEEE Transactions on Neural Networks and Learning Systems 33 (2022) 494–514.

[11] S.K. Mohamed, A. Nounu, V. Nováček, Biological applications of knowledge graph embedding models, Briefings in Bioinformatics 22 (2021) 1679–1693.

[12] A. Paganelli, M. Spadafora, C. Navarrete-Dechent, S. Guida, G. Pellacani, C. Longo, Natural language processing in dermatology: A systematic literature review and state of the art, Journal of the European Academy of Dermatology and Venereology 38 (2024) 2225–2234.

[13] Lun Yu, Fan Yang, Amiao Pan, Yu Zhao, Qichang Dong, Meijie Wang, Miao Guo, Zheng Zhou, Heng Luo, Anti-aging skin care formulation development guided by knowledge graph and artificial intelligence, 33rd IFSCC Congress, Barcelona (2023).

[14] M. Haga, K. Iida, M. Okada, Positive and negative feedback regulation of the TGF-β1 explains two equilibrium states in skin aging, iScience 27 (2024) 109708.

[15] T. Chin, X.E. Lee, P.Y. Ng, Y. Lee, O. Dreesen, The role of cellular senescence in skin aging and age-related skin pathologies, Front. Physiol. 14 (2023).

[16] F. Yang, Z. Zhou, M. Guo, Z. Zhou, The study of skin hydration, anti-wrinkles function improvement of anti-aging cream with alpha-ketoglutarate, J Cosmet Dermatol 21 (2022) 1736–1743.

[17] R.M. Martinez, F.A. Pinho-Ribeiro, V.S. Steffen, C.V. Caviglione, D. Pala, M.M. Baracat, S.R. Georgetti, W.A. Verri, R. Casagrande, Topical formulation containing hesperidin methyl chalcone inhibits skin oxidative stress and inflammation induced by ultraviolet B irradiation, Photochem Photobiol Sci 15 (2016) 554–563.