Bioactive Materials 46 (2025) 37–54

Contents lists available at ScienceDirect

Bioactive Materials

journal homepage: www.keaipublishing.com/en/journals/bioactive-materials

Throw out an oligopeptide to catch a protein: Deep learning and natural language processing-screened tripeptide PSP promotes Osteolectin-mediated vascularized bone regeneration

Yu Chen [a,1], Long Chen [c,1], Jinyang Wu [a,1], Xiaofeng Xu [a], Chengshuai Yang [a], Yong Zhang [a,**], Xinrong Chen [b,***], Kaili Lin [a,****], Shilei Zhang [a,*]

[a] Department of Oral and Cranio-maxillofacial Surgery, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine; College of Stomatology, Shanghai Jiao Tong University; National Center for Stomatology; National Clinical Research Center for Oral Diseases; Shanghai Key Laboratory of Stomatology; Shanghai Research Institute of Stom, Shanghai, 200011, China

[b] Academy for Engineering and Technology, Fudan University, Shanghai Key Laboratory of Medical Image Computing and Computer Assisted Intervention, Shanghai, 200000, China

[c] Stomatology Hospital, School of Stomatology, Zhejiang University School of Medicine, Zhejiang Provincial Clinical Research Center for Oral Diseases, Key Laboratory of Oral Biomedical Research of Zhejiang Province, Cancer Center of Zhejiang University, Engineering Research Center of Oral Biomaterials and Devices of Zhejiang Province, Hangzhou, China

ARTICLE INFO

Keywords:
Deep learning
Natural language processing
Bioactive peptides
Vascular-osteo communication
Osteolectin

ABSTRACT

Angiogenesis is imperative for bone regeneration, yet the conventional cytokine therapies have been constrained by prohibitive costs and safety apprehensions. It is urgent to develop a safer and more efficient therapeutic alternative. Herein, utilizing the methodologies of Deep Learning (DL) and Natural Language Processing (NLP), we proposed a paradigm algorithm that amalgamates *Word2vec* with a *TF-IDF* variant, *TF-IIDF*, to deftly discern potential pro-angiogenic peptides from intrinsically disordered regions (IDRs) of 262 related proteins, where are fertile grounds for developing safer and highly promising bioactive peptides. After the evaluation of the candidate oligopeptides, one tripeptide, PSP, emerged as particularly notable for its exceptional ability to stimulate the vascularization of endothelial cells (ECs), enhance vascular-osteo communication, and then boost the osteogenic differentiation of bone marrow stem cells (BMSCs), evidenced in mouse critical-sized cranial model. Moreover, we found that PSP serves as a 'priming' agent, activating the body's innate ability to produce Osteolectin (Oln) — prompting ECs to release small extracellular vesicles (sEVs) enriched with Oln to facilitate bone formation. In summary, our study established a precise and efficient composite model of DL and NLP to screen bioactive peptides, opening an avenue for the development of various peptide-based therapeutic strategies applicable to a broader range of diseases.

1. Introduction

Bone regeneration is an intricate dynamic process with multiple cells involvement, including inflammatory response, angiogenesis, bone formation and remodeling [1]. Prompt initiation of angiogenesis and restoration of blood supply at the bone injury site are essential for reinvigorating vascular-osteo communication and ensuring effective bone healing [2]. Neovascularization not only ferries oxygen and nutrients but also delivers a suite of growth factors to bone cells, fueling the proliferation and differentiation of osteoblasts and fostering the renewal

Peer review under responsibility of KeAi Communications Co., Ltd.

* Corresponding author.
** Corresponding author.
*** Corresponding author.
**** Corresponding author.
E-mail addresses: zhangyong1362@163.com (Y. Zhang), chenxinrong@fudan.edu.cn (X. Chen), linkaili@sjtu.edu.cn, lklecnu@aliyun.com (K. Lin), leinnymd@hotmail.com (S. Zhang).
[1] These authors have contributed equally to this work and share first authorship.

https://doi.org/10.1016/j.bioactmat.2024.11.011
Received 26 February 2024; Received in revised form 26 September 2024; Accepted 6 November 2024

and restoration of bone tissue [3,4]. Hence, stimulating angiogenesis is paramount for the ideal restoration of bone defects.

Currently, one of the primary strategies of enhancing vascularization to improve bone defect repair involves the delivery of various pro-angiogenesis growth factors, such as vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF), and platelet-derived growth factor BB (PDGF-BB) et al. [5–7]. However, these cytokines face challenges such as high cost, poor stability, antigenicity, and tumorigenicity, which limit their widespread clinical applications [8]. Consequently, researchers have turned their attentions to angiogenic peptides extracted from proteins, such as KLT peptide (15 amino acids), IKVAV peptide (5 amino acids), PHSRN peptide (5 amino acids) and so on [9–11]. These peptides offer considerable benefits over their larger protein counterparts, boasting a lower molecular weight, reduced synthesis costs, better stability, higher bioavailability, lack of antigenicity, and ease of transportation, modification, and recombination [12,13]. These attributes bioactive peptides as highly promising agents in the field of bone regeneration. The traditional methodology for isolating bioactive peptides from protein sequences typically demands laborious manual alignment and segmentation, followed by extensive experimental validation, which is time-consuming, labor-intensive, and prone to errors [14]. Consequently, to streamline the identification process of bioactive peptides from their intricate and heterogeneous structures, it is of paramount importance to distill the essential elements of proteins and refine the screening methodologies for enhanced efficiency and precision.

Proteins serve as the fundamental executors of biological processes. Historically, the "sequence-structure-function" paradigm postulated that the biological function of a protein is contingent upon a stable three-dimensional conformation dictated by its amino acid sequence [15]. However, since the 21st century scientists have begun to notice that protein segments lacking a fixed three-dimensional structure in their native state, known as intrinsically disordered regions (IDRs) [16,17]. These IDRs have been recognized as crucial regulators of protein functionality and play pivotal roles in a vast array of biological processes, including transcription and translation regulation, cell signaling, protein phosphorylation, the storage of small molecules and other aspects [18–20]. Studies suggested that IDRs function through the condensation and partner interactions, mediated by specific peptide motifs [20,21]. The advancing insights into IDRs help us concentrate critical biological segments from bulk protein sequences, narrowing the horizons in the exploration and identification of bioactive peptides.

The advent of artificial intelligence (AI) technologies has ushered in unprecedented convenience for the screening of potential bioactive peptides. DL is an advanced machine learning methodology designed to establish and simulate neural networks for analytical learning in the human brain, and to interpret data by mimicking the mechanisms of the human brain [22]. It is particularly effective at identifying trends, patterns, and clusters within datasets, which extensively applied in areas such as drug discovery and structural prediction [23,24]. NLP is a branch of AI that focuses on the interaction between computers and human language. It involves the development of algorithms and systems that enable computers to understand, interpret, and generate human language in a valuable way [25]. NLP can marry computational linguistics with DL models, enabling the nuanced processing of text and voice data, reaching a full grasp of meaning, intention, and sentiment [26]. Therefore, by constructing appropriate algorithms, DL and NLP can be synergistically employed to mine and characterize potential bioactive peptides within the IDRs of angiogenic proteins. For example, Cai et al. [27] successfully generated osteogenic oligopeptides from the IDRs of osteogenesis-related proteins by *N-gram* model and *Monte carlo* simulation. They tested 28 candidate osteogenic oligopeptides by ARS staining and singled out a pentapeptide, ESSES, with remarkable osteogenic effects both *in vitro* and *in vivo*. Yet, the exploration of angiogenic oligopeptides remains an untapped area of research.

The present study introduced a DL and NLP composite model that amalgamates *Word2vec* with *TF-IIDF*, enabling the identification of a suite of promising bioactive peptides from the IDRs of 262 angiogenic proteins. Based on the result ranking and the greater inherent advantages of the shorter the amino acid (AA) sequences, the tripeptide PSP was picked out for its exceptional angiogenic properties, as determined by qRT-PCR analysis among the top five candidates. *In vitro* and *in vivo* experiments validated that PSP can promote the vascularization of endothelial cells (ECs), enhance vascular-osteo communication, and ultimately benefit osteogenic differentiation of bone marrow stem cells (BMSCs). When incorporated into gelatin methacrylate (GelMA), PSP demonstrated an impressive capacity for repairing critical-sized cranial defects in mice. Furthermore, we discovered that sEVs secreted by ECs after PSP treatment (PSP-sEVs) effectively facilitate the osteogenic differentiation BMSCs. Label-free proteomics analysis revealed that PSP induce ECs to secrete sEVs rich in more Osteolectin (Oln). The comprehensive schematic representation of the experimental procedure of this research was shown in Diagram 1. Overall, our research not only demonstrated the effectiveness and efficiency of the DL and NLP composite model in the strategic identification of angiogenic peptides, but also highlighted the pivotal role of DL and NLP-screened tripeptide PSP in the orchestration of vascular-osteo communication, which provided an innovative therapeutical strategy for bone defects and related pathologies.

2. Materials and methods

2.1. Protein collection and IDR prediction

To collect proteins related to blood vessel formation, we downloaded all proteins listed in the GO terms of "vasculogenesis" and "angiogenesis" from *UniProt* (https://www.prot.org) and yielded 262 of them (Table S1), which were subsequently uploaded on *Metascape* (https://metascape.org/) for GO analysis Biological Process to evaluate whether there is potential interference from unrelated biological function clusters. Then amino acid sequences of all proteins were submitted to *IUPred2A* to identify their IDRs with the parameter "context-dependent predictions (default ANCHOR2)" under the "long disorder" module. Those sequences with IUPred scores larger than 0.5 were regarded as the IDRs for each protein for further analysis (Table S2).

2.2. Angiogenic oligopeptide learning using a DL and NLP model

A DL and NLP composite model based on improved word embedding is constructed in this paper to analyze and understand the above IDRs. In the first step, screen out possible words based on the frequency of their combination occurrence. Secondly, through the pre-trained *Word2vec*, the initial embedding vectors of all words are obtained. Then the *TF-IIDF* values of each word are computed. Finally, in order to remove the unimportant words on the classification results, the words in IDRs with *TF-IIDF* value lower than the threshold are set to zero.

2.2.1. Word2vec

Word2vec, as one of the word embedding methods, is a shallow neural network model that converts words into computable and structured vectors. In *Word2vec*, there are two training modes, which are Continuous Bag-of-Words Model (CBOW) and Continuous Skip-gram Model (Skip-gram). The probability of occurrence of a central word is calculated by CBOW based on consecutive words before and after it, which is selected to obtain the embedding vectors of words. The objective function to predict the center word ω_n by context ω_c can be defined as follows.

$$p(\omega_n|\omega_c) = \frac{exp(\omega_n h_n)}{\sum_{\omega' \in corpus} exp(\omega' h_n)}$$

where h_n represents the mean vector in the context of word ω_n.

2.2.2. TF-IIDF

TF-IDF is a commonly used text mining technique. It can evaluate the importance of words to a document in a corpus by examining the frequency of their appearance in the document and their frequency in a corpus. *TF* indicates a word's frequency in a document, which can be used to evaluate the importance of oligopeptides, and the calculation formula is as follows.

$$TF(t,d) = \frac{n_{t,d}}{\sum_k n_{k,d}}$$

where $n_{t,d}$ represents the number of occurrences of the tth word in the document d.

IDF indicates a word's the rarity in the whole corpus, which is opposite to potential angiogenic oligopeptide. So this indicator has been changed to Inverse Inverse Document Frequency (*IIDF*) and the

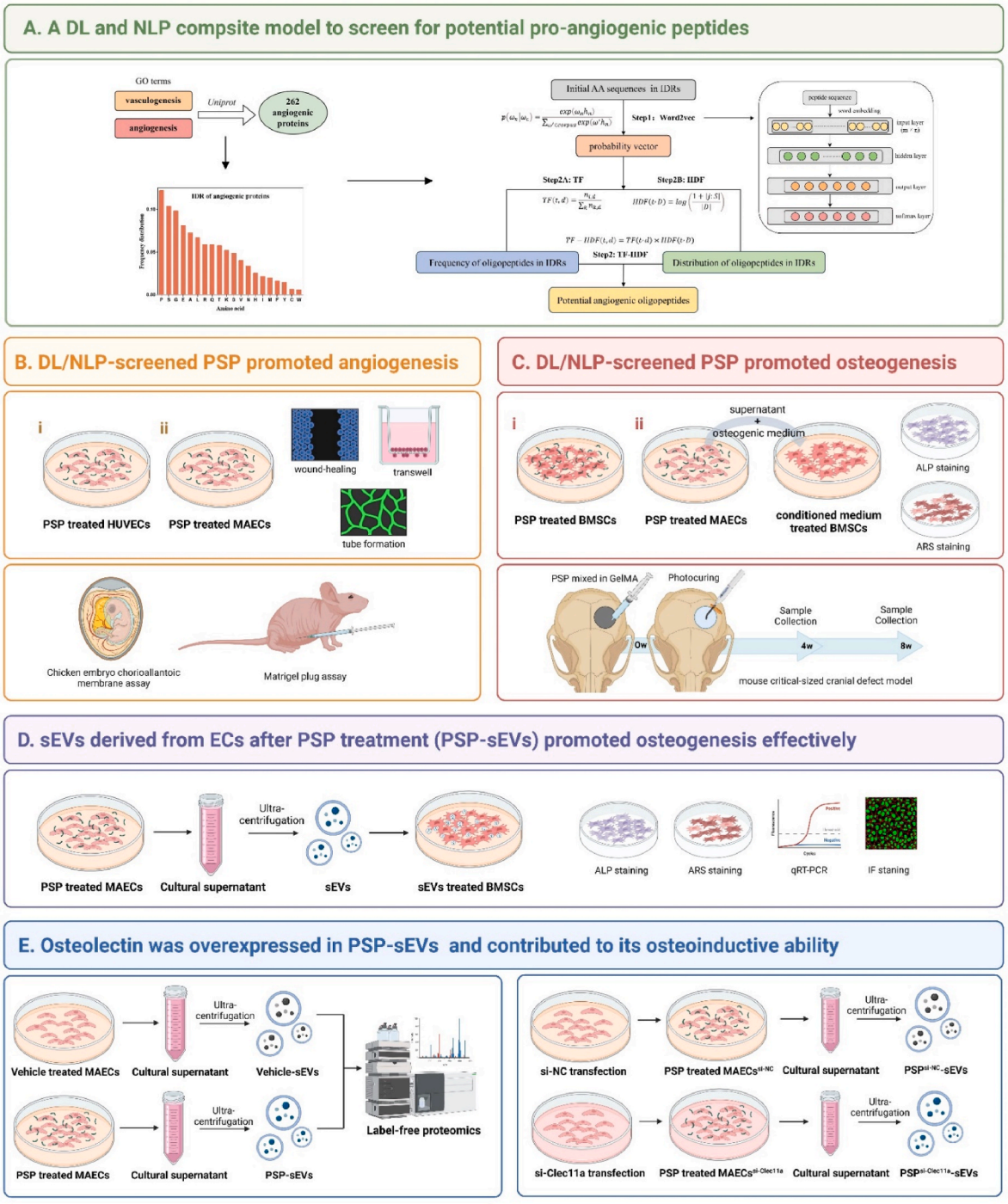

Diagram 1. Comprehensive schematic representation of the experimental procedure. (A) The procedure of AI screening and the framework of our DL and NLP composite model; (B) Experimental verification of the *in vitro*, *ex vivo* and *in vivo* effects of PSP on angiogenesis; (C) Experimental verification of the *in vitro*, and *in vivo* effects of PSP on angiogenesis; (D) Experimental verification of the osteogenic effects of PSP-sEVs; (E) Experimental verification of the osteogenic effects of Oln in PSP-sEVs.

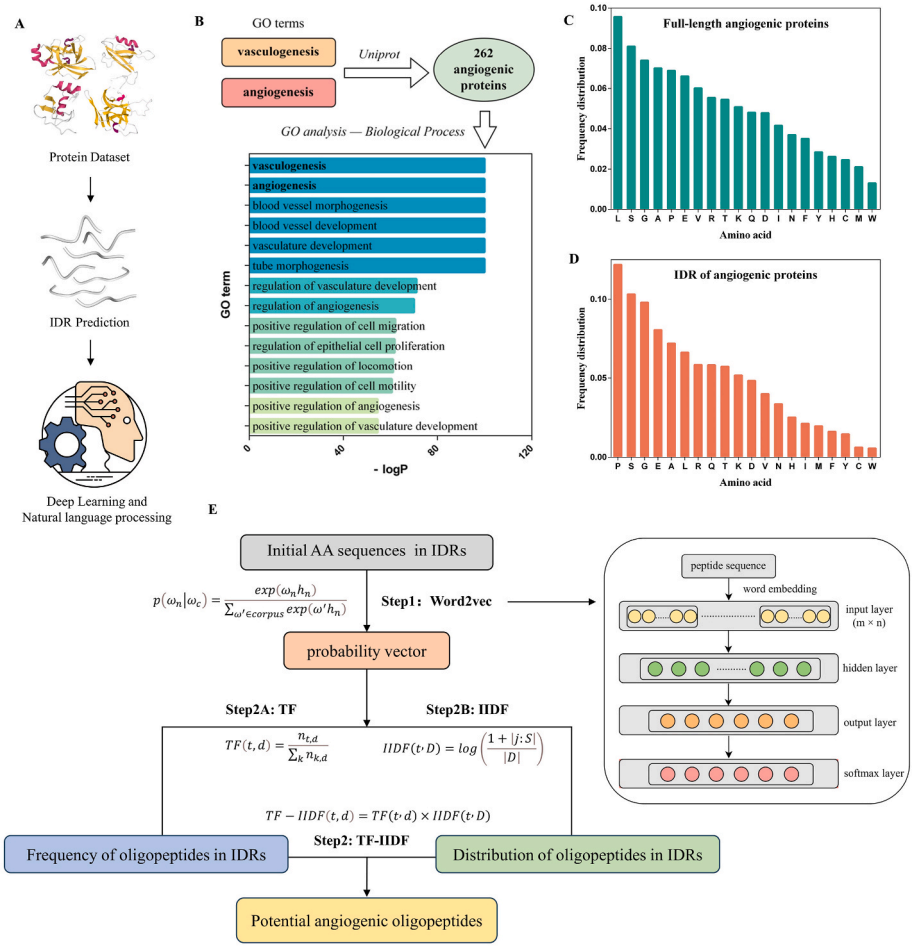

Fig. 1. A DL and NLP model to screen for potential pro-angiogenic peptides. (A) Illustration of the procedural workflow; (B) Selection of 262 protein candidates from UniProt annotated with the GO terms "vasculogenesis" and "angiogenesis"; (C) AA frequency distribution in full-length sequences of the above proteins; (D) AA frequency distribution in IDRs of the above proteins; (E) Outline of the architecture of the DL and NLP model, which consists of two complementary modules: *Word2vec* and *TF-IIDF*.

calculation formula is as follows.

$$IIDF(t,D) = log\left(\frac{1+|j:S|}{|D|}\right)$$

where $|D|$ represents the total number of IDRs segments in angiogenic proteins and $|j:S|$ denotes the total number of IDRs segments containing the potential angiogenic oligopeptide. Here, for the convenience of calculation, *IIDF* can be simplified as

$$IIDF(t,D) = \frac{1+|j:S|}{|D|}$$

The *TF-IIDF* value of word t in document d is then calculated by:

$$TF-IIDF(t,d) = TF(t,d) \times IIDF(t,D)$$

TF-IIDF value is calculated to represent the importance of word t in document d.

2.2.3. Filter of embedding vector

The larger the *TF-IIDF* value, the more important the word is. Therefore, removing the words with smaller *TF-IIDF* values can increase the classification accuracy.

2.3. Oligopeptides synthesis

The oligopeptides were synthesized from Sangon Biotech (Shanghai) Co.,Ltd.(purity≧98 %). High-performance liquid chromatography (HPLC) and mass spectrometry (MS) data were provided. Peptides applied for further cellular experiments were dissolved at the concentration of 10 mM. The random peptide AEP and angiogenic peptide KLTWQELYQLKYKGI were used as controls.

2.4. Animals

The animals employed in this study were sourced from the Laboratory Animal Center at the 9th People's Hospital, affiliated to the Shanghai Jiao Tong University School of Medicine. All experimental protocols were reviewed and approved by the Animal Care and Use Committee of Shanghai Jiao Tong University, ensuring adherence to ethical guidelines for animal research.

2.5. Cell culture

BMSCs were isolated and cultured as follow. Femurs and tibias from 4-week-old male C57BL/6 mice were harvested. The metaphyses were excised and the marrow cavities were flushed with Minimum Essential Medium Alpha (aMEM) with 10 % fetal bovine serum (FBS) (Gibco) until white. The single-cell suspension was prepared by repeated aspiration and then plated onto 100 cm^2 culture dishes. MAECs were isolated and cultured with the following procedure. Under a dissecting microscope, the aorta was extracted from 4-week-old male C57BL/6 mice and thoroughly washed with PBS. The inner endothelial layer was

carefully turned inside out and placed against the surface of a culture dish. Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10 % FBS was added in sufficient volume to just submerge the tissue. HUVECs was purchased from ATCC (Manassas, VA, USA) and also cultured in DMEM containing 10 % FBS. All cells were cultivated in a humidified atmosphere with 5 % CO_2 at 37 °C.

2.6. Cell migration assay

The wound-healing assay and Transwell assay were carried out to explore the pro-migration influence of the PSP on HUVECs and MAECs. For the wound-healing assay, culture-inserts (ibid, 80209, Germany) were utilized to generate a uniform scratch. The cells were treated with 0.2 mM and 0.4 mM concentrations of PSP and KLT peptides, respectively, and incubated for 24 h. Subsequently, the cells underwent fixation and were stained with DAPI. Images documenting the extent of wound closure were captured and analyzed utilizing Image J software. In Transwell assay, add 0.2 mM or 0.4 mM PSP and KLT peptides respectively in the lower chamber of a 24-well Transwell plate (Corning, 3422, NY) and culture for 24 h. Then the cells were fixed and stained with crystal violet. Non-migratory cells on the upper surface were carefully removed, and the migratory cells adhering to the bottom surface were photographed and quantified with Image J.

2.7. Tube formation assay

To execute the tube formation assay, a 48-well plate was prepared by adding 150 μl of Matrigel matrix to each well. After solidification, HUVECs and MAECs were then seeded onto the Matrigel at a density of 3×10^5 cells/DL, with wells receiving 0.2 mM or 0.4 mM PSP and KLT peptides respectively. Following a 6-h incubation period, images capturing the network of tube-like structures were taken. The structures were highlighted through Calcein AM staining, and the degree of tube formation was quantitatively assessed using Image J.

2.8. Chicken embryo chorioallantoic membrane (CAM) assay

Fertilized sterile eggs were incubated under stringent conditions at 37 °C with 60 % humidity for a period of 7 days. The eggs' air chambers were delicately opened to reveal the CAM with great care to avoid any damage. A GelMA hydrogel globule, mixed with 10 mM concentrations of PSP or KLT peptides, was gently placed onto the CAM surface. The openings were then securely resealed using a transparent sealing film and the eggs were returned to the incubator for an additional 3 days. Photograph the vessels on the CAM and analyze with Image J.

2.9. Matrigel plug assay

To evaluate the pro-angiogenic differentiation potential of PSP *in vivo*, Matrigel plug assays were performed according to the previous report [28]. 5-week-old female BALB/c mice were randomly assigned into three distinct groups. For the assay preparation, 100 μl of MAECs at a concentration of 1×10^7 cells/ml were thoroughly blended with 400 μl of Matrigel matrix, which was premixed with 10 mM of either PSP or KLT peptides. These cell-matrix admixtures were then carefully injected subcutaneously into the right ventral region of the mice. After 7 days, the Matrigel plugs were removed for HE staining and immunohistochemical staining of CD31 and VEGF.

2.10. Osteogenic differentiation assay

BMSCs were seeded in a 24-well plate and cultured in the osteogenic induction medium (Cyagen, USA) containing corresponding stimuli. The medium was refreshed every 2–3 days. Following a 7-day culture period, we performed alkaline phosphatase (ALP) staining and carried out a semi-quantitative assessment according to the guidelines provided with the BCIP/NBT ALP color development kit (Beyotime, China) and the ALP assay kit (Nanjing Jiancheng, China). The total protein content of the corresponding wells was determined by BCA Protein Assay Kit (Sangon Biotech, China). Relative ALP activity was calculated as OD_{520nm}/total protein content (mg/ml). After 21 days of culture, alizarin red (ARS) staining and its quantitation were performed. Cells were fixed with 4 % PFA and stained with ARS solution. After photography, the calcium nodules were completely dissolved using a 10 % cetylpyridinium chloride solution and recorded the absorbance at 562 nm. The total protein content of the corresponding wells was determined. Quantitative ARS staining was calculated as OD_{562nm}/total protein content (mg/ml).

2.11. qRT-PCR analysis of angiogenesis and osteogenesis

Total RNA was exacted from HUVECs, MAECs, and BMSCs post-cultivation under respective conditions utilizing the RNA fast200 extraction kit (Fastagen, China). Subsequent reverse transcription of RNA into cDNA was executed employing PrimeScript™ RT Master Mix kit (Perfect Real Time) (Takara, Japan) in a 20-μl reaction volume. qPCR analyses were conducted in LightCycler@96 Instrument (Roche) using the TB Green Premix Ex Taq™ kit (Tli RNaseH Plus) (Takara). β-action served as the internal control for normalization purposes. Primer sequences employed for the qRT-PCR assays are delineated in Supplemental Table 1. The relative expression level of the gene was calculated using the $2^{-\triangle\triangle CT}$ method.

2.12. Cellular immunofluorescence staining

Cells were seeded on slides pre-treated with polylysine under respective conditions. After fixation, the cells were treated with rabbit anti-Runx2 antibody (1:200, 20700-1-AP, Proteintech, USA), rabbit anti-Osterix antibody (1:200, ab209484, Abcam, UK), rabbit anti-OCN antibody (1:200, 23418-1-AP, Proteintech, USA), rabbit anti-Col1 antibody (1:200, PA5-95137, Invitrogen, USA) respectively. Subsequently, Alexa Fluor®594-labeled anti-rabbit secondary antibody (1:1000, 8889, CST, USA) and FITC-labeled phalloidin were sequentially added. The slide was then covered with an anti-fade mountant containing DAPI. All immunofluorescence staining images were captured using an Olympus fluorescence microscope (Germany).

2.13. Western blotting (WB) assay

Following their respective stimulatory treatments, all cells were processed to extract total proteins using RIPA lysis buffer. The protein samples were then separated using ExpressPlus™ PAGE Gels (Genscript, China), and subsequently transferred onto NC membranes. The Western blot assay was performed using the following primary antibodies: rabbit anti-Hsp70 antibody (1:1000, GB11241, Servicebio, China), mouse anti-Tsg101 (1:1000, sc-7964, Santa Cruz, USA), rabbit anti-CD31 antibody (1:1000, GB11063, Servicebio, China), rabbit anti-Oln antibody (1:1000, YT5719, Immunoway, USA), and rabbit anti-β-actin antibody (1:1000, 4967, CST, USA). The HRP-conjugated Goat Anti-Rabbit secondary antibody (1:10000, SA00001-2, Proteintech, USA) and the HRP-conjugated Goat Anti-Mouse secondary antibody (1:10000, RGAM001, Proteintech, USA) was then applied. The semi-quantitative statistical analysis of the Western blot was conducted using Image J.

2.14. Critical-sized cranial defect model

The GelMA, PSP-GelMA and KLT-GelMA hydrogels (100 mM PSP or KLT) were applied in the mouse critical-sized cranial defect model. After freeze-drying, the morphology was captured using scanning electron microscopy (SEM, ZEISS, Germany).

Forty 8-week-old male C57BL/6 mice were randomly assigned into four groups: Blank, GelMA, PSP-GelMA and KLT-GelMA, to undergo respective experimental interventions. Following anesthesia induction,

the surgical site was aseptically prepared. A longitudinal incision was made along the skull's midline to expose the underlying periosteum. The periosteum was carefully incised to provide unobstructed access to the area of interest. Subsequently, a 4 mm diameter defect was meticulously created using a precision dental drill. UV curing after inserting different hydrogels. The incision was then closed with 4–0 surgical sutures (Johnson, USA). At 4- and 8-weeks post-surgery, the mice were euthanized to facilitate subsequent analyses.

2.15. Micro-CT characterization

4-and 8- weeks following the surgical procedure, the mice were anesthetized, humanely euthanized. We then carefully extracted the mouse skulls and preserved them in a 4 % PFA solution. These samples were processed using a Micro-CT scanner (with a resolution of 9 μm, at 48 kV and 201 mA, utilizing SkyScan 1076, Belgium). The raw scan data was reconstructed using CTAn software, which was subsequently employed to analyze the formation of new bone. A cylinder with a diameter of 4 mm and a thickness of 0.5 mm centered on the defect center was set as the region of interest (ROI) with a bone segmentation threshold set at 2000.

2.16. Histological staining

Following decalcification, the mice skull samples were enshrined in paraffin. The skull sections were then treated with Hematoxylin and Eosin (H&E) staining and Masson staining. Primary mouse anti-CD31 antibody (1:200, 66065-2-lg, Proteintech, USA), primary rabbit anti-EMCN antibody (1:200, YT1557, Immunoway, USA) and primary rabbit anti-OCN antibody (1:200, GB11233, Servicebio, China) were used for immunofluorescence (IF) and immunohistochemistry (IHC) staining. Alexa Fluor®594-labeled anti-mouse secondary antibody (1:1000, 8890, CST, USA) and Alexa Fluor®488-labeled anti-rabbit secondary antibody (1:1000, 4412, CST, USA) were used as secondary antibodies.

2.17. Isolation and characterization of sEVs

The cell supernatants were gathered after a 48-h incubation period of MAECs in a culture medium supplemented with 2 % exosome-depleted FBS. This medium was then put through a filtration process using a 0.22 μm filter and subjected to centrifugation at 10,000×*g* for 30 min. This was followed by an additional centrifugation at 100,000×*g* for 90 min at a temperature of 4 °C. The result of this process provided us with MAECs derived small extracellular vesicles (sEVs). We utilized Transmission Electron Microscopy (TEM, Hitachi, Japan) to examine the microstructure of the sEVs. To evaluate the particle concentration and diameter distribution, we employed Nanoparticle Tracking Analysis (NTA, NanoSight300, Malvern, UK). We used WB to identify the presence of sEV surface markers Hsp70 and Tsg101.To investigate their internalization, sEVs were labeled with Dil dye and centrifuge to remove unadhered dyes. BMSCs were then co-cultured with labeled sEVs for 8 h, subsequently fixed and stained with DAPI.

2.18. Label-free proteomics of sEVs

The sEVs were extracted from the supernatant of three MAECs cultures with or without PSP treatment. After protein extraction, the samples were subjected to label-free proteomics, which was performed by Applied Protein Technology (Shanghai) Co., Ltd. LC-MS/MS analysis was performed on a Q Exactive mass spectrometer (ThermoFisher Scientific) that was coupled to Easy nLC (Proxeon Biosystems, ThermoFisher Scientific). Related parameters and instructions of identification and quantitation of proteins are as follows: Enzyme = Trypsin, Fixed modification: Carbamidomethyl (C), Variable modification: Oxidation (M), Peptide mass tolerance = 20 ppm, MS/MS tolerance = 0.1 Da, Missed cleavage = 2, FDR ≤0.01.

2.19. Enzyme linked immunosorbent assay (Elisa)

The samples from each sEV group were tested according to the method indicated by the Mouse CLEC11A ELISA Kit (Shkbrbio, China). The results were determined by measuring the OD values of each sample at 450 nm and converting the concentrations of the samples through the standard curve.

2.20. Inhibitor transfection

Small interfering RNA (siRNA) sequences targeting the genes in the experiment, as well as negative controls, were supplied by Genomeditech (Shanghai) Co., Ltd. (see Supplementary Table 2 for details). The siRNA can bind to the target mRNA, leading to its degradation. MAECs were seeded in a 12-well plate at a density of $5x10^3$ cells per well. Once the cells achieved 40 %–50 % confluence, they were transfected with siRNAs employing Lipofectamine 3000 (Invitrogen, USA). The success of the transfection procedure was verified using qRT-PCR and WB analysis.

2.21. Statistical analysis

All data were measured at least in triplicate and were presented as the mean value accompanied by the standard deviation (mean ± SD). The data was analyzed using GraphPad 8.0 software. Student's t-test and one-way analysis of variance (ANOVA) were applied. A significance level of $P < 0.05$ was used to determine statistical significance.

3. Results

3.1. A DL and NLP model to screen for potential pro-angiogenic peptides

We have developed a DL and NLP composite model dedicated to screening for oligopeptides with the potential to promote angiogenesis. Our approach involved extracting IDRs from proteins that are functionally annotated as being associated with angiogenic processes. These IDRs were then employed as inputs for the DL and NLP model to identify candidate pro-angiogenic peptides. The entire procedural workflow is depicted in Fig. 1A. Initially, proteins implicated in blood vessel formation were selected from the UniProt database using two angiogenic Gene Ontology (GO) terms: "vasculogenesis" and "angiogenesis", as demonstrated in Fig. 1B, yielding a collection of 262 protein candidates (Table S1). Subsequently, the IDRs of these proteins were pinpointed using the IUPred2A tool (Table S2). A comparative analysis revealed a significant disparity in the frequency distribution of AA residues between the complete protein sequences and the IDRs, as illustrated in Fig. 1C&D. The DL and NLP model utilized the *Word2vec* model to discern the characteristic patterns of IDRs and assess the correlation between amino acids (AAs), along with the *TF-IIDF* model, a modified *TF-IDF* algorithm, tailored to value the importance of the oligopeptide segments based on the frequency and distribution of potential angiogenesis oligopeptides within the IDRs respectively, which is outlined in Fig. 1E. The execution of the word2vec model is based on the Gensim library's implementation, which provides multiple parameters to adjust the training process. The basic implementation code is "*model = Word2Vec(LineSentence(inp), vector_size=400, window=1, min_count=5, workers=multiprocessing.cpu_count())*", in which size is the feature vector dimension, window is the context window size, min_comunt filters low-frequency vocabulary, and workers control the number of parallel training threads. Here, CBOW is used and amino acids are used as the central word, and IDR sequences are used as the context. The results of *Word2vec* and *TF-IIDF* were shown in Tabe S3 and S4, respectively. Those oligopeptide sequences containing consecutive correlated AAs had been excluded.

On the one hand, we arranged the selected candidate oligopeptides in descending order based on their *TF-IIDF* values, with tripeptides generally ranking at the top of the list. On the other hand, shorter AA

sequences can better leverage the inherent advantages of bioactive peptides — such as smaller the molecular weight, lower the synthesis cost, easier to transport, modify, and recombine [29,30]. Consequently, in the first batch of experiments, we validated the top five oligopeptides — SPP, SSP, SSG, PPS, and PSP. The assessment employed qRT-PCR analysis to scrutinize the expression levels of key angiogenic genes (*VEGF* and *FGF*) in HUVECs after treatment with these tripeptides. For comparative purposes, KLT peptide was utilized as a positive control, while a random peptide sequence served as the negative control. We found that SPP, SSP, PPS, and PSP were capable of enhancing the expression of angiogenic genes to some extent compared to the blank group. Remarkably, the PSP sequence demonstrated stronger efficacy in promoting this expression even compared to the KLT peptide (Fig. S1A). We also used different concentrations of PSP to treat HUVECs to evaluate the optimal concentration range of the tripeptide and found that 200 μM PSP exerts an outstanding upregulation of genes *VEGF* and *FGF* (Fig. S1B). These results demonstrated the practical application of our AI system in pinpointing bioactive peptides with a high potential for pro-angiogenic activity efficiently and accurately. Based on the preliminary verification, tripeptide PSP performed better than the positive

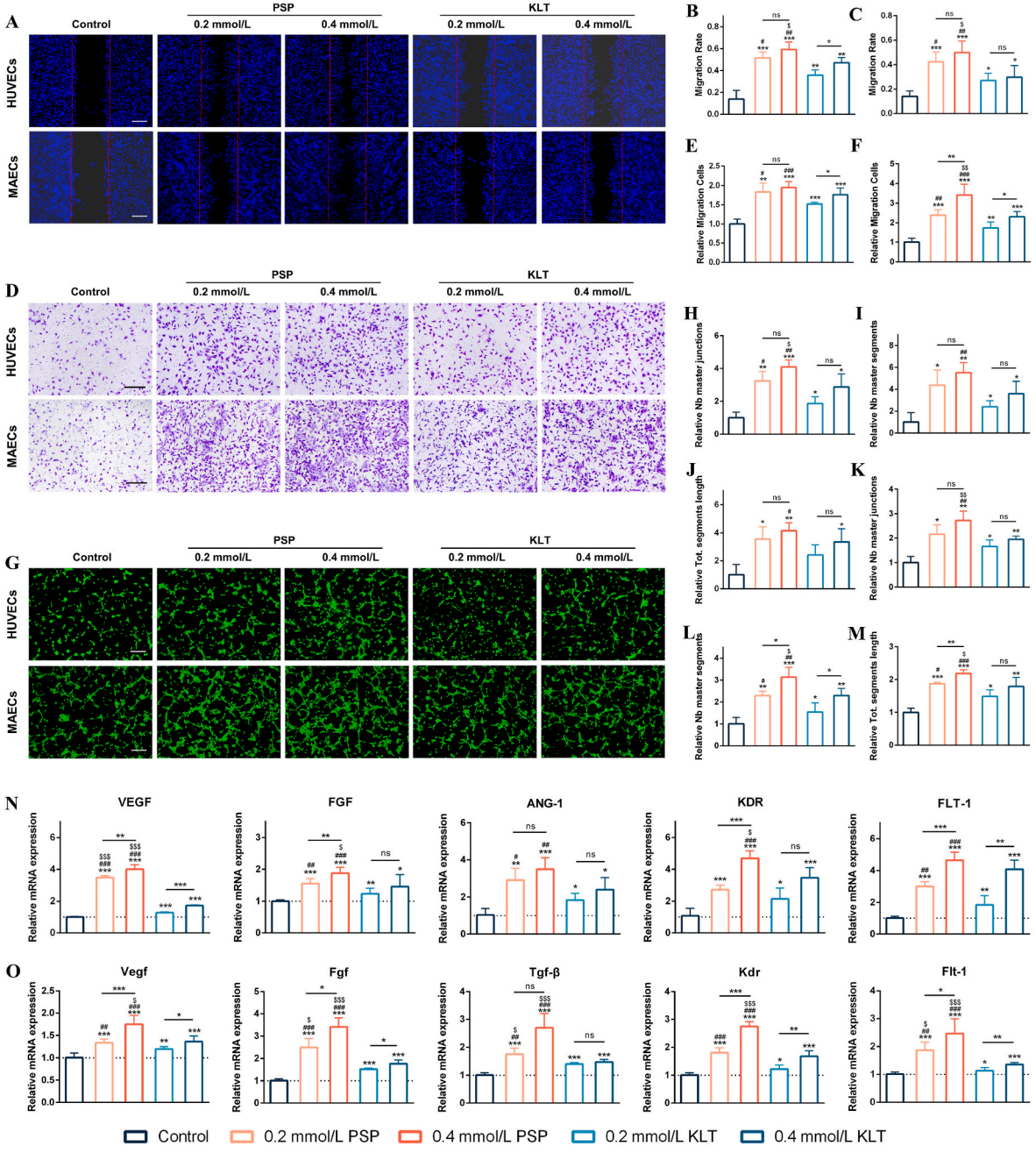

Fig. 2. DL and NLP-screened tripeptide PSP facilitated the angiogenesis of ECs in vitro. (A) Representative images of wound-healing assay after 24 h culture (scale bar, 200 μm), and the quantitative analyses of migration rate of HUVECs (B) and MAECs (C) according to wound-healing assays; (D) Representative images of Transwell assay after 24 h culture (scale bar, 500 μm), and the quantitative analyses of migration cells of HUVECs (E) and MAECs (F) according to Transwell assays. (G) Representative images of tube formation after 6 h culture (scale bar, 200 μm), and quantitative analyses of number of master junctions, number of master segments and total segments length of HUVECs (H–J) and MAECs (K–M); mRNA expression of angiogenic-related genes of HUVECs (N) and MAECs (O) by qRT-PCR after 3 d culture. Data are presented as means ± SD. Significant difference with respect to the control groups (*) or to 0.2 mM KLT group (#) or to 0.4 mM KLT group ($), *, #, $ $P < 0.05$, **, ##, $$ $P < 0.01$, ***, ###, $$$ $P < 0.001$.

control KLT peptide and was used for the following researches.

3.2. DL and NLP-screened tripeptide PSP facilitated the angiogenesis of ECs in vitro

To further investigate the promotive effect of PSP on angiogenesis, we carried out a comprehensive set of experiments *in vitro* using two kinds of ECs: HUVECs and MAECs. Recognizing the migration of ECs as an indispensable step in angiogenesis, we assessed the impact of PSP on ECs motility using wound-healing assay and Transwell assay. ECs were treated with PBS or with 0.2 mM, 0.4 mM PSP or KLT peptide. Fig. 2A–C and Fig. 2D–F illustrated the results of wound-healing assay and Transwell assay, respectively. It was shown that PSP remarkably enhances the migratory activities of both HUVECs and MAECs, surpassing the levels observed in PBS group and KLT groups at identical concentrations. Further, we conducted tube-formation assays to evaluate the *in vitro* angiogenic potential of the peptides. Fig. 2G–M demonstrated that PSP notably increased tube formation compared to both the control and KLT-treated groups. An upregulation in the mRNA expression of angiogenesis-related genes—*VEGF*, *FGF*, *ANG-1*, *KDR*, and *FLT-1* in HUVECs, and *Vegf*, *Fgf*, *Tgf-β*, *Kdr*, and *Flt-1* in MAECs—was observed in response to both PSP and KLT treatments when compared to the PBS group. Notably, PSP induced a more pronounced upregulation of these genes than KLT at the same concentrations, as shown in Fig. 2N&O. *In vitro* experiments collectively revealed that PSP could notably enhances angiogenesis of ECs, exhibiting a superior effect compared to the KLT peptide.

3.3. DL and NLP-screened tripeptide PSP facilitated the angiogenesis of ECs ex vivo and in vivo

To substantiate the pro-angiogenic findings observed *in vitro*, we

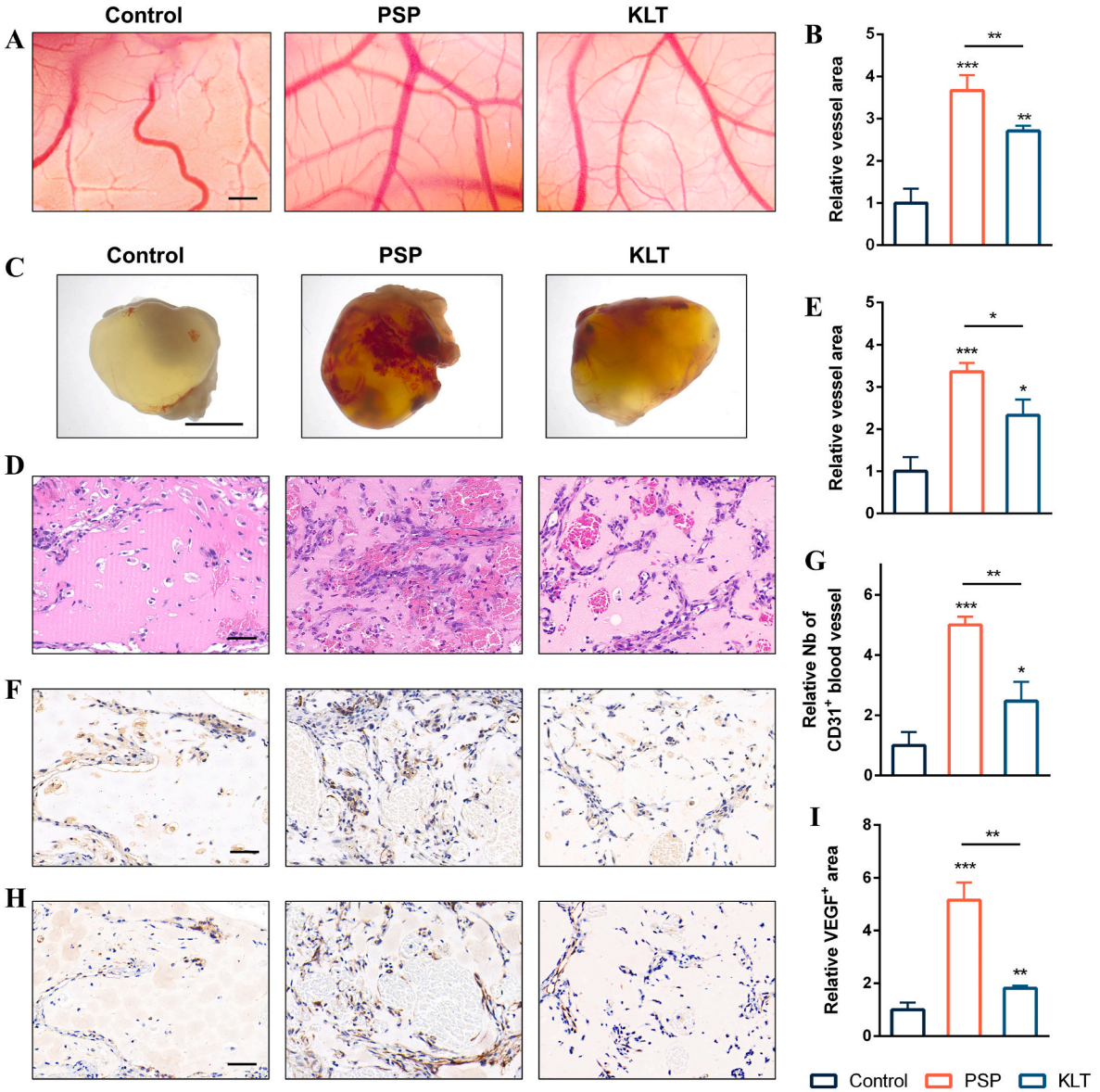

Fig. 3. DL and NLP-screened tripeptide PSP facilitated the angiogenesis of ECs *ex vivo* and *in vivo*. (A) Representative images of CAM assay (scale bar, 500 μm), and quantitative analysis of the vessel area (B); (C) Representative images of Matrigel plug assay (scale bar, 500 μm); (D) Representative images of the HE staining of the gel (scale bar, 50 μm) and quantitative analysis of the vessel area (E); (F) Representative images of CD31 immunohistochemistry staining of the gel (scale bar, 50 μm) and (G) quantitative analysis of the number of CD31 positive blood vessel; (H) Representative images of VEGF immunohistochemistry staining of the gel (scale bar, 50 μm) and (I) quantitative analysis of the VEGF positive area. Data are presented as means ± SD. Significant difference with respect to the control groups (*), *P < 0.05, **P < 0.01, ***P < 0.001.

performed *ex vivo* and *in vivo* angiogenic assays. Illustrated in Fig. 3A&B, CAM assay outcomes revealed that the vessel area was significantly expanded in the PSP-treated group relative to both the control and KLT groups. Subsequently, the Matrigel plug assay was employed to evaluate the *in vivo* angiogenic potential of PSP (as shown in Fig. 3C). HE staining of the Matrigel plugs uncovered a substantial increase in the density of blood vessels filled with red blood cells (Fig. 3D). Quantitative analysis further confirmed this observation, indicating that PSP markedly elevated the number of vessels compared to the control and KLT groups (Fig. 3E). In addition, CD31 immunostaining was performed to verify the presence of ECs and the formation of new blood vessels (Fig. 3F). The accompanying quantitative analysis showed that PSP significantly boosted the count of CD31-positive blood vessels relative to the control and KLT groups (Fig. 3G). Meanwhile, by VEGF staining and its quantitative analysis, we confirmed PSP can notably increase the expression of VEGF relative to the control and KLT groups (Fig. 3H&I). Collectively, these findings underscored PSP's more potent capacity to stimulate angiogenesis *ex vivo* and *in vivo* in comparison to the control and KLT peptides.

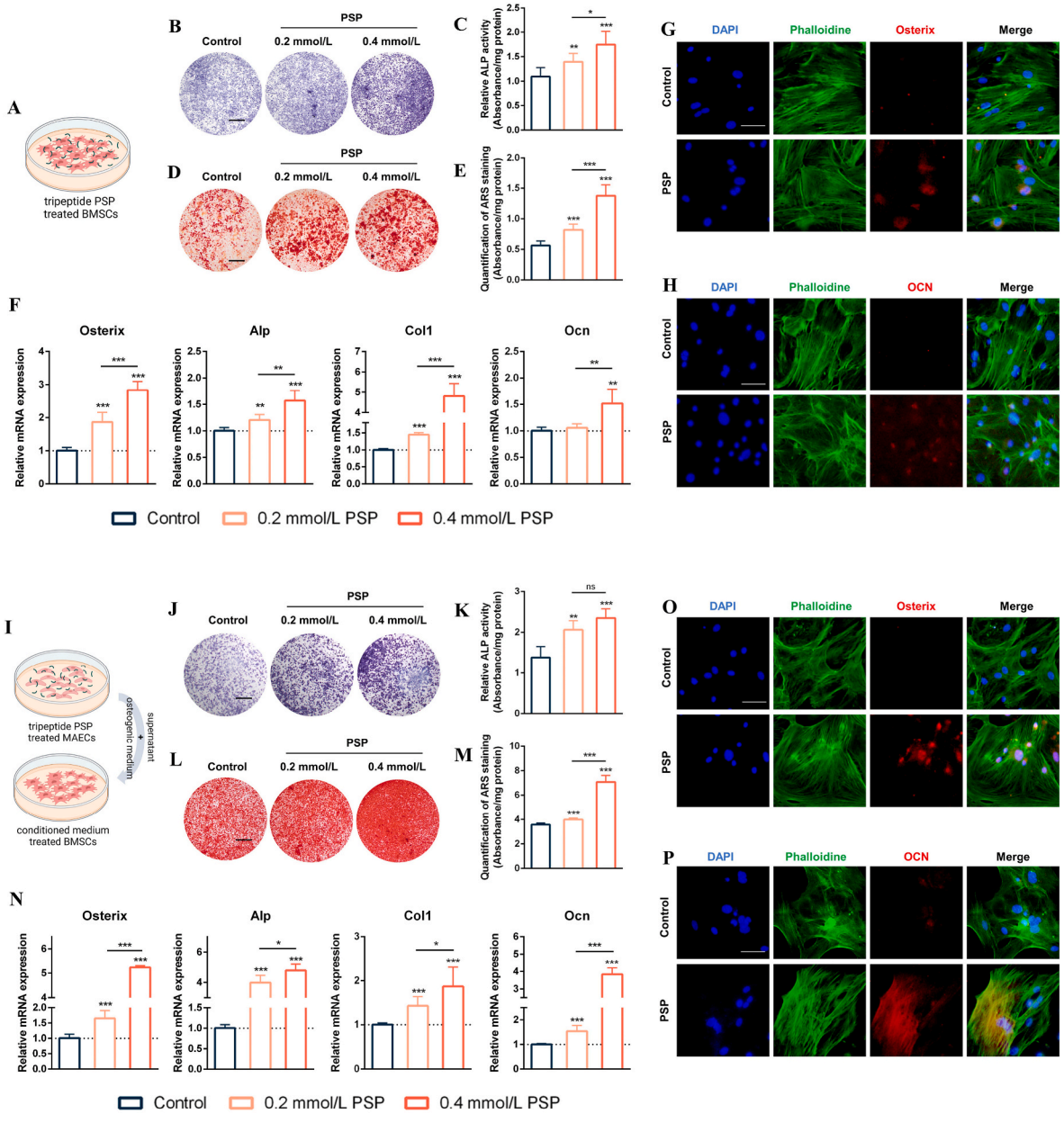

Fig. 4. DL and NLP-screened tripeptide PSP promoted osteogenesis in vitro. (A) Illustration of the experimental condition; (B) Representative images of ALP staining after 7 days culture (scale bar, 1 mm), and its semi-quantitative analysis (C); (D) Representative images of ARS staining after 21 days culture (scale bar, 1 mm), and its semi-quantitative analysis (E); (F) mRNA expression of osteogenic-related genes of BMSCs by qRT-PCR after 7 d culture; (G) Representative images of immunofluorescence staining of Osterix (DAPI in blue, Phalloidin in green, Osterix in red, scale bar, 50 μm); (H) Representative images of immunofluorescence staining of OCN (DAPI in blue, Phalloidin in green, OCN in red, scale bar, 50 μm); (I) Illustration of the experimental condition; (J) Representative images of ALP staining after 7 days culture (scale bar, 1 mm), and its semi-quantitative analysis (K); (L) Representative images of ARS staining after 21 days culture (scale bar, 1 mm), and its semi-quantitative analysis (M); (N) mRNA expression of osteogenic-related genes of BMSCs by qRT-PCR after 7 d culture; (O) Representative images of immunofluorescence staining of Osterix (DAPI in blue, Phalloidin in green, Osterix in red, scale bar, 50 μm); (P) Representative images of immunofluorescence staining of OCN (DAPI in blue, Phalloidin in green, OCN in red, scale bar, 50 μm); Data are presented as means ± SD. Significant difference with respect to the control groups (*), *P < 0.05, **P < 0.01, ***P < 0.001.

3.4. DL and NLP-screened tripeptide PSP promoted osteogenic differentiation of BMSCs in a direct way and in an angiogenesis-mediated way

We performed a comprehensive analysis of the osteogenic effect of PSP on BMSCs through both direct contact and angiogenesis-driven way. As shown in Fig. 4A, BMSCs directly underwent treatment with PBS, 0.2 mM PSP or 0.4 mM PSP. Fig. 4B&C illustrated the ALP staining and semi-quantitative analysis of ALP activity respectively after 7 days of osteogenic induction. The early osteogenic differentiation of BMSCs was enhanced under PSP exposure. Fig. 4D&E demonstrated the ARS staining and the accompanying semi-quantitative analysis after 21-day osteogenic induction. In the late stage of osteogenesis, PSP was also favorable to the osteogenic mineralization of BMSCs. As shown in Fig. 4F, PSP treatment significantly upregulated the expression of the osteogenic markers *Osterix*, *Alp*, *Col1* and *Ocn*. Our findings collectively indicated that a higher concentration of PSP (0.4 mM) exhibited a more substantial osteogenic effect compared to the lower concentration (0.2 mM). Furthermore, as evidenced by the immunofluorescent staining shown in Fig. 4G&H, treatment with 0.4 mM PSP notably boosted the protein expression levels of Osterix and OCN.

Fig. 4I illustrates that MAECs were exposed to PBS, 0.2 mM PSP, or 0.4 mM PSP for a duration of 3 days. Subsequently, the collected supernatants were blended with osteogenic induction medium in a 1:1 ratio and applied to treat BMSCs. Fig. 4J&K illustrated the ALP staining and semi-quantitative analysis of ALP activity respectively after a week of exposure to the conditioned medium. The results indicated that conditional cultural inclusive of PSP significantly enhanced the early osteogenic differentiation of BMSCs. Additionally, Fig. 4L&M displayed the ARS staining and its semi-quantitative analysis following a 21-day period of conditional cultural, revealing that PSP notably promoted osteogenic mineralization in BMSCs through a mechanism mediated by angiogenesis. As shown in Fig. 4N, PSP treatment significantly upregulated the expression of the osteogenic markers *Osterix*, *Alp*, *Col1* and *Ocn*. It is worth noting that the conditional medium with 0.4 mM PSP exhibited a dramatic osteogenic effect. Additionally, immunofluorescent staining images portrayed in Fig. 4O&P illustrated that the involvement of 0.4 mM PSP markedly augmented the protein expression of Osterix and OCN.

To differ and compare the direct osteogenic effect and the angiogenesis-mediated (indirect) osteogenic effect of PSP on BMSCs, we conducted the following experiments. As shown in Fig. S2A, FITC labeled PSP (FITC-PSP) was added in medium and cultured either without MAECs or with MAECs for 3 days. The resulting supernatants were then mixed with osteogenic induction medium to culture BMSCs, named direct culture group and indirect culture group, respectively. Representative images of FITC-PSP interacting with BMSCs after 12 h culture were exhibited in Fig. S2B. A large amount of green fluorescence was observed in direct culture group. Conversely, minimal green fluorescence was observed in indirect culture group, indicating that there was little or no PSP in the supernatant of the collected MAECs medium. Fig. S2C&D illustrated the ALP staining and semi-quantitative analysis of ALP activity respectively, revealing that indirect culture of PSP is more effective in promoting osteogenesis than direct culture. Additionally, Fig. S2E&F displayed the ARS staining and its semi-quantitative analysis also shown the same trend. As shown in Fig. S2G, angiogenesis-mediated (indirect) treatment of PSP significantly upregulated the expression of the osteogenic markers *Osterix*, *Alp*, *Col1* and *Bsp* relative to direct treatment.

Through the above experiment, we found that PSP could facilitated the osteogenic differentiate of BMSCs both in a direct way and in an angiogenesis-mediated way. Furthermore, angiogenesis-mediated treatment of PSP was more effective in promoting osteogenesis than direct treatment.

3.5. DL and NLP-screened tripeptide PSP promoted vascularized bone regeneration in critical-sized cranial defect model

To further investigate the *in vivo* effect of PSP on vascularized bone regeneration, we incorporated PSP into GelMA hydrogel (PSP-GelMA) before introducing it into critical-sized cranial defects in mice, using blank and unmodified GelMA hydrogel and GelMA hydrogel with KLT incorporation (KLT-GelMA) as controls. Fig. 5A illustrated the experimental procedure. As shown in Fig. 5B, the unmodified GelMA hydrogel, PSP-GelMA hydrogel and KLT-GelMA hydrogel all exhibited the similar porous structures. Micro-CT scans of the mouse skulls from each group were conducted at 4- and 8-weeks post-implantation. Micro-CT reconstructions from the top view, coronal view and sagittal view were shown in Fig. 5C&D. Newly formed bone was highlighted in pink in the top view of the reconstructed images, providing a clear visual assessment of the regenerative capacity imparted by the PSP-GelMA treatment compared to the other groups. Notably, the PSP-treated group exhibited a more substantial amount of newly formed bone tissue at both 4- and 8-weeks post-implantation. The semi-quantitative analysis of Micro-CT (Fig. 5E) confirmed that there are no statistically significant differences in bone volume (BV), bone surface/bone volume (BS/BV), and bone mineral density (BMD) between the blank group and GelMA group. Importantly, the PSP-GelMA group exhibited the highest BV among all groups studied. Additionally, the PSP-GelMA group demonstrated a decrease in BS/BV and an increase in BMD compared to both the blank and GelMA groups, although no significant differences in BS/BV and BMD were observed when compared to the KLT-GelMA group. In histological analysis, HE staining along with Masson's trichrome staining were employed to evaluate new bone formation. Illustrated in Fig. 5F–I, the PSP-GelMA group displayed a greater amount of new bone formation than the KLT-GelMA group, in contrast to the minimal bone formation observed in the blank and GelMA groups. Additionally, the application of PSP-GelMA contributed to a remarkable increase in the expression of OCN compared to the other groups both after 4- and 8-weeks of the implantation (Fig. S3).

Type H vessels, characterized by their CD31 and Endomucin (EMCN)-positive endothelium, are crucial for orchestrating the concomitant processes of angiogenesis and osteogenesis [31]. As shown in Fig. 5H&I, there was a higher prevalence of CD31 and EMCN co-labeled type H vessels in the PSP-GelMA group compared to the other groups, indicating the vital role of PSP in fostering the synergistic interplay of vascular-osteo communication. It is worth noting that the number of type H vessels in the PSP-GelMA group at 8 weeks was not as significant as that at 4 weeks. This decline may be attributed to the recruitment and differentiation of angiogenic cells in the defect area during the initial several weeks of the repair phase, which leads to increased blood flow in the tissue repair area [32]. As bone regeneration advances into the remodeling phase, the initially elevated number of blood vessels begins to gradually decrease [33]. This transition is a natural part of the healing process, as the body optimizes the vascular network to support long-term stability and functionality of the regenerated tissue. Consequently, the observed reduction in type H vessels at 8 weeks likely reflects this adaptive response rather than a failure of the regenerative process. In summary, these findings suggested that PSP could effectively promote vascularized bone regeneration in mouse critical-sized cranial defect model, exhibiting even greater efficacy than KLT.

Furthermore, *in vivo* toxicity evaluation of PSP was conducted following 8 weeks post-implantation. As shown in Fig. S4A, HE staining revealed no notable differences in the histology of vital organs, including the heart, liver, spleen, lung, and kidney, across the experimental groups. Additionally, as shown in Fig. S4B, routine blood tests encompassing a range of parameters—white blood cells (WBC), red blood cells (RBC), hemoglobin (HGB), hematocrit (HTC), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), and platelets

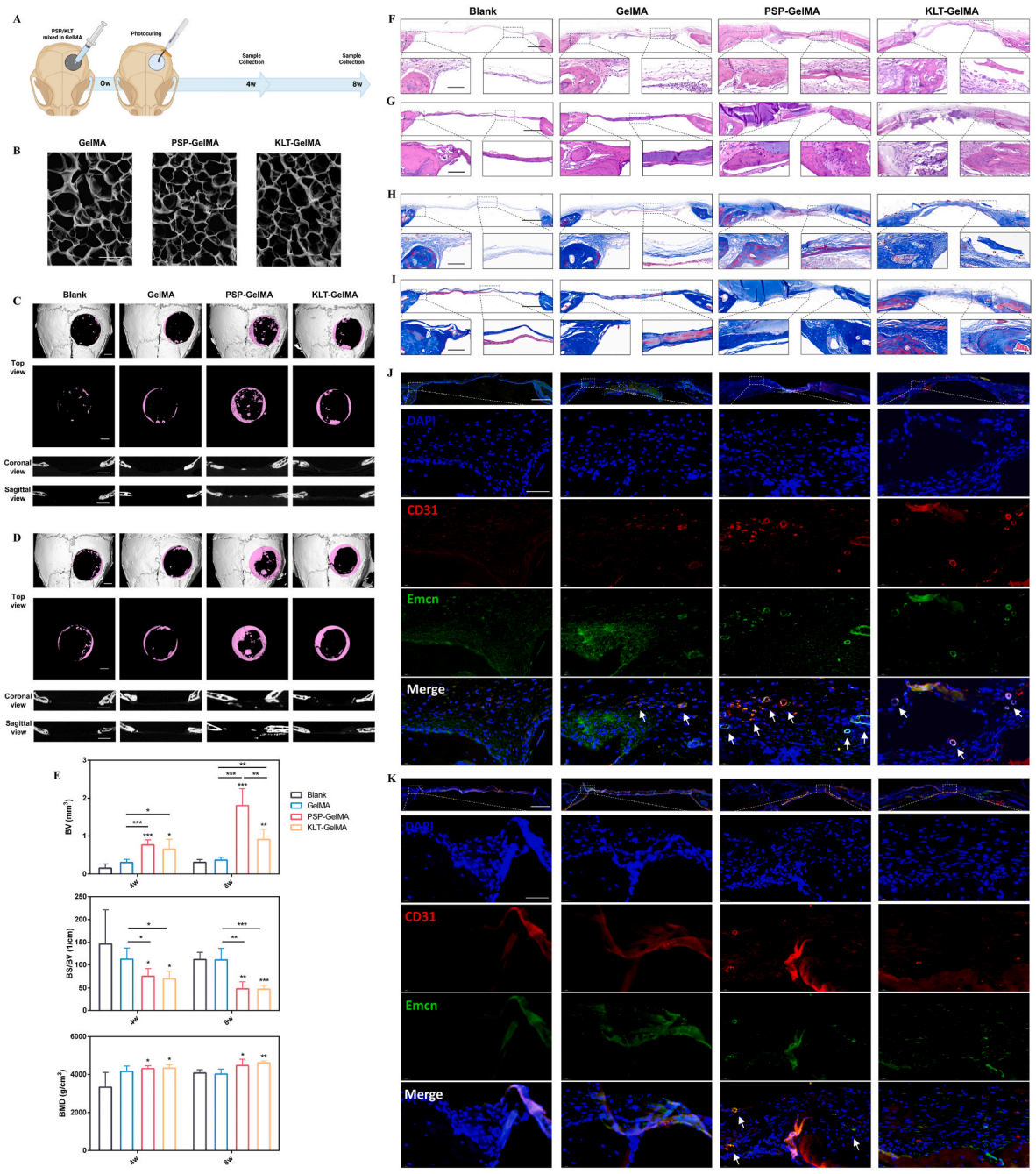

Fig. 5. DL and NLP-screened tripeptide PSP promoted vascularized bone regeneration in critical-sized cranial defect model. (A) Illustration of the experimental procedure; (B) Representative SEM images of GelMA, PSP-GelMA and KLT-GelMA hydrogels (scale bar, 20 μm); Micro-CT reconstruction from top view, coronal view and sagittal view 4-weeks (C) and 8-weeks (D) after implantation (scale bar, 1 mm); (E) Semi-quantitative analysis of Micro-CT; Representative images of HE staining 4-weeks (F) and 8-weeks (G) after implantation (scale bar, low magnification 500 μm/high magnification 100 μm); Representative images of Masson's trichrome staining 4-weeks (H) and 8-weeks (I) after implantation (scale bar, low magnification 500 μm/high magnification 100 μm); Representative images of immunofluorescence staining of CD31 and EMCN 4-weeks (J) and 8-weeks (K) after implantation (DAPI in blue, CD31 in red, EMCN in green, scale bar, low magnification 500 μm/high magnification 50 μm, arrow indicated CD31 and EMCN co-labeled type H vessels). Data are presented as means ± SD. Significant difference with respect to the control groups (*), *P < 0.05, **P < 0.01, ***P < 0.001.

(PLT)—showed no statistically significant variations, indicating the absence of systemic toxicity attributable to PSP.

3.6. sEVs derived from ECs after PSP treatment (PSP-sEVs) promoted osteogenesis effectively

Building on the research conducted, we discovered that PSP significantly augments the osteogenic differentiation of BMSCs through an angiogenesis-driven way. Furthermore, PSP was found to play a pivotal role in stimulating the synergistic interaction between angiogenesis and osteogenesis, thereby fostering bone regeneration. Considering the essential role of sEVs in cellular communication, we hypothesized that PSP may induce ECs to secrete distinct sEVs to exert a unique regulatory effect on the vascular-osteo communication. Hence, we proceeded to isolate sEVs from ECs (ECs-sEVs) treated with PSP (termed PSP-sEVs) and investigated their impact on the osteogenic differentiation of BMSCs. Fig. 6A illustrated the procedure of PSP-sEVs isolation and extraction, with sEVs derived from ECs treated with a standard vehicle

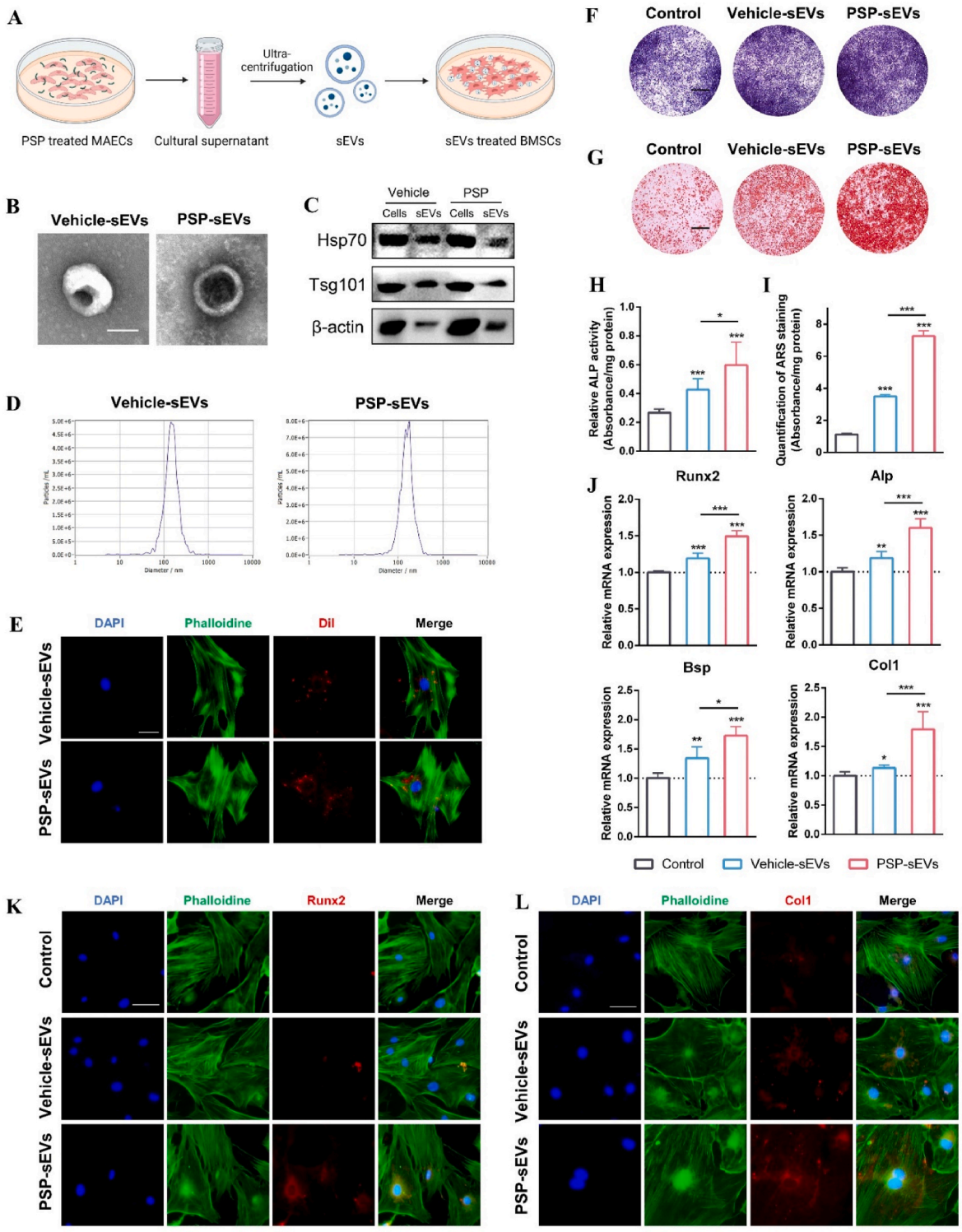

Fig. 6. sEVs derived from ECs after PSP treatment (PSP-sEVs) promoted osteogenesis effectively. (A) Illustration of the experimental process of PSP-sEVs isolation and extraction; (B) Representative TEM images of ECs-sEVs (scale bar, 100 nm); (C) WB results of ECs-sEVs; (D) NTA analyses of ECs-sEVs; (E) Internalization of Dil-labeled sEVs after 8 h co-incubation with BMSCs (DAPI in blue, Phalloidin in green, sEVs in red, scale bar, 50 μm); (F) Representative images of ALP staining after 7 days culture (scale bar, 1 mm), and its semi-quantitative analysis (H); (G) Representative images of ARS staining after 21 days culture (scale bar, 1 mm), and its semi-quantitative analysis (I); (J) mRNA expression of osteogenic-related genes of BMSCs by qRT-PCR after 7 d culture; (K) Representative images of immunofluorescence staining of Runx2 (DAPI in blue, Phalloidin in green, Runx2 in red, scale bar, 50 μm); (L) Representative images of immunofluorescence staining of Col1 (DAPI in blue, Phalloidin in green, Col1 in red, scale bar, 50 μm); Data are presented as means ± SD. Significant difference with respect to the control groups (*), *P < 0.05, **P < 0.01, ***P < 0.001.

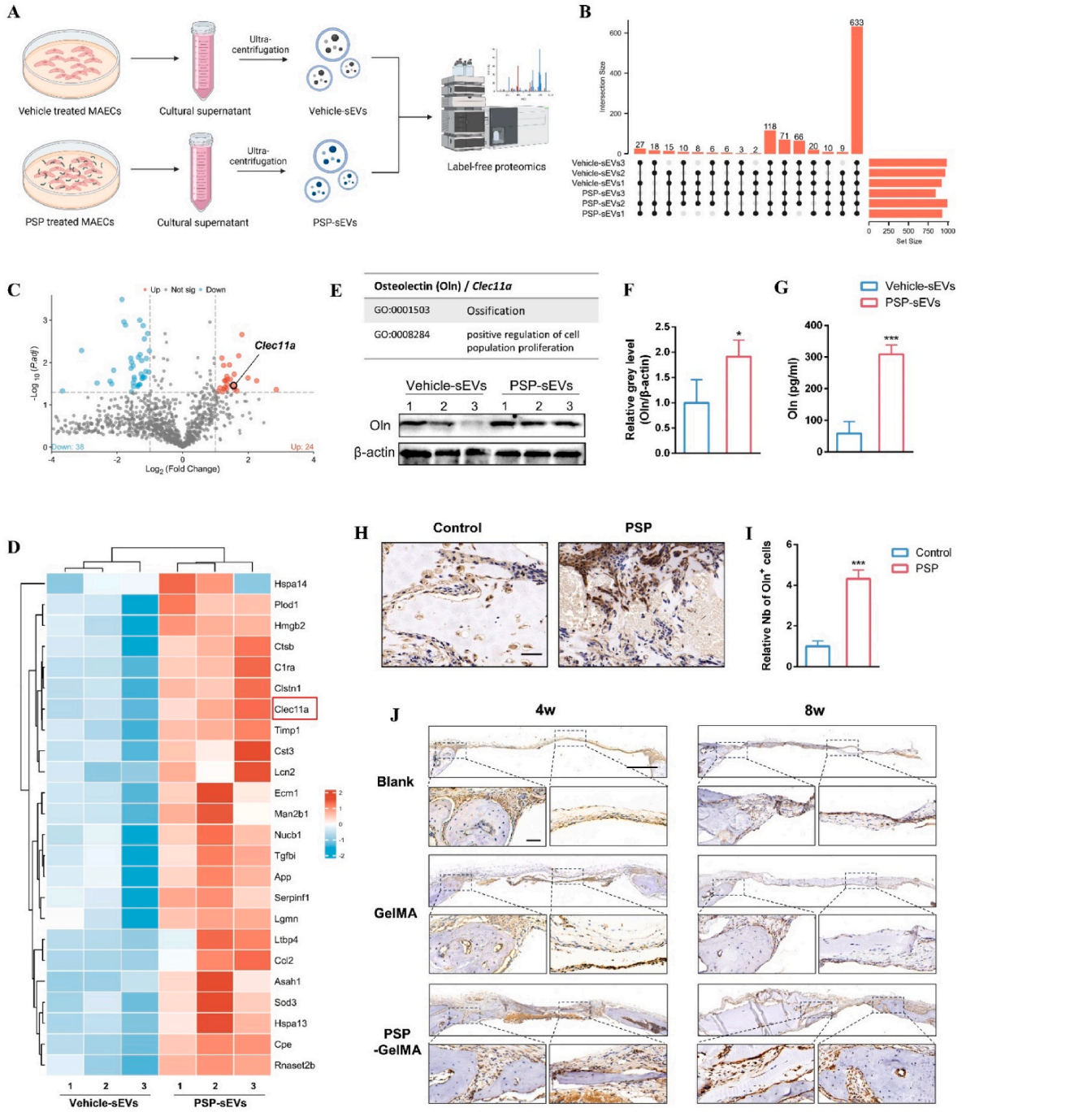

Fig. 7. Osteolectin was overexpressed in PSP-sEVs and after PSP treatment. (A) Illustration of the experimental procedure of Label-free quantitative proteomics of ECs-sEVs. (B) Upset graph of Label-free quantitative proteomics; (C) Volcano diagram of differentially expressed proteins; (D) Heatmap of upregulated proteins in PSP-sEVs; (E) WB analysis of Oln in ECs-sEVs; (F) Semi-quantitative analysis of Oln according to the WB; (G) Elisa analysis of Oln in ECs-sEVs; (H) Respective images of immunohistochemical staining of Oln and its semi-quantitative analysis (I) in the Matrigel plug assay (scale bar, 50 μm); (J) Respective images of immunohistochemical staining of Oln in the critical-sized cranial model (scale bar, low magnification 500 μm/high magnification 100 μm); Data are presented as means ± SD. Significant difference with respect to the control groups (*), *P < 0.05, **P < 0.01, ***P < 0.001.

(termed Vehicle-sEVs) serving as the control. Fig. 6B presents the TEM images, revealing that the ECs-sEVs exhibit the characteristic cup-shaped or spherical nanoparticle morphology. The presence of established sEVs markers, Hsp70 and Tsg101, was confirmed in the sEVs as illustrated in Fig. 6C. For concentration and size distribution analysis, the ECs-sEVs were suspended in an aqueous solution and subjected NTA. This analysis showed that the average diameters of the Vehicle-sEVs and PSP-sEVs were 156.8 nm and 163.1 nm, respectively, as detailed in Fig. 6D. To investigate the uptake of ECs-derived sEVs by BMSCs, DiI-labeled sEVs were co-cultured with BMSCs for 8 h. Subsequent imaging revealed red fluorescence in proximity to the cellular nuclei as depicted in Fig. 6E, indicating successful internalization of the ECs-sEVs into the target cells, thereby setting the stage for further investigative analysis.

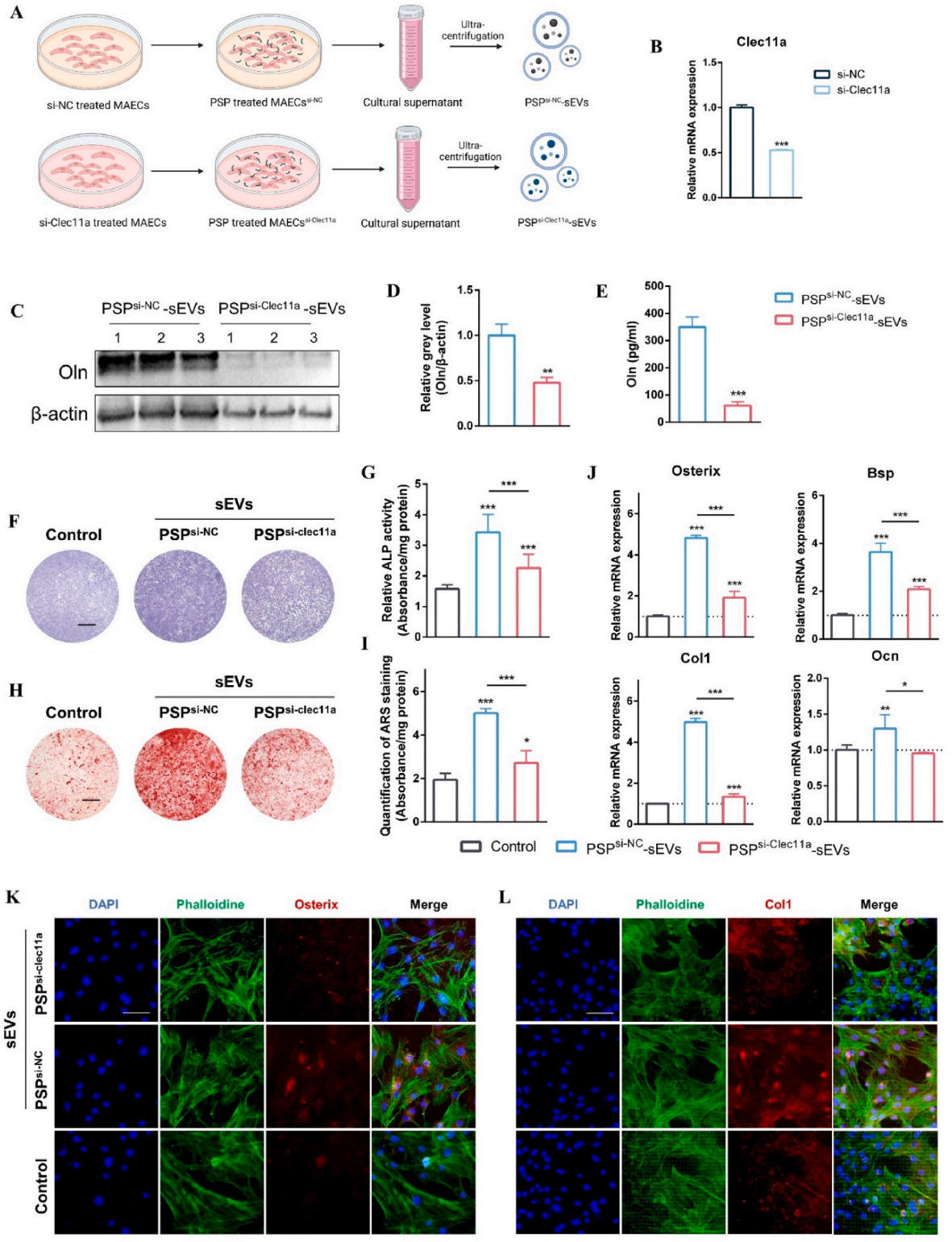

Fig. 8. Enhanced osteoinductive capacity of PSP-sEVs was attributed to the overexpressed Osteolectin. (A) Illustration of the experimental procedure; (B) qRT-PCR analysis of MAECs after si-Clec11a transfection; (C) WB analysis of Oln in sEVs; (D) semi-quantitative analysis of Oln according to the WB; (E) Elisa analysis of Oln in sEVs; (F) Representative images of ALP staining after 7 days culture (scale bar, 1 mm), and its semi-quantitative analysis (G); (H) Representative images of ARS staining after 21 days culture (scale bar, 1 mm), and its semi-quantitative analysis (I); (J) mRNA expression of osteogenic-related genes of BMSCs by qRT-PCR after 7 d culture; (K) Representative images of immunofluorescence staining of Osterix (DAPI in blue, Phalloidin in green, Osterix in red, scale bar, 100 μm); (L) Representative images of immunofluorescence staining of Col1 (DAPI in blue, Phalloidin in green, Col1 in red, scale bar, 100 μm); Data are presented as means ± SD. Significant difference with respect to the control groups (*), *P < 0.05, **P < 0.01, ***P < 0.001.

To explore the osteogenic effect of PSP-sEVs, BMSCs were exposed to PBS, 50 μg/ml Vehicle-sEVs or PSP-sEVs. Fig. 6F&H presented the ALP staining and semi-quantitative analysis of ALP activity respectively after one-week treatment of ECs-sEVs. The findings suggested that ECs-sEVs promote early osteogenic differentiation of BMSCs, with PSP-sEVs demonstrating a significantly enhanced osteogenic effect relative to Vehicle-sEVs. Furthermore, Fig. 6G&I displayed the ARS staining and its semi-quantitative analysis following a 21-day period of ECs-sEVs treatment, indicating that PSP-sEVs markedly augmented osteogenic mineralization in BMSCs when compared to both the control and Vehicle-sEVs groups. As evidenced in Fig. 6J, treatment with ECs-sEVs substantially increased the expression of osteogenic markers including *Runx2*, *Alp*, *Bsp*, and *Col1*. PSP-sEVs, in particular, presented a remarkable increase in osteogenic activity over Vehicle-sEVs. This trend was further corroborated by the immunofluorescent staining for Runx2 and Col1, as illustrated in Fig. 6K&L, which also reflected the superior osteogenic influence of PSP-sEVs. The aforementioned studies have demonstrated that ECs-sEVs are capable of inducing the osteogenic differentiation of BMSCs. Crucially, PSP-sEVs have been found to have more robust bone-promoting effects compared to Vehicle-sEVs, suggesting PSP as a potent enhancer of the osteogenic effects mediated by ECs-sEVs.

3.7. Osteolectin was overexpressed in PSP-sEVs and after PSP treatment

To delve deeper into the mechanisms of PSP-sEVs amplifying ECs-EVs' osteoinductive effects, we employed Label-free quantitative proteomics on Vehicle-sEVs and PSP-sEVs to obtain the differential changes in protein expression in the comparative group. Fig. 7A illustrates the experimental procedure. The Upset graph in Fig. 7B showed the quantity of protein expression in sEVs for each group, as well as the intersection and number of intersections in the set. By applying the criteria of fold change (FC) greater than 2 and a P value less than 0.05 (T-test or other), we discovered that there were 24 upregulated and 38 downregulated proteins in PSP-sEVs compared to Vehicle-sEVs (Fig. 7C). In Fig. 7D, a heatmap was presented, showing the expression levels of the 24 upregulated proteins in PSP-sEVs, along with their corresponding gene names. Notably, we observed that Oln/*Clec11a* exhibited significantly higher expression in PSP-sEVs compared to Vehicle-sEVs. Oln/*Clec11a* has been proven to be involved in the biological processes of "Ossification" and "positive regulation of cell population proliferation". Fig. 7E&F illustrated the WB and its semi-quantitative analysis of Oln expression in ECs-sEVs. The results indicated that Oln was significantly upregulated in PSP-sEVs compared to Vehicle-sEVs. This finding was further supported by the Elisa analysis in Fig. 7G, which also revealed the same trend of increased Oln expression in PSP-sEVs.

Furthermore, we examined the expression of Oln in animal models after administering PSP treatment. As shown in Fig. 7H&I, in the Matrigel plug assay, we performed immunohistochemical staining of Oln and its semi-quantitative analysis, which revealed a significant increase in the expression of Oln following PSP treatment. Additionally, in the critical-sized cranial model, the application of PSP-GelMA resulted in a notable elevation in the expression of Oln compared to the blank and GelMA groups both after 4- and 8-weeks of the implantation (Fig. 7J). After 8 weeks of implantation, the control and GelMA groups exhibited minimal expression of Oln, while the PSP group maintained a significant level of Oln expression. These findings indicated that PSP treatment exerts a significant increase on the expression of Oln *in vivo*. Overall, Oln was found to be overexpressed in PSP-sEVs and after PSP treatment, suggesting that Oln may exert a pivotal role in the enhanced osteoinductive effects of PSP-sEVs.

3.8. Enhanced osteoinductive capacity of PSP-sEVs was highly attributed to the overexpressed Osteolectin

We then turned our attention to the role of Oln in the osteoinductive capacity of PSP-sEVs. By deploying miRNA to silence Clec11a in MAECs, we managed to decrease Oln levels. The subsequent extraction of sEVs after PSP stimulation (to form $PSP^{si\text{-}clec11a}$-sEVs) would then provide a population of sEVs with reduced Oln content. We compared the osteoinductive ability of these $PSP^{si\text{-}clec11a}$-sEVs with that of $PSP^{s\text{-}NC}$-sEVs, helping in understanding whether Oln is indeed the key factor for the enhanced osteoinductive ability of PSP-sEVs. Fig. 8A illustrates the experimental procedure. Fig. 8B and Fig. S5 clarified the downregulated gene and protein expression of *Clec11a* in MAECs after si-Clec11a transfection respectively. Fig. S6A showcased the TEM images of $PSP^{s\text{-}NC}$-sEVs and $PSP^{si\text{-}clec11a}$-sEVs. The existence of recognized sEV markers, Hsp70 and Tsg101, was ratified in the sEVs, as depicted in Fig. S6B. For concentration and size distribution analysis, the sEVs were subjected to NTA. This examination revealed that the average diameters of the $PSP^{si\text{-}NC}$-sEVs and $PSP^{si\text{-}clec11a}$-sEVs were 181.4 nm and 178.7 nm respectively, as elaborated in Fig. S6C. As shown in Fig. 8C&D, WB and its semi-quantitative analysis indicated that Oln was significantly downregulated in $PSP^{si\text{-}clec11a}$-sEVs compared to $PSP^{s\text{-}NC}$-sEVs. This finding was further supported by the Elisa analysis in Fig. 8E, which also revealed the same trend of decreased Oln expression in $PSP^{si\text{-}clec11a}$-sEVs. Fig. 8F and G illustrated the ALP staining and semi-quantitative analysis of ALP activity respectively after a week of exposure to the above sEVs. The outcomes suggested that $PSP^{si\text{-}clec11a}$-sEVs considerably diminish the early osteogenic differentiation of BMSCs in contrast to $PSP^{si\text{-}NC}$-sEVs. Moreover, Fig. 8H and I displays the ARS staining and its semi-quantitative analysis after a 21-day cultivation period, unveiling that $PSP^{si\text{-}clec11a}$-sEVs significantly hinder osteogenic mineralization in BMSCs compared to $PSP^{s\text{-}NC}$-sEVs. As demonstrated in Fig. 6J, treatment with $PSP^{s\text{-}NC}$-sEVs notably amplified the expression of osteogenic markers *Osterix*, *Bsp*, *Col1*, and *Ocn*, while $PSP^{si\text{-}clec11a}$-sEVs substantially reduced the expression of the corresponding genes relative to $PSP^{s\text{-}NC}$-sEVs. Furthermore, immunofluorescent staining images depicted in Fig. 8K and L revealed that $PSP^{si\text{-}clec11a}$-sEVs treatment significantly degraded the protein expression of Osterix and Col1 when compared to $PSP^{s\text{-}NC}$-sEVs. Overall, the osteogenic potential of $PSP^{si\text{-}clec11a}$-sEVs is markedly inferior to that of $PSP^{s\text{-}NC}$-sEVs, suggesting that Oln indeed plays an indispensable role in the osteoinductive ability of PSP-sEVs.

4. Discussion

Here, we presented a DL and NLP composite model to successfully screen angiogenic oligopeptides from the IDRs of the related proteins. IDRs are regions in protein sequences that have remarkable structural flexibility. These regions are abundant in the eukaryotic proteome and have been recognized as crucial regulators of protein functionality [20]. Consequently, IDRs have emerged as a hotbed of discovery for bioactive peptides. By screening and analyzing IDRs, novel bioactive peptides can be uncovered, offering valuable insights for the development of new therapeutic strategies and drugs. Ma et al. [34] evaluated all 12 oligopeptide segments within IDRs of the conserved domains of MHC class II protein. Among the oligopeptides examined, they successfully discovered a specific bioactive peptide called QSHGPS, which demonstrated the ability to promote angiogenesis and accelerate wound healing. This discovery highlighted the feasibility of identifying functional oligopeptides within IDRs.

However, the exhaustive method proves beneficial in the selection of peptides that meet established criteria when dealing with a limited quantity. When the peptide count escalates substantially, this approach becomes inherently inefficient, necessitating an excessive expenditure of time, effort, and resources. Such a scenario also exacerbates the probability of encountering errors, making it a less optimal strategy for handling large-scale peptide selection. As a result, there is a pressing need to explore more convenient, efficient, and accurate screening methods, allowing for a more streamlined and cost-effective process in the identification of functional oligopeptides within IDRs. One potential approach is the utilization of AI algorithms, which have the capability to

deduce hidden structures from unlabeled data. By applying AI models, researchers could identify and gather subsets, trends, or clusters within IDRs that may exert biological activities, which provide a more targeted and efficient method for the identification of potential therapeutic peptides.

By employing our DL and NLP composite model with the combination of *Word2vec* and *TF-IIDF*, potential pro-angiogenesis oligopeptides were identified within IDRs of 262 proteins annotated with the GO terms "vasculogenesis" and "angiogenesis". *Word2vec*, a method for generating word embeddings, creates vector representations that enable words with similar meanings to have analogous representations [35]. In the context of our research, *Word2vec* was employed to identify the patterns, trends, and specific features of AAs inherent in the IDRs, subsequently excluding those oligopeptide sequences containing consecutive correlated AAs. When compared to the *N-gram* model proposed by previous researchers [27], our *Word2vec* model demonstrates superior intelligence, precision, efficiency, compactness, and it better harnesses the relationships amidst neighboring AA sequences. Primarily, *Word2vec* is capable of providing a better representation for previously unseen data. Moreover, *Word2vec* employs continuous distributed representations to depict the context of a word, leading to a more efficient contextual representation. In contrast, *N-gram* model use discrete, sparse representations, which can be less efficient and expressive. Furthermore, *Word2vec* accomplishes dimensionality reduction. The size of an *N-gram* model escalates exponentially with the increase of 'n', whereas the size of a *Word2vec* model solely relies on the chosen size of the vectors. More importantly, *Word2vec* captures semantic information and relationships between AAs, as it represents words in a high-dimensional space where semantically similar words are placed closer together, while the *N-gram* model merely considers the frequency of AA sequences and does not inherently capture semantic relationships. Hence, the *Word2vec* model outperforms the *N-gram* model in downstream NLP tasks, making it a more adept choice for IDRs processing.

Secondly, we utilized *TF-IIDF*, a modified *TF-IDF* model which commonly used for text mining [36], to evaluate the significance of potential angiogenic oligopeptides within the IDRs. It consists of two parts: *TF* and *IIDF*. *TF* quantified the number of occurrences of an oligopeptide segment within the IDRs. The more frequently the oligopeptide appears, the higher its *TF* value. *IIDF*, inverse to *IDF*, indicated the number of IDRs that contain the potential angiogenic oligopeptides, emphasizing the distribution of oligopeptide segments in the IDRs. The *TF-IIDF* value was calculated by multiplying these two metrics. It increases proportionally to the number of times an oligopeptide appears and the number of different IDRs segments in which the oligopeptide can be found. A higher *TF-IIDF* score implied greater importance of that specific oligopeptide in the IDRs. Taken together the power of *Word2vec* and *TF-IIDF*, the AI model was able to uncover hidden insights within the IDRs data and make predictions about the angiogenic properties of the oligopeptides.

Based on the tripeptides ranking among the top according to the *TF-IIDF* values and the potential benefits of shorter AA chains for bioactive peptides—such as smaller the molecular weight, lower the synthesis cost, easier to transport, modify, and recombine [29,30], we conducted a qRT-PCR analysis of angiogenic gene expression for the top five oligopeptides — SPP, SSP, SSG, PPS, and PSP in the first batch of experiments. The KLT peptide, also referred to as the QK peptide, is known for its VEGF-mimicking properties and has been extensively validated as an angiogenesis promoter both *in vivo* and *in vitro* [28]. For our experimental framework, the KLT peptide was employed as a benchmark positive control, a randomly generated peptide sequence served as the negative control. Our findings revealed that four of five tripeptides successfully enhanced angiogenic gene expression to some extent. Notably, the PSP sequence outperformed the others, exhibiting the most potent effect in stimulating this expression, even surpassing that of the KLT peptide. Consequently, the PSP sequence was selected for more in-depth investigation based on its promising efficacy. Upon analyzing the migration and differentiation of ECs, alongside the results from the CAM assay and the Matrigel plug assay, it was evident that PSP significantly promoted angiogenesis *in vitro*, *ex vivo*, and *in vivo*. Delving deeper into the biological functions of the PSP, our research uncovered that it not only directly stimulates osteogenic differentiation of BMSCs but also enhances this process through the mediation of angiogenesis. Furthermore, when integrated into GelMA hydrogels, the tripeptide PSP exhibited a remarkable proficiency in repairing mouse critical-sized cranial defects. Given the effectiveness of PSP in angiogenesis and osteogenesis, its potential application extends to be cooperated with a variety of reparative materials across a broader spectrum of tissue regeneration scenarios in the future.

Angiogenesis and osteogenesis are intricately coupled in bone [37]. sEVs, rich in proteins and microRNAs, are 30–200 nm nanoscale lipid bilayer-enclosed structures secreted by diverse cell types. They play an indispensable role in maintaining and facilitating vascular-osteo communication. Evidence has established that sEVs derived from BMSCs possess the ability to facilitate angiogenesis [38]. Expanding upon this knowledge, our study contributed the missing piece regarding the function of sEVs in vascular-osteo interactions. We have discovered, for the first time, that sEVs derived from ECs (ECs-sEVs) can, in turn, promote the osteogenic differentiation of BMSCs. Additionally, we found that the DL and NLP-screened tripeptide PSP could enhanced the angiogenesis-mediated osteogenesis of BMSCs. Notably, sEVs from ECs treated with PSP (PSP-sEVs) significantly bolster the osteogenic differentiation of BMSCs, outperforming those from untreated ECs (Vehicle-sEVs). To elucidate the mechanisms underpinning the efficacy of PSP-sEVs, we turned to label-free proteomics and discovered that PSP treatment stimulates ECs to secrete sEVs enriched with Osteolectin.

Osteolectin (Oln), also known as the C-type lectin domain family 11, member A (Clec11a), or Stem Cell Growth Factor (Scgf), is a protein with a C-type lectin domain that is secreted by various cells including matrix cells, osteoblasts, osteocytes, and hypertrophic chondrocytes [39]. It has garnered attention as an osteogenic growth factor crucial for maintaining adult bone mass [39–41]. Yue et al. [39] observed that mice lacking Clec11a suffer from accelerated bone loss, reduced bone strength, and impaired fracture healing with age, along with diminished osteogenic differentiation of BMSCs. Conversely, recombinant Clec11a was shown to promote osteogenic differentiation of BMSCs into mature osteoblasts and potentially counteract osteoporosis in mice. Shen et al. [40] further investigated the mechanism through which Oln promotes osteogenesis, revealing that Oln binds to Integrin α11, subsequently activating Wnt pathway and facilitating osteogenic differentiation of BMSCs. Zhang et al. [41] expanded on this by demonstrating that parathyroid hormone (PTH) enhances Oln expression, with Oln mediating part of PTH's effect on bone formation, underscoring their synergistic relationship. Additionally, Zhang et al. [42] identified Oln as a regulator of bone elongation and body length in both mice and humans, with its absence resulting in shorter bone development. Research also shown that Oln deficiency markedly exacerbates the progression of osteoarthritis, while the administration of Oln into joints can mitigate the advancement of this disease [43]. Shen et al. [44] highlighted that physical exercise stimulates Oln production from bone marrow cells in mice, which results in increased bone formation, enhanced immune responses, and a boost in the production of immune progenitor cells, aiding bacterial clearance after fracture. Given its wide-ranging effects, Oln and Oln-positive cells are seen as promising agent for bone regeneration to address various clinical skeletal challenges, including osteoporosis, delayed or non-union fracture healing, osteoarthritis, and associated infections [45]. Yet the synthetic production of Oln poses challenges similar to those encountered with other cytokines, such as high costs, potential antigenicity, and tumorigenic risks [46,47]. Our study introduced a novel tripeptide PSP, discovered through our DL and NLP composite model, which overcomes these drawbacks. PSP offers several advantages: it has a low molecular weight, is less costly to synthesize, and is simpler to transport, modify, and recombine. Importantly,

PSP has a 'priming' effect, activating the body's innate ability to produce endogenous Oln and prompting ECs to release Oln-rich sEVs.

Hu et al. [48] have found that human umbilical cord mesenchymal stromal cell-derived sEVs (hucMSC-EVs) are particularly rich in Oln. These hucMSC-EVs have been shown to enhance the osteogenic differentiation of BMSCs and suppress the formation of osteoclasts, thereby presenting a promising strategy for both the prevention and treatment of osteoporosis. In our research, we initially discovered the presence of Oln in sEVs derived from ECs. Furthermore, our DL and NLP-screened tripeptide PSP was observed to enhance the secretion of sEVs containing a higher abundance of Oln by ECs. These PSP-sEVs enriched with Oln exhibit dramatically improved capabilities in promoting osteogenesis. We attenuated Oln expression in in MAECs by leveraging siRNA to target and suppress the *Clec11a* gene. The subsequent extraction of sEVs after PSP stimulation ($PSP^{si\text{-}clec11a}$-sEVs) would then provide a population of sEVs with reduced Oln content. We observed a significant reduction in the osteogenic capabilities of $PSP^{si\text{-}clec11a}$-sEVs comparted to that of $PSP^{si\text{-}NC}$-sEVs, suggesting that the enhanced osteoinductive capacity of PSP-sEVs is attributed to the overexpressed Osteolectin.

5. Conclusion

In this study, we introduced a composite AI model that merges DL and NLP methodologies, integrating the *Word2vec* and *TF-IIDF* models. This fusion enabled us to identify a range of potent bioactive peptides from the IDRs of 262 angiogenic proteins. Recognizing that shorter AA sequences often possess superior inherent advantages, we distinguished the tripeptide PSP for its extraordinary angiogenic properties, as determined by qRT-PCR analysis. Our *in vitro* and *in vivo* experiments confirmed that PSP can stimulate ECs' vascularization, enhance vascular-osteo communication, and ultimately boost the osteogenic differentiation of BMSCs. When embedded into GelMA, PSP demonstrated a notable proficiency in repairing critical-sized cranial defects. Moreover, we discovered that the sEVs secreted by ECs post-PSP treatment (PSP-sEVs) effectively amplify the osteoinductive potential of EC-sEVs. Employing label-free proteomics analysis, we found that PSP induces ECs to secrete sEVs that are notably rich in Osteolectin, thereby contributing to the heightened osteoinductive capacity of PSP-sEVs. Our research demonstrated the effectiveness and efficiency of the DL and NLP composite model in pinpointing angiogenic peptides, providing an innovative strategy for bone defect treatment based on the PSP therapy. Furthermore, this research paved the way for the development of various peptide-based therapeutic strategies applicable to a broader range of diseases.

CRediT authorship contribution statement

Yu Chen: Writing – original draft, Visualization, Methodology, Conceptualization. **Long Chen:** Writing – original draft, Methodology. **Jinyang Wu:** Supervision. **Xiaofeng Xu:** Writing – review & editing. **Chengshuai Yang:** Writing – review & editing. **Yong Zhang:** Supervision. **Xinrong Chen:** Methodology, Conceptualization. **Kaili Lin:** Methodology, Conceptualization. **Shilei Zhang:** Supervision.

Ethics approval and consent to participate

All animal experiments in this study obtained approval from the Laboratory Animal Ethics Committee of the Ninth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine (Approval No: SH9H-2023-A924-1).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Natural Science Fundation of Shanghai, Shanghai Action Plan for Science, Technology and innovation (24ZR1443700), Fundamental Research Funds for the Central Universities (YG2023LC05), SJTU Trans-med Awards Research (20230104), the National Clinical Research Center for Oral Diseases (NCRCO202307 and NCRCO202106), Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine (JS22A03), Shanghai's Top Priority Research Center (2022ZZ01017), CAMS Innovation Fund for Medical Sciences (CIFMS) (2019-I2M-5-037), the Discipline Cluster Fund (XKQJS202401) from Shanghai Ninth People's Hospital Shanghai Jiao Tong University Schoolof Medicine, Project 21DZ2271700-1 from Shanghai Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bioactmat.2024.11.011.

References

[1] G.N. Duda, et al., The decisive early phase of bone regeneration, Nat. Rev. Rheumatol. 19 (2023) 78–95.
[2] A. Marrella, et al., Engineering vascularized and innervated bone biomaterials for improved skeletal tissue regeneration, Mater. Today 21 (2018) 362–376.
[3] S.K. Ramasamy, et al., Regulation of hematopoiesis and osteogenesis by blood vessel-derived signals, Annu. Rev. Cell Dev. Biol. 32 (2016) 649–675.
[4] J. Tuckermann, R.H. Adams, The endothelium-bone axis in development, homeostasis and bone and joint disease, Nat. Rev. Rheumatol. 17 (2021) 608–620.
[5] L. Yang, L. Fan, X. Lin, Y. Yu, Y. Zhao, Pearl powder hybrid bioactive scaffolds from microfluidic 3D printing for bone regeneration, Adv. Sci. (2023) e2304190.
[6] K.D. Hankenson, K. Gagne, M. Shaughnessy, Extracellular signaling molecules to promote fracture healing and bone regeneration, Adv. Drug Deliv. Rev. 94 (2015) 3–12.
[7] Z. Lv, et al., A MgFe-LDH nanosheet-incorporated smart thermo-responsive hydrogel with controllable growth factor releasing capability for bone regeneration, Adv. Mater. 35 (2023) e2206545.
[8] E.R. Oliveira, et al., Advances in growth factor delivery for bone tissue engineering, Int. J. Mol. Sci. 22 (2021).
[9] R. Li, et al., Synergistic osteogenic and angiogenic effects of KP and QK peptides incorporated with an injectable and self-healing hydrogel for efficient bone regeneration, Bioact. Mater. 18 (2022) 267–283.
[10] F. Gonzalez-Perez, M. Alonso, I. Gonzalez de Torre, M. Santos, J.C. Rodriguez-Cabello, Protease-sensitive, VEGF-mimetic peptide, and IKVAV laminin-derived peptide sequences within elastin-like recombinamer scaffolds provide spatiotemporally synchronized guidance of angiogenesis and neurogenesis, Adv. Healthcare Mater. 11 (2022) e2201646.
[11] C.C. Wu, L.C. Wang, Y.T. Su, W.Y. Wei, K.J. Tsai, Synthetic alpha5beta1 integrin ligand PHSRN is proangiogenic and neuroprotective in cerebral ischemic stroke, Biomaterials 185 (2018) 142–154.
[12] L. De Rosa, R. Di Stasi, L.D. D'Andrea, Pro-angiogenic peptides in biomedicine, Arch. Biochem. Biophys. 660 (2018) 72–86.
[13] J. Fetse, S. Kandel, U.F. Mamani, K. Cheng, Recent advances in the development of therapeutic peptides, Trends Pharmacol. Sci. 44 (2023) 425–441.
[14] M. Nomizu, et al., Cell binding sequences in mouse laminin alpha1 chain, J. Biol. Chem. 273 (1998) 32491–32499.
[15] M.I. Sadowski, D.T. Jones, The sequence-structure relationship and protein function prediction, Curr. Opin. Struct. Biol. 19 (2009) 357–362.
[16] A.K. Dunker, et al., Intrinsically disordered protein, J. Mol. Graph. Model. 19 (2001) 26–59.
[17] M. Guharoy, K. Pauwels, P. Tompa, SnapShot: intrinsic structural disorder, Cell 161 (2015) 1230, 1230 e1231.
[18] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, Nat. Rev. Mol. Cell Biol. 6 (2005) 197–208.
[19] P.E. Wright, H.J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation, Nat. Rev. Mol. Cell Biol. 16 (2015) 18–29.
[20] A.S. Holehouse, B.B. Kragelund, The molecular basis for cellular function of intrinsically disordered protein regions, Nat. Rev. Mol. Cell Biol. 25 (2024) 187–211.
[21] A. Patil, et al., A disordered region controls cBAF activity via condensation and partner recruitment, Cell 186 (2023) 4936–4955, e4926.
[22] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature 559 (2018) 547–555.
[23] W.P. Walters, R. Barzilay, Applications of deep learning in molecule generation and molecular property prediction, Acc. Chem. Res. 54 (2021) 263–270.
[24] M.W. Mullowney, et al., Artificial intelligence for natural product drug discovery, Nat. Rev. Drug Discov. 22 (2023) 895–916.

[25] J. Hirschberg, C. Manning, Advances in natural language processing, Science (New York, N.Y.) 349 (2015) 261–266.
[26] D. Miller, A. Stern, D. Burstein, Deciphering microbial gene function using natural language processing, Nat. Commun. 13 (2022) 5731.
[27] M. Cai, et al., Generation of functional oligopeptides that promote osteogenesis based on unsupervised deep learning of protein IDRs, Bone Res 10 (2022) 23.
[28] D. Zhao, et al., Tetrahedral framework nucleic acid carrying angiogenic peptide prevents bisphosphonate-related osteonecrosis of the jaw by promoting angiogenesis, Int. J. Oral Sci. 14 (2022) 23.
[29] R. Visser, G.A. Rico-Llanos, H. Pulkkinen, J. Becerra, Peptides for bone tissue engineering, J. Contr. Release 244 (2016) 122–135.
[30] I.W. Hamley, Small bioactive peptides for biomaterials design and therapeutics, Chem. Rev. 117 (2017) 14015–14041.
[31] A.P. Kusumbe, S.K. Ramasamy, R.H. Adams, Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone, Nature 507 (2014) 323–328.
[32] Y. Peng, S. Wu, Y. Li, J.L. Crane, Type H blood vessels in bone modeling and remodeling, Theranostics 10 (2020) 426–436.
[33] T.A. Einhorn, L.C. Gerstenfeld, Fracture healing: mechanisms and interventions, Nat. Rev. Rheumatol. 11 (2015) 45–54.
[34] S. Ma, et al., Modification of the small intestinal submucosa membrane with oligopeptides screened from intrinsically disordered regions to promote angiogenesis and accelerate wound healing, Biomater. Adv. 148 (2023) 213360.
[35] W.-t. Y. Tomas Mikolov, Geoffrey Zweig, Linguistic regularities in continuous space word representations, in: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Atlanta, Georgia, 2013, pp. 746–751.
[36] G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval, Inf. Process. Manag. 24 (1988) 513–523.
[37] S.K. Ramasamy, A.P. Kusumbe, L. Wang, R.H. Adams, Endothelial Notch activity promotes angiogenesis and osteogenesis in bone, Nature 507 (2014) 376–380.
[38] Y. Zhang, et al., Umbilical mesenchymal stem cell-derived exosome-encapsulated hydrogels accelerate bone repair by enhancing angiogenesis, ACS Appl. Mater. Interfaces 13 (2021) 18472–18487.
[39] R. Yue, B. Shen, S.J. Morrison, Clec11a/osteolectin is an osteogenic growth factor that promotes the maintenance of the adult skeleton, Elife 5 (2016).
[40] B. Shen, et al., Integrin alpha11 is an Osteolectin receptor and is required for the maintenance of adult skeletal bone mass, Elife 8 (2019).
[41] J. Zhang, et al., The effect of parathyroid hormone on osteogenesis is mediated partly by osteolectin, Proc. Natl. Acad. Sci. U.S.A. 118 (2021).
[42] J. Zhang, et al., Osteolectin increases bone elongation and body length by promoting growth plate chondrocyte proliferation, Proc. Natl. Acad. Sci. U.S.A. 120 (2023) e2220159120.
[43] A. Fan, et al., Inhibition of fibroblast activation protein ameliorates cartilage matrix degradation and osteoarthritis progression, Bone Res 11 (2023) 3.
[44] B. Shen, et al., A mechanosensitive peri-arteriolar niche for osteogenesis and lymphopoiesis, Nature 591 (2021) 438–444.
[45] S. Ehnert, T. Histing, A.K. Nussler, Osteolectin(+) stromal cells: mechanical stimulation improves bone regeneration and supports bacterial clearance after fracture, Signal Transduct. Targeted Ther. 6 (2021) 257.
[46] R.A. Saxton, C.R. Glassman, K.C. Garcia, Emerging principles of cytokine pharmacology and therapeutics, Nat. Rev. Drug Discov. 22 (2023) 21–37.
[47] T. Yao, et al., Thiol-ene conjugation of VEGF peptide to electrospun scaffolds as potential application for angiogenesis, Bioact. Mater. 20 (2023) 306–317.
[48] Y. Hu, et al., Human umbilical cord mesenchymal stromal cells-derived extracellular vesicles exert potent bone protective effects by CLEC11A-mediated regulation of bone metabolism, Theranostics 10 (2020) 2293–2308.