Osteoarthritis and Cartilage 34 (2026) 1231–1242

Osteoarthritis and Cartilage

SaintGSE: Transformer-based efficient and explainable gene set enrichment analysis

Min-Seung Jeon [a 1], Jiho Nam [b 1], Minseok Lee [a], Chanmi Cho [c d], Siyoung Yang [b *], Seong-il Eyun [a *]

[a] Department of Life Science, Chung-Ang University, Seoul 06974, South Korea
[b] Department of Biological Sciences, Sungkyunkwan University, Suwon 16419, Korea
[c] Center for Systems Biology, Harvard Medical School, Boston, MA 02114, USA
[d] Department of Molecular Medicine, Hanyang University-ERICA, Ansan 15588, Korea

ARTICLE INFO

Article history:
Received 17 June 2025
Accepted 9 April 2026

Keywords:
Self-Attention and Intersample Attention Transformer (SAINT)
eXplainable Artificial Intelligence (XAI)
Signaling pathway
Osteoarthritis (OA)
Natural product

ABSTRACT

Objective: Gene set enrichment analysis (GSEA) is widely used to interpret genome-wide expression data, but pathway inference from differentially expressed gene (DEG) signatures remains limited by data scarcity, model incompatibility with tabular data, and limited interpretability. We aimed to develop an efficient and explainable model that predict gene-pathway associations from a large curated DEG signature compendium and supports potential therapeutic target discovery in complex diseases such as osteoarthritis (OA) and other disease contexts.

Method: We developed SaintGSE, a supervised prediction framework combining an autoencoder for dimensionality reduction and the SAINT for modeling tabular DEG signatures. The model was trained on a large public DEG compendium with pathway labels derived from EnrichR and interpreted using Integrated Gradients (IG) to prioritize driver genes. Experimental validation was conducted using mouse (n=5) and human (n=3) chondrocyte cultures, and in vivo OA-induced mouse models (n=5 per group).

Results: SaintGSE analysis identified key OA-related signaling pathways potentially modulated by natural extract from *Senna obtusifolia*, including NF-κB and p38. In vitro, the extract reduced OA catabolic factors compared with IL-1β–treated controls, including Mmp3 (mean difference 52.06, 95% CI 32.8–71.3, $p < 0.0001$), Mmp13 (mean difference 5.13, 95% CI 2.9–7.4, $p < 0.0001$), and Cox2 (mean difference 1.39, 95% CI 1.1–1.7, $p < 0.0001$). In vivo, the extract significantly reduced cartilage damage compared with the DMM-operated PBS–treated group (effect estimate [mean rank difference] 13.3, p=0.0009). IG-based driver genes provided compact, condition-specific candidates supporting each pathway prediction.

Conclusions: SaintGSE offers scalable, explainable pathway prediction from DEG signatures and enables mechanism- and target-oriented follow-up for disease studies and drug candidate screening.

© 2026 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Introduction

Signaling pathway analysis, traditionally based on differentially expressed genes (DEGs), is increasingly incorporating artificial intelligence (AI) to analyze gene sets from large-scale gene expression profiles. Previous studies have significantly increased the accessibility of biological insights through user-friendly web-based tools, offering novel ways to explore transcription factor (TF) interactions and discover related genes or drugs in signaling pathways [1–4]. However, traditional gene set enrichment analysis (GSEA) or over-

* Corresponding authors.
E-mail addresses: msjeon27@cau.ac.kr (M.-S. Jeon), yjtnng3043@g.skku.edu (J. Nam), wltjr0103@cau.ac.kr (M. Lee), ccho11@hanyang.ac.kr (C. Cho), yangsy@skku.edu (S. Yang), eyun@cau.ac.kr (S.-i. Eyun).
[1] Contributed equally to this work.

https://doi.org/10.1016/j.joca.2026.04.009
1063-4584/© 2026 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

representation analysis (ORA) still face limitations due to its reliance on predefined gene sets, while AI models often struggle with issues like insufficient training data and difficulty in handling tabular gene expression data and due to the black-box nature of the models, interpreting the results can be difficult.

To address challenges in gene expression analysis, we utilized fold change instead of raw gene expression data, allowing us to reduce outliers, streamline normalization, and enhance data augmentation for a more robust analysis [5–7]. To handle the computational cost of large-scale datasets, we implemented autoencoder (AE) for dimensionality. Additionally, application of the Self-Attention and Intersample Attention Transformer (SAINT) to tabular expression data generated superior performance compared to traditional machine learning approaches by considering relationships between features and samples at the same time [8,9]. Finally, the application of eXplainable AI (XAI) interpreted model outputs, using Integrated Gradients (IG) for each stage of AE and SAINT, identifying influential genes within signaling pathways, thus providing insight into our model's predictions [3,10].

This entire modeling process, named as SaintGSE, outperformed other models in evaluation metrics, particularly for imbalanced datasets, and was used to analyze signaling networks involved in osteoarthritis (OA). Osteoarthritis, whose incidence generally increases with age and various contributing factors, is characterized by pain, inflammation, and cartilage deformity [11,12]. The representative pro-inflammatory cytokine IL-1β activates downstream signaling pathways, such as the Erk, NF-κB, P38, and JNK pathways, leading to the secretion of OA-related factors, including metalloproteinases (Mmp3 and Mmp13) and cyclooxygenase 2 (Cox2), contributing to joint inflammation and cartilage damage [13,14]. While no curative drugs exist for OA, natural products like *Senna obtusifolia* have shown promise due to their multi-compound nature, which is also beneficial for alleviating several diseases [15–20]. *S. obtusifolia* is composed of compounds such as anthraquinones and naphthopyrone glycosides [21,22] and anthraquinones have been studied for their anti-inflammatory and anti-arthritic effects, which include Cox2 regulation and their role in drug development [23–26]. Our previous study also identified that obtusifolin, an anthraquinone from *S. obtusifolia*, can reduce OA severity by specifically regulating NF-κB signaling under OA conditions [27]. Hypothesizing that the natural extract would modulate more OA-related pathways than the single isolated compound, obtusifolin, due to synergistic effects, we used SaintGSE to identify key OA-related genes and signaling pathways. This research demonstrated that SaintGSE could efficiently predict candidate signaling pathways and showcased its broader applicability in predicting potentially effective drugs.

Methods

We compiled a large collection of public DEG signatures and trained SaintGSE to predict pathway involvement and identify driver genes from DEG profiles. We curated 1295 GEO experiments (Oct 2023–Apr 2024; ≥2 replicates per group) and performed within-study DEG analysis using pyDESeq2 (ver. 0.5.3), yielding 29,065 signatures; 929 additional human DEG signatures from the Gene Expression Atlas were included (total 29,994; summarized in Supplementary Fig. 1 and Supplementary Table 1) [28,29]. Each signature was represented as a $\log_2 FC$ vector over 36,926 genes (matched NCBI/Ensembl IDs); genes with adjusted $p > 0.05$ or not expressed were assigned $\log_2 FC = 0$ [30].

Pathway labels were derived by ORA in EnrichR using the "Elsevier_Pathway_Collection" library: genes with $|\log_2 FC| \geq 1$ were submitted to ORA, and pathways with adjusted $p < 0.05$ were labeled positive, producing 1719 pathway-wise binary tasks [31]. For application analyses, we additionally performed pre-ranked GSEA and ssGSEA using the same pathway library to compare ORA-, GSEA-, ssGSEA-, and SaintGSE-based pathway signals [32,33].

SaintGSE is a two-stage model (AE and SAINT) trained per pathway. Inputs were standardized using a training-only scaler. The AE was trained with MSE reconstruction plus an L1 penalty and encoded each signature into a 256-dimensional latent vector (selected by consistent performance; Supplementary Table 2). A [*CLS*] token was prepended to latent tokens and a SAINT-based classifier was pretrained and fine-tuned with BCEWithLogitsLoss. Data were split at the signature level into non-overlapping train/validation/test sets (70/10/20 for AE; 65/15/20 for SAINT). We compared SaintGSE with representative gradient-boosting and deep-learning baselines using the same AE latents, identical splits, early stopping, and area under the receiver operating characteristic curve (AUROC).

For interpretation, we report post-hoc calibrated pathway scores and IG-based gene attributions. SaintGSE pathway signals were compared with conventional ORA(EnrichR)/GSEA/ssGSEA outputs using the same pathway library [34]. Gene-level attributions were computed with IG on the pathway logit, using a zero baseline in the standardized input space. Chondrocyte and animal validation experiments, RNA-seq processing, and downstream pathway analyses, as well as replicate-dependent stability analysis in paired prostate cancer RNA-seq (GSE22260; paired DESeq2 (ver. 1.42.1) with nested downsampling N=5→2, seed=1), are described in the Supplementary Information [35].

Results

SaintGSE's transformer architecture is robust under severe class imbalance in signaling pathway prediction

SaintGSE formulates pathway involvement prediction as independent binary classification tasks (1719 total; one model per pathway) (Fig 1A; Methods). We benchmarked SaintGSE against nine baselines (XGBoost, LightGBM, CatBoost, CNN, RNN, LSTM, FT-Transformer, TabNet, and NODE) across four pathway occurrence groups (rare, low, balanced, and high) using 5-fold stratified cross-validation on the same fixed AE-encoded inputs for all models. Performance was evaluated using AUROC, area under the precision–recall curve (AUPRC) and F1 score (Fig. 1B–D) [36–38]. In low-to-high occurrence pathways, SaintGSE showed discrimination and ranking performance comparable to that of the strongest baseline models. For example, in "Chronic Myeloid Leukemia" (n=4588), SaintGSE achieved mean AUROC/AUPRC/F1 values of 0.868 (95% CI 0.862–0.874)/0.695 (95% CI 0.681–0.709)/0.629 (95% CI 0.618–0.640). Corresponding mean AUROC/AUPRC/F1 values were 0.883 (95% CI 0.876–0.890)/0.712 (95% CI 0.705–0.719)/0.643 (95% CI 0.628–0.658) for XGBoost, 0.874 (95% CI 0.868–0.880)/0.699 (95% CI 0.690–0.708)/0.629 (95% CI 0.618–0.640) for LightGBM, and 0.864 (95% CI 0.855–0.873)/0.687 (95% CI 0.678–0.696)/0.624 (95% CI 0.610–0.638) for CatBoost, indicating performance comparable to the gradient-boosting baselines in this low-occurrence pathway.

However, performance differences became more evident under extreme imbalance, a setting of particular importance because 400 of 1719 pathways (23.3%) were in the rare-occurrence group and 1305 of 1719 pathways (75.9%) were in the low-occurrence group (Supplementary Fig. 2A). In the rare pathway "Proteins with Altered Expression in Cancer-Associated DNA Methylation" (n=105), AUPRC was low across all models, as expected under extreme imbalance. SaintGSE achieved mean AUROC/AUPRC/F1 values of 0.766 (95% CI 0.715–0.817)/0.111 (95% CI 0.061–0.161)/0.192 (95% CI 0.125–0.259). Corresponding mean AUROC/AUPRC/F1 values were 0.678 (95% CI 0.580–0.776)/0.029 (95% CI 0.000–0.059)/0.073 (95% CI 0.018–0.128) for FT-Transformer, 0.506 (95% CI 0.424–0.588)/0.006 (95% CI 0.000–0.013)/0.005 (95% CI 0.000–0.011) for TabNet, and 0.478 (95%

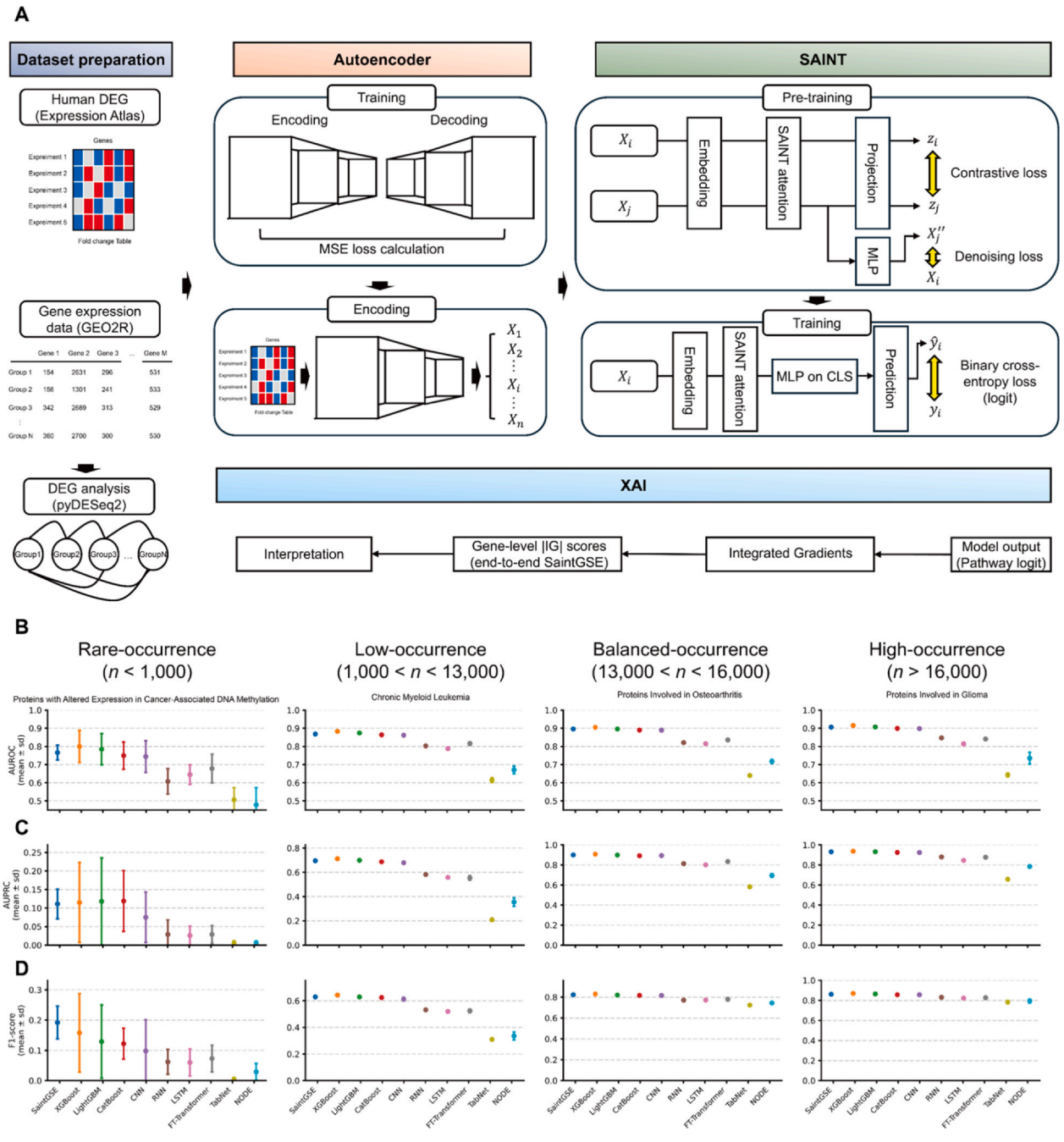


Fig. 1

Osteoarthritis and Cartilage

Schematic representation of the SaintGSE architecture and the evaluation metrics of seven distinct models across four occurrence groups. (A) In the dataset preparation stage, human differentially expressed gene (DEG) datasets from Expression Atlas and GEO are merged and submitted to an autoencoder (AE). The AE is pre-trained by minimizing the mean squared error (MSE) loss between the raw DEG data and its reconstructed data after encoding and decoding. Following pre-training, the AE encodes the merged DEG dataset into a latent gene set, which is subsequently used for the pretraining of the SAINT model. SAINT fine-tunes its parameters during pretraining and learns to predict using the [*CLS*] token during the training phase. For model interpretation, gene-level contribution scores were computed on the trained end-to-end SaintGSE model using Integrated Gradients (IGs). (B–D) Predictive performance of SaintGSE and baseline models was evaluated across four pathway occurrence groups (rare, low, balanced, and high) using 5-fold stratified cross-validation and is reported as mean ± standard deviation (s.d.) for (B) AUROC, (C) AUPRC, and (D) F1-score. All baseline models were trained using the same AE-encoded latent gene set as SaintGSE. The occurrence groups were defined by the number of positive DEG signatures in the full dataset (29,996 signatures): rare ($n < 1000$), low ($1000 < n < 13{,}000$), balanced ($13{,}000 < n < 16{,}000$), and high ($n > 16{,}000$). The occurrence counts for the representative pathways shown are n=105 for “Proteins with Altered Expression in Cancer-Associated DNA Methylation,” n=4588 for “Chronic Myeloid Leukemia,” n=15,399 for “Proteins Involved in Osteoarthritis,” and n=17,987 for “Proteins Involved in Glioma.”.

CI 0.361–0.595)/0.007 (95% CI 0.003–0.011)/0.029 (95% CI 0.000–0.064) for NODE, indicating less severe degradation than in these tabular deep-learning baselines. This robustness may reflect SaintGSE's design, which combines feature-wise attention with an additional mini-batch inter-sample attention block that may capture sample-level context beyond within-sample feature interactions (Supplementary Fig. 3). Although tree-based boosting models matched or exceeded SaintGSE in AUROC for some tasks (Fig. 1B–D), their explanations are naturally defined in the AE latent feature space, whereas SaintGSE supports end-to-end, gradient-based gene attribution through the encoder and classifier within a single differentiable framework. To test whether imbalance could be mitigated beyond architectural choices, we additionally evaluated imbalance-handling loss functions (class re-weighting and focal loss) and mixup/CutMix augmentation. Class re-weighting and focal loss did not consistently improve performance relative to the unaugmented model (Supplementary Table 3). Mixup/CutMix-based augmentation consistently improved AUROC across the five representative rare-occurrence pathways and improved AUPRC in four of five pathways, but improved F1 in only three of five pathways (Supplementary Table 4), indicating metric-dependent trade-offs rather than a uniformly superior performance profile. Consequently, we retained the unaugmented SaintGSE model as the final predictive model, prioritizing consistency across evaluation metrics and a training setup directly aligned with downstream gene-level interpretation.

Applying SaintGSE to mouse chondrocyte RNA-seq reveals OA-related pathway involvement by an independent scoring framework

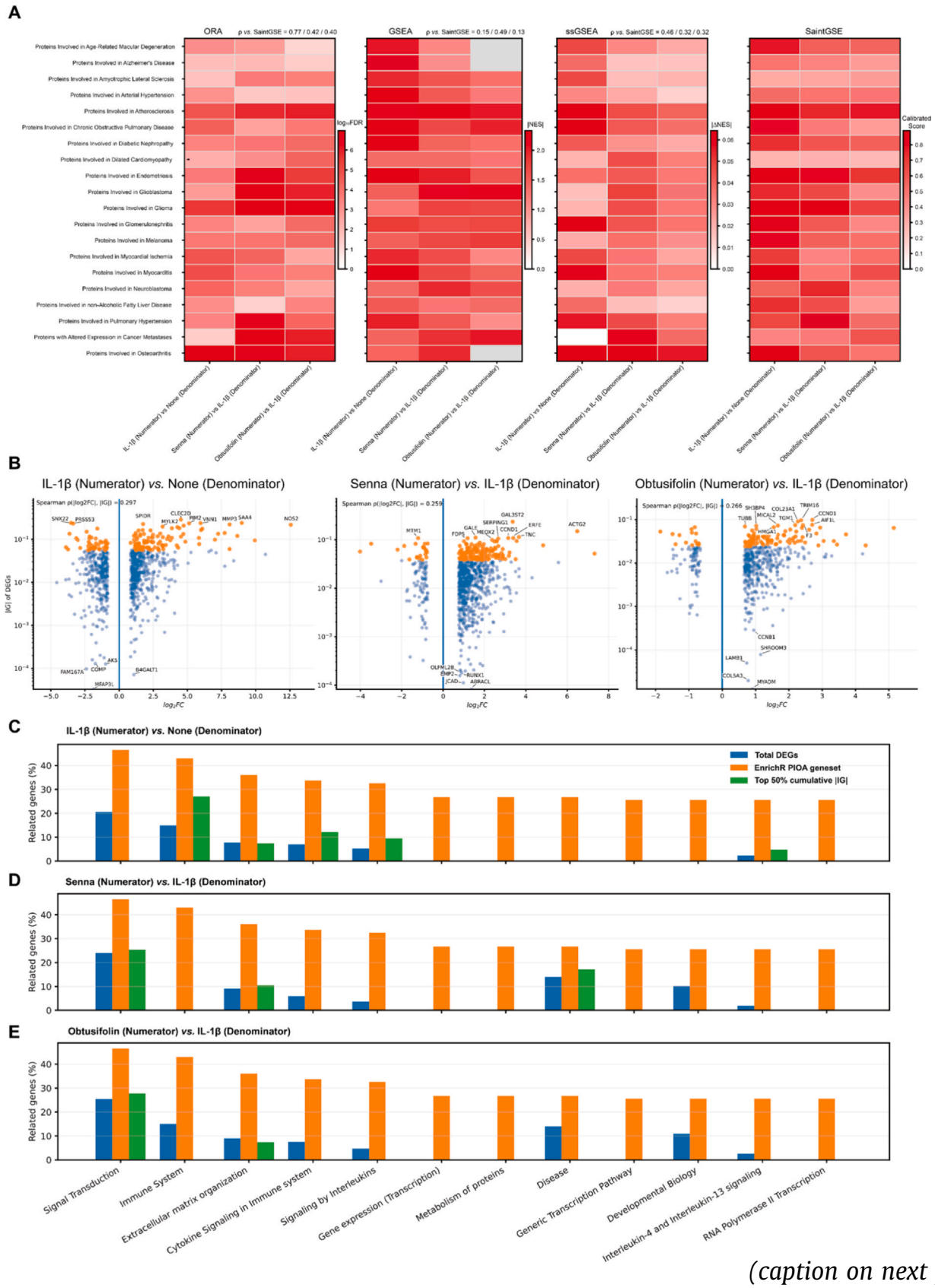
We subsequently applied SaintGSE to an external RNA-seq dataset of mouse chondrocytes by converting mouse genes to human orthologs and evaluated three contrasts: (1) IL-1β-treated (Numerator) *vs*. None (Denominator) mouse chondrocytes, (2) IL-1β-treated (Numerator) *vs*. IL-1β and *S. obtusifolia*-cotreated mouse chondrocytes (Denominator), and (3) IL-1β-treated (Numerator) *vs*. IL-1β and obtusifolin-cotreated mouse chondrocytes (Denominator). Using the same Elsevier Pathway Collection, we compared ORA, pre-ranked GSEA, and ssGSEA against SaintGSE on a harmonized pathway set. For a like-for-like comparison, we retained only terms detectable by all three conventional methods (ORA adjusted $p \le 0.05$, GSEA FDR $q \le 0.25$, ssGSEA max $|z| \ge 1.5$ in any sample), resulting in 20 pathways (Fig. 2A). For each contrast, we quantified agreement by Spearman rank correlation across pathways using method-specific scores: ORA = $-\log_{10}$(FDR), GSEA = |NES|, ssGSEA = |ΔNES|, and SaintGSE = calibrated score (post-hoc calibrated on validation set). Because these approaches summarize pathway signals using different assumptions (gene-set aggregation *vs.* learned scoring), pathway rankings differed across methods, particularly for IL-1β (Numerator) *vs.* None (Denominator), where rank similarity between SaintGSE and GSEA was low (Spearman ρ=0.153), whereas similarity between SaintGSE and ORA was higher (ρ=0.769).

To understand why pathway rankings diverged between enrichment-based methods (predefined gene sets) and SaintGSE, we compared curated pathway members with SaintGSE gene-level attributions derived by IG. IG attributions were highly concentrated across signatures: only 17.2% of DEGs in IL-1β-treated (Numerator) *vs.* None (Denominator), 20.3% in IL-1β-treated (Numerator) *vs. S. obtusifolia*-cotreated mouse chondrocytes (Denominator), and 20.4% in IL-1β-treated (Numerator) *vs.* obtusifolin-cotreated mouse chondrocytes (Denominator) were required to account for 50% of cumulative |IG| (Supplementary Fig. 4A). We therefore adopted top 50% of cumulative |IG|—the minimal DEG set explaining 50% of cumulative |IG|—as a compact definition of high-contribution genes and applied it to discordant cases. In the IL-1β-treated (Numerator) *vs.* None (Denominator) contrast, where the rank similarity between SaintGSE and GSEA was lowest (ρ=0.153), we examined a representative discordant pathway ("Proteins Involved in Glioma"). Overlap between the curated Elsevier gene set and top 50% of cumulative |IG| was minimal (Jaccard J=0.016; Supplementary Fig. 4B). Notably, top IG-ranked drivers were predominantly out-of-set yet plausible in an inflammatory response (e.g., *NOS2*, *ARG2*, *ANGPTL4*, *MMP3*), indicating that SaintGSE can prioritize condition-responsive drivers even when overlap with curated pathway members is limited (Supplementary Fig. 4C) [14].

Across all three contrasts, SaintGSE assigned higher calibrated scores to "Proteins Involved in Osteoarthritis" (PIOA; 0.99, 0.61, and 0.49) than to OA (0.08, 0.19, and 0.07) (Supplementary Fig. 4D). To interpret what supported these predictions in each signature, we examined IG-based gene attributions and focused on the top 50% of cumulative |IG|, which yields a compact, pathway-directed driver set (Fig. 2B). In IL-1β-treated (Numerator) *vs.* None (Denominator), multiple strongly induced inflammatory/catabolic genes also carried high attribution (e.g., *NOS2*, $\log_2FC$=12.56, |IG|=0.217; *MMP3*, $\log_2FC$=8.09, |IG|=0.221), consistent with an IL-1β–driven inflammatory program supporting the PIOA prediction [14,39]. Importantly, SaintGSE did not simply prioritize fold change: some genes showed only modest expression shifts yet ranked among the strongest decision-driving contributors (e.g., *SPIDR*, $\log_2FC$=1.75, |IG|=0.226). *SPIDR* has been implicated in DNA damage response/repair-linked regulatory programs, which is mechanistically compatible with emerging evidence that genome stress, senescence, and inflammatory signaling are intertwined in OA-like chondrocyte states, making it a plausible context-specific driver candidate despite moderate fold change [40]. In contrast, other DEGs with non-trivial fold changes contributed negligibly (e.g., *COMP*, $\log_2FC$=−1.74 but |IG|=1.3×10^{-4}), indicating selective, context-dependent attribution rather than uniform weighting of DEGs.

Under co-treatment conditions, the IG-based driver profiles differed from those in IL-1β-treated (Numerator) *vs.* None (Denominator). In IL-1β-treated (Numerator) *vs. S. obtusifolia*-cotreated (Denominator), high-attribution genes were enriched for remodeling-associated components (e.g., *TNC* and *ACTG2* among top contributors), together with regulatory nodes such as *CCND1*, suggesting a shift from dominant inflammatory markers toward tissue remodeling and signaling axes [39,41]. In IL-1β-treated (Numerator) *vs.* obtusifolin-cotreated (Denominator), top attributions emphasized structural/regulatory factors (e.g., *COL23A1*, *TRIM16*, *TGM1*), and included candidates with relatively small fold changes but notable attribution (e.g., *SH3BP4*, $\log_2FC$=1.02, |IG|=0.072) [27]. *SH3BP4* has been reported as a functional OA susceptibility gene in in vivo screening and its known roles in receptor trafficking and growth/nutrient signaling modules (*e.g.*, FGFR/mTOR/Wnt-linked regulation) provide a coherent interpretation for why it emerges as a treatment-specific, model-prioritized driver [42]. Across contrasts, bottom-ranked genes could retain measurable fold changes but near-zero attribution, supporting that |IG| reflects signature-specific contribution to the PIOA involvement prediction rather than differential expression magnitude.

To connect model-derived drivers with conventional gene-set interpretation, we compared (1) the top 50% of cumulative |IG| genes, (2) the full DEG list, and (3) the EnrichR PIOA gene set, and summarized their functional orientation by DAVID/Reactome (Fig. 2C–E). The curated PIOA gene set showed broad, largely invariant coverage across immune/signaling/ECM categories, reflecting fixed prior membership [14]. In contrast, SaintGSE drivers were more selective and contrast-dependent: IL-1β stimulation emphasized immune/cytokine-related categories, whereas co-treatment signatures shifted toward signaling and extracellular matrix organization, consistent with distinct transcriptional programs underlying each condition. Together, these results support that SaintGSE complements enrichment outputs by providing a calibrated pathway score and a compact, context-specific driver set, which can guide downstream validation of OA-relevant catabolic and remodeling factors.



(caption on next page)

Fig. 2

Osteoarthritis and Cartilage

Application of SaintGSE to an external mouse chondrocyte RNA-seq dataset highlights OA-related pathway involvement and context-specific driver genes. Mouse genes were converted to human orthologs, and DEG signatures were generated for three contrasts: IL-1β-treated (Numerator) *vs.* None (Denominator), IL-1β-treated (Numerator) *vs. Senna obtusifolia*-cotreated (Denominator), and IL-1β-treated (Numerator) *vs.* Obtusifolin-cotreated (Denominator). (A) Benchmark comparison of ORA (EnrichR), GSEA, ssGSEA, and SaintGSE across the three DEG signatures. Pathways shown (n=20) were selected as the set of terms detectable by all three conventional methods under the analysis thresholds (ORA adjusted $p \leq 0.05$; GSEA FDR q≤0.25; ssGSEA max |z|≥1.5 in any sample). Pathway scores were method-specific (ORA: -$\log_{10}$(FDR); GSEA: |NES|; ssGSEA: |ΔNES|; SaintGSE: calibrated score), with missing values shown as NA (gray). Spearman's rank correlation (ρ) between SaintGSE and each conventional method was computed across the 20 pathways for each contrast and is shown beneath each contrast. (B) Gene-level interpretation of the "Proteins involved in osteoarthritis" (PIOA) prediction. DEGs were ranked by |IG| and summarized by cumulative |IG|; high-contribution genes (orange) were defined as the minimal set of DEGs accounting for the top 50% of cumulative |IG|. (C-E) Functional orientation of three gene sets—total DEGs, the EnrichR-derived PIOA gene set, and SaintGSE drivers (top cumulative 50% |IG|)—for three contrasts using DAVID Reactome annotations. Bars show related genes (%; y-axis), defined as the fraction of genes in each input set annotated to each Reactome category (x-axis).

Generalizability and replicate-dependent stability of SaintGSE

We further analyzed an external prostate cancer RNA-Seq dataset with matched tumor–benign pairs (GSE22260). To evaluate the effect of biological replicate number on pathway involvement prediction, we generated DEG signatures for primary prostate cancer (Numerator) *vs.* matched benign tissues (Denominator) while varying the number of replicates (N=2–5). The calibrated score of SaintGSE for the "Prostate cancer" pathway (n=10,123) was 0.03, 0.07, 0.98, and 0.05 at N=2, N=3, N=4, and N=5, respectively, whereas the corresponding score for "Proteins involved in prostate cancer" (n=13,379) was 0.11, 0.14, 0.42, and 0.28 (Fig. 3A), indicating sensitivity of pathway scores to replicate number. Although the number of detected DEGs increased with additional replicates, DEG composition varied across replicate settings (Fig. 3B; Fig. S3A), indicating that replicate choice affected the inferred transcriptional signature.

We next assessed the stability of gene-level interpretation across replicate settings by defining an anchor set as the top 50% cumulative |IG| genes from the N=4 signature (58 genes). These anchor genes were rarely recovered in the DEG lists at N=2–3, but showed partial retention at N=5 (41/58 present as DEGs, with 27/58 remaining within the top cumulative |IG| set; Fig. 3C), consistent with additional biological heterogeneity across sample subsets reshaping gene prioritization. Notably, representative genes highlighted in Fig. 3D include *IFNAR2–IL10RB* (readthrough/chimeric transcript candidate), *SAA2* (immune/inflammatory signal), and transcriptional regulators such as *SETD7* and *TP53INP1*, illustrating that |IG| reflects model reliance on expression features rather than a simple function of $\log_2$*FC* [41,43–45]. Consistently, |IG| exhibited limited association with $\log_2$*FC* (ρ=0.26; Fig. 3D). Finally, pathway enrichment based on the total DEG set was dominated by immune-related terms (*e.g.*, "Innate Immune System" and "Neutrophil degranulation"), whereas enrichment restricted to the SaintGSE-prioritized genes (top 50% cumulative |IG| among DEGs) retained the immune signal but additionally highlighted programs such as small-molecule transport (including "SLC-mediated transmembrane transport") and the "RHO GTPase cycle" (Fig. 3E), supporting that SaintGSE concentrates model decision on decision-driving signals within the expression signature [46].

Senna obtusifolia extract attenuates the levels of OA catabolic factors in chondrocytes

To validate the result of SaintGSE using DEG comparisons, we treated cells with *S. obtusifolia* at 10, 50, 100, or 200 μg/mL. After confirming cell damage at a concentration of 200 μg/mL through LDH (Fig. 4A and Supplementary Table 5), we further investigated the effect of *S. obtusifolia* under OA-mimicking conditions. Mmp3, Mmp13, and Cox2 are well-known as key molecules in the degradation of cartilage ECM in OA (15). When chondrocytes were treated with IL-1β in the presence of *S. obtusifolia* extract under OA-mimicking conditions, the extract significantly reduced the expression of OA catabolic factors compared with IL-1β–treated controls, including *Mmp3* (mean difference 52.06, 95% CI 32.871.3, $p < 0.0001$), *Mmp13* (mean difference 5.13, 95% CI 2.9–7.4, $p < 0.0001$), and *Cox2* (mean difference 1.39, 95% CI 1.1–1.7, $p < 0.0001$) (Fig. 4B, C and Supplementary Table 6). Cox2 protein expression was also reduced compared with IL-1β-treated controls, consistent with the mRNA results (Fig. 4B, D and Supplementary Table 7). The magnitude of reduction was comparable to, and in some cases numerically greater than, that observed with obtusifolin treatment under the same conditions. In OA, cartilage extracellular matrix (ECM) degradation is mediated in part by collagenase activity and PGE_2 secretion [47]. Under IL-1β stimulation, *S. obtusifolia* extract significantly reduced PGE_2 secretion compared with IL-1β–treated controls (effect estimate [mean rank difference] 22.6, p=0.0002) (Fig. 4E and Supplementary Table 8). Collagenase activity was also significantly reduced compared with IL-1β–treated controls (mean difference 1.33, 95% CI 0.93–1.73, $p < 0.0001$) (Fig. 4F and Supplementary Table 9). The magnitude of reduction was numerically greater than that observed with obtusifolin treatment. To assess translational relevance, we evaluated whether similar effects were observed in human chondrocytes (Supplementary Fig. 6A). In human chondrocytes, *S. obtusifolia* extract did not induce significant cytotoxicity and significantly reduced OA catabolic factors and collagenase activity compared with IL-1β–treated controls, demonstrating a consistent pattern of modulation (Supplementary Fig. 6B–6E and Supplementary Tables 10–12).

Prediction and verification of the signaling pathways regulated by S. obtusifolia extract under OA conditions

To predict the OA pathways regulated by *S. obtusifolia* extract and obtusifolin, an RNA-seq analysis was performed. From the sequencing results, various gene-related drugs were identified and compared to predict signaling using ingenuity pathway analysis (IPA). Among them, IL-1β–related signaling drugs were identified. First, we compared the number of IL-1β–related signaling modulators associated with genes regulated by the whole extract and obtusifolin (Fig. 5A). Representative pathways for IL-1β signaling include the Erk, NF-κB, P38, and JNK pathways, and we identified drugs that regulate genes involved in each pathway (Fig. 5B) [14]. Overlapping drugs (Fig. 5A and B) with similar effects (Fig. 5C) were found between the two substances. A greater proportion of non-overlapping drugs was associated with *S. obtusifolia* extract across multiple signaling pathways (Fig. 5D and E), suggesting broader pathway modulation compared with obtusifolin. Western blot analysis confirmed

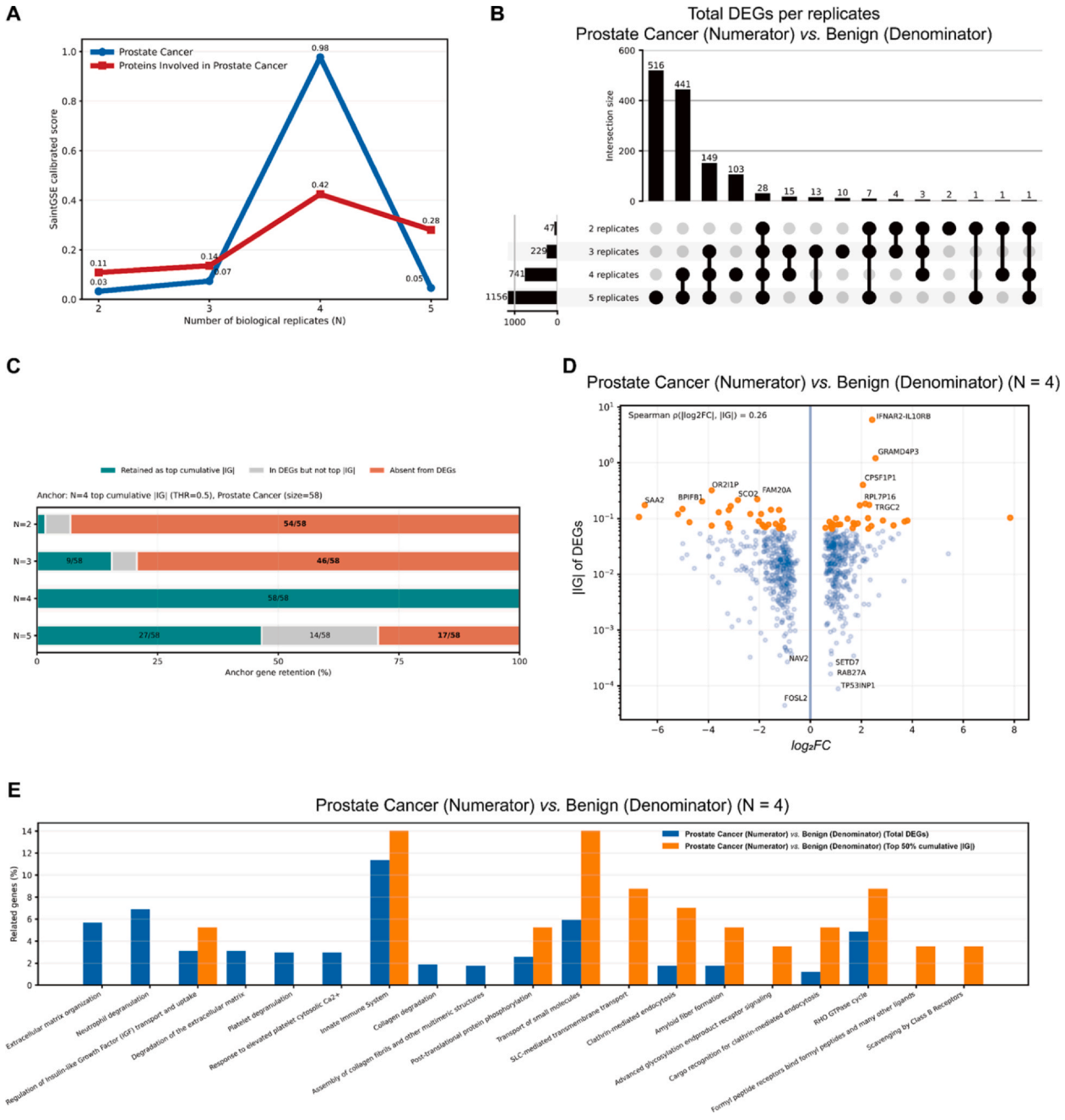


Fig. 3

Osteoarthritis and Cartilage

Replicate number affects stability of SaintGSE pathway scores and IG-attributed genes in an external prostate cancer dataset. An external prostate cancer RNA-seq dataset (GSE22260; tumor and matched benign samples) was used to assess replicate-dependent stability of SaintGSE predictions. Differential expression was computed with a paired design (tumor (Numerator) *vs.* matched benign (Denominator) within each patient), and $\log_2FC$ signatures were used for SaintGSE inference using two pre-trained prostate-related pathways ("Prostate cancer" and "Proteins involved in prostate cancer"). (A) Replicate-dependent changes in SaintGSE calibrated pathway scores across nested downsampling (N=2–5 matched pairs). (B) Total number of DEGs (adjusted $p < 0.05$) detected at each replicate count and their overlap structure across nested subsets, visualized as an UpSet-style intersection summary. (C) Retention of the anchor gene set defined by the N=4 top 50% cumulative |IG| genes (anchor size=58). For each replicate count (N=2, 3, 4, and 5), anchor genes were categorized as: (1) retained as top cumulative |IG|, (2) present among DEGs but not in the top cumulative |IG| set, or (3) absent from the DEG list; values are reported as fractions of the anchor set. (D) Relationship between gene-level attribution and effect size at N=4. Each point represents a DEG, plotted by $\log_2FC$ (x-axis) and |IG| (y-axis); genes in the N=4 top 50% cumulative |IG| set are highlighted. The annotated value indicates the Spearman correlation (ρ) between |IG| and $|\log_2FC|$. (E) Reactome over-representation analysis comparing pathway composition between all DEGs ("Total DEGs") and the top 50% cumulative |IG| subset at N=4. Bars show the percentage of genes in each set (y-axis) associated with the displayed Reactome terms (Related genes, %; x-axis); terms are shown as the union of top-ranked results from the two gene sets.

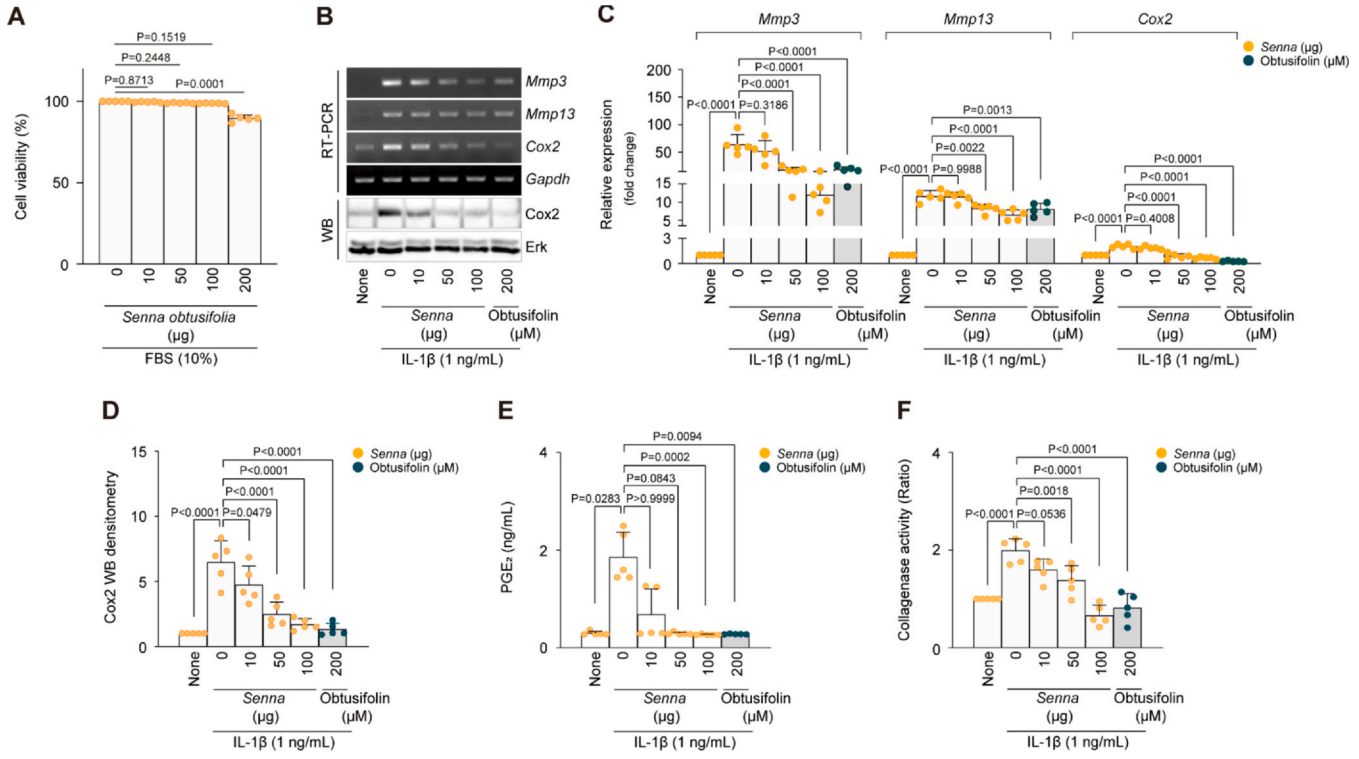


Fig. 4

Osteoarthritis and Cartilage

Senna obtusifolia extract reduces the levels of osteoarthritis (OA)-related factors in chondrocytes. (A) Mouse chondrocytes were treated with 10, 50, 100, or 200 μg/mL of *S. obtusifolia* extract for 24 h, and cell viability was measured using a lactate dehydrogenase assay. (B-F) Interleukin-1β (IL-1β; 1 ng/mL) was co-administered with either *S. obtusifolia* extract (10, 50, or 100 μg/mL) or obtusifolin (200 μM) for 24 h. The expression of OA catabolic factors *Mmp3*, *Mmp13*, and *Cox2* were confirmed using PCR (B-upper) and quantified using qPCR (C). The expression of Cox2 protein was determined using western blotting (B-lower) and densitometry (D). (E and F) Assessments of prostaglandin E2 (PGE_2) and collagenase activities. The data for each group are shown as the mean ± SD (n=5). Panels A, C, D and F were analyzed using one-way ANOVA followed by Dunnett's multiple comparison test after confirmation of normality using the Shapiro-Wilk test. Panel E was analyzed using the Kruskal-Wallis test with Dunn's post hoc test. Adjusted p are indicated in the graphs.

these predictions. Under IL-1β stimulation, *S. obtusifolia* extract significantly reduced pp65 levels compared with IL-1β–treated controls (mean difference 6.05, 95% CI 3.58–8.52, $p < 0.0001$), reduced pp38 levels (mean difference 4.51, 95% CI 3.43–5.59, $p < 0.0001$), and increased IκB expression (mean difference 45.7, 95% CI 37.67–53.72, $p < 0.0001$) (Fig. 5F, G and Supplementary Table 13). The magnitude of modulation was numerically greater than that observed with obtusifolin treatment under the same conditions. This suggests that the predictions of the signaling pathways regulated by related drugs are supported by experimental validation. Consistently, similar modulation of NF-κB and p38 signaling pathways was observed in human chondrocytes compared with IL-1β–treated controls (Supplementary Fig. 6F, 6G, and Supplementary Table 14).

Inhibition of cartilage destruction by S. obtusifolia extract

We have so far demonstrated that *S. obtusifolia* extract downregulates OA at the cellular level and previous studies have shown that obtusifolin reduces cartilage damage in mice [27]. Thus, we conducted an animal experiment to see if it could prevent damage to cartilage tissue. First, DMM surgery was performed on 12-week-old mice to induce OA and then *S. obtusifolia* extract was orally administered daily for six weeks starting four weeks after surgery (Fig. 6A). As the administered concentrations increased, cartilage damage was reduced compared with DMM-operated PBS-treated group (Fig. 6B). Specifically, OARSI scores were significantly reduced (effect estimate [mean rank difference] 13.3, p=0.0009), SBP thickness was decreased (mean difference 79.26, 95% CI 51.82–106.7, $p < 0.0001$), and Osteophyte formation was suppressed (effect estimate [mean rank difference] 12.6, p=0.0019) compared with DMM-operated PBS-treated group, demonstrating a consistent protective effect of *S. obtusifolia* across all scoring parameters (Fig. 6C-E, and Supplementary Tables 15–17).

Discussion

Architectural design and interpretability of SaintGSE

We developed SaintGSE to predict the association between DEG results and specific signaling pathways and to identify key genes in these pathways using large gene expression datasets. Although gradient-boosting methods often outperform neural networks on conventional tabular benchmarks [9], our setting differs because all models operate on an AE-compressed representation of DEG signatures. Training directly on raw gene-level features (36,926 genes) would be computationally prohibitive at this scale in both memory

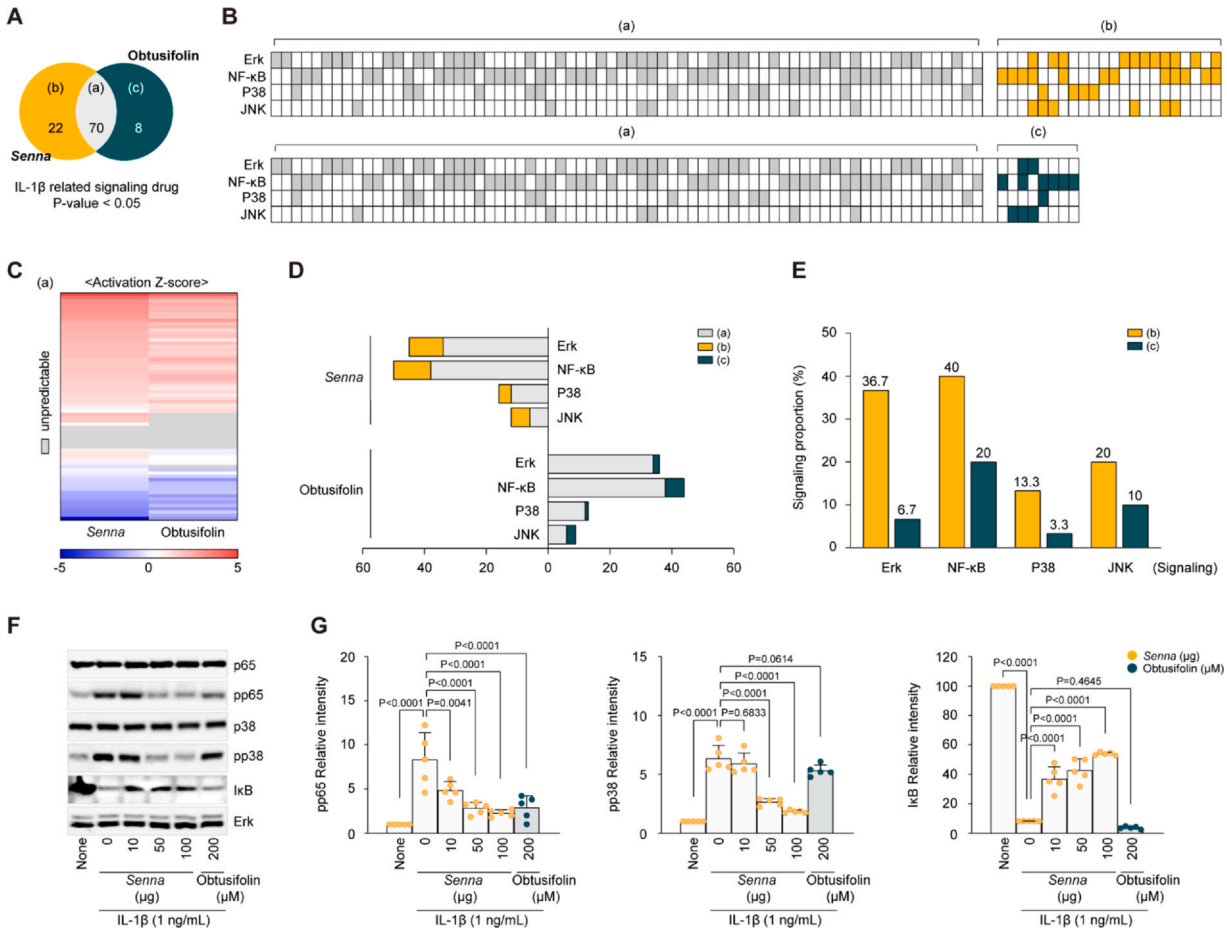


Fig. 5

Osteoarthritis and Cartilage

Osteoarthritis (OA) signaling is regulated by compounds from *Senna obtusifolia.* An ingenuity pathway analysis (IPA) was performed using RNA-seq data from chondrocytes treated with *S. obtusifolia* extract and obtusifolin, and the signaling regulation that these compounds produced was predicted through signaling drugs. After sorting the interleukin-1β (IL-1β) signaling drug and upstream drug data related to the genes regulated by each extract through IPA analysis, the numbers of overlapping and unique signaling drugs associated with each treatment were determined. (A) The number of IL-1β signaling drugs associated with the extract (b), obtusifolin (c), and both (a) is displayed as a Venn diagram. (B) A graph showing IL-1β signaling (Erk, NF-κB, P38, and JNK)-associated drugs belonging to parts (a), (b), and (c) of the Venn diagram. (C) The Z-scores of the downstream activation of the drugs in part (a) of the Venn diagram, presented as a heatmap. (D) The number of IL-1β signaling (Erk, NF-κB, P38, and JNK)-drugs in each extract belonging to parts (a), (b), and (c) of the Venn diagram. (E) The proportion of the signaling involved in the drugs in (b) and (c) of the Venn diagram. The signaling proteins regulated in each extract were identified using western blotting (F) and quantified by densitometry (G). The data are shown as mean ± SD (n=5). All statistical data were analyzed using one-way ANOVA followed by Dunnett's multiple comparison after confirmation of normality using the Shapiro-Wilk test. Adjusted p are indicated in the graph.

and runtime. Moreover, because pathway labels are extremely imbalanced, accuracy can be misleading, and generic remedies (e.g., mixup/CutMix, class re-weighting, or focal loss) showed no consistent improvements in rare-pathway ranking performance across tasks. In this context, SaintGSE's attention-based architecture can model non-linear interactions among AE latent features and showed robust behavior under class imbalance in our benchmarks. In addition, the SAINT architecture uses both feature-wise attention and inter-sample attention, which may contribute to stable ranking performance when positive signals are sparse and distributed across correlated dimensions. For interpretation, tree-based baselines explain importance in the AE latent space (e.g., TreeSHAP on latent dimensions). By contrast, SaintGSE supports end-to-end, gradient-based gene-level attributions through a single differentiable pipeline using IG, which we report as our primary interpretation.

Application of SaintGSE across external datasets

Leveraging a large and diverse training compendium (Supplementary Table 1; Supplementary Fig. 1), SaintGSE is designed to generalize across heterogeneous DEG contexts and to prioritize context-specific drivers from each input signature. Because pathway labels are derived from EnrichR ORA, they serve as weak supervision, and calibrated scores of SaintGSE and IG-prioritized drivers should be interpreted as complementary, context-specific evidence alongside enrichment outputs and experimental validation. To assess

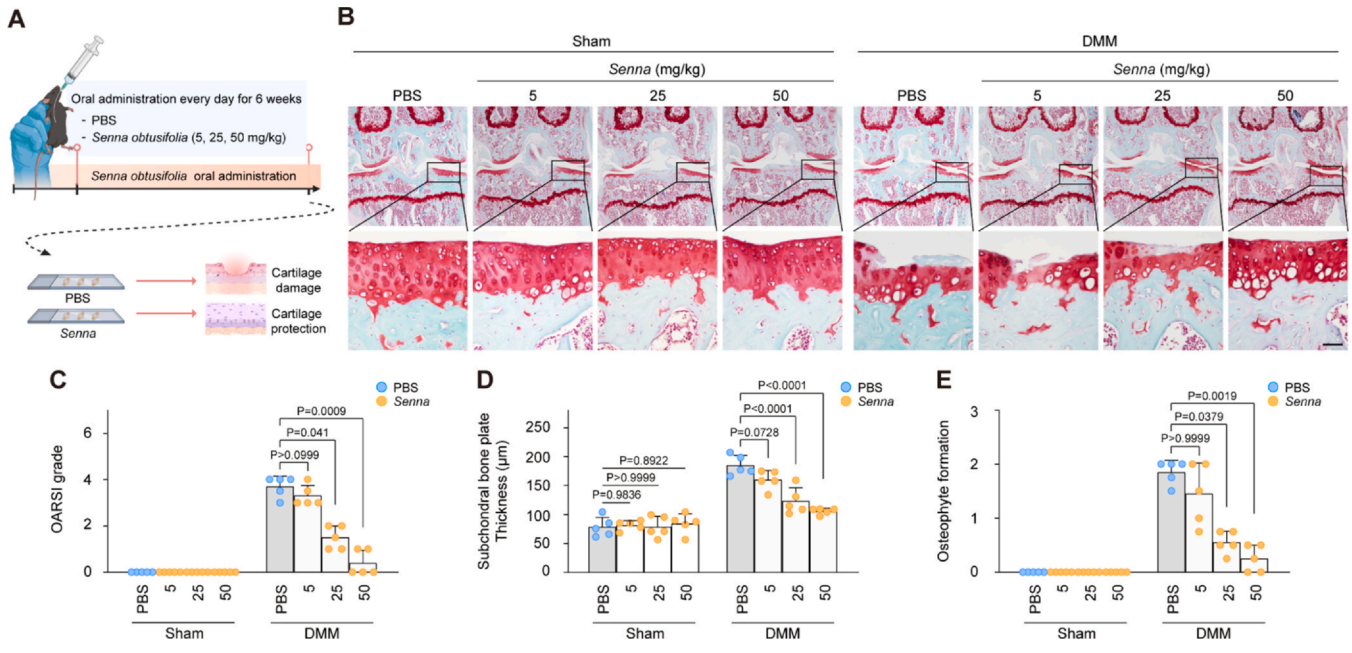


Fig. 6

Osteoarthritis and Cartilage

Protective effects of *Senna obtusifolia* on mouse cartilage tissue. Osteoarthritis (OA) was induced in mice by performing destabilization of the medial meniscus (DMM) surgery, which was followed by oral administration of PBS (control) or *S. obtusifolia* extract (5, 25, or 50 mg/kg). Sham-operated mice were used as controls. (A) The time schedule for oral administration: OA-induced mice were orally administered PBS or *S. obtusifolia* every day for six weeks starting four weeks after the surgery. (B) Histological images of cartilage destruction using safranin O staining. Cartilage destruction was scored based on the OARSI grade (C), subchondral bone plate thickness (D), and osteophyte formation (E). The data for each group are shown as the mean ± SD (n=5). All in vivo experiments were performed using independent animals. For statistical analysis panels C and E were analyzed using the Kruskal-Wallis test followed by Dunn's post hoc test, whereas panel D was analyzed using one-way ANOVA with Dunnett's post hoc test after confirming normality with the Shapiro-Wilk test. Adjusted p are indicated in the graph.

practical generalizability and the stability of these gene-level interpretations under realistic perturbations, we evaluated two external datasets.

In addition to performance benchmarking, the external mouse chondrocyte dataset illustrates how SaintGSE can be used as a complementary readout to conventional enrichment analyses. ORA and pre-ranked GSEA summarize pathway-level signals by aggregating evidence over predefined gene sets, whereas SaintGSE produces a calibrated pathway involvement score directly from the DEG signature and enables a model-based interpretation of which DEGs support that prediction through |IG|. In the mouse OA setting, SaintGSE consistently assigned higher calibrated scores to PIOA than to OA, and the IG profiles showed that the PIOA prediction relied on a compact subset of DEGs rather than the full DEG list. This concentration is practically useful because it focuses mechanistic interpretation on decision-driving genes; conversely, DEGs with measurable fold changes can still receive low |IG| when they are not primary contributors under the specific perturbation context, indicating that SaintGSE prioritization is not a simple proxy for $|\log_2 FC|$. At the functional level, the EnrichR PIOA gene set exhibited a stable, broad Reactome composition, whereas SaintGSE drivers shifted across contrasts—immune/cytokine-oriented under IL-1β stimulation but more signal-transduction/ECM-oriented under co-treatment—supporting context-dependent driver selection rather than a direct reflection of static gene-set membership. Finally, this framework can surface candidate therapeutic targets that may be missed by fold-change-first prioritization; for example, genes with modest $|\log_2 FC|$ but high |IG| (e.g., *SPIDR* or *SH3BP4* in our analyses) merit follow-up as potential regulators of OA-relevant programs, which can be directly tested in further molecular assays.

We further evaluated generalizability and the impact of replicate number using an external prostate cancer dataset with matched tumor–benign pairs (GSE22260). This analysis highlights an important practical limitation of DEG-derived inputs: increasing replicate number can substantially change not only the number of detected DEGs, but also the composition of the DEG signature itself. Consistent with this, SaintGSE pathway scores and gene-level prioritization were replicate-dependent, with the calibrated score of "Prostate cancer" pathway peaking at N=4 but decreasing at N=5, and with partial loss of the N=4 anchor driver set when additional replicates were included. This behavior is not necessarily a model failure; rather, it reflects that DEG signatures are summaries of underlying transcriptional states, and changes in replicate composition (including biological heterogeneity) can shift the apparent signature that both enrichment methods and SaintGSE operate on. Importantly, the weak association between |IG| and $\log_2 FC$ observed in this dataset further supports that |IG| captures model reliance on specific features in the learned representation instead of simply ranking the largest fold changes. Finally, restricting enrichment to SaintGSE-prioritized genes retained dominant immune signals while revealing additional programs that were less evident from total-DEG enrichment alone, suggesting that SaintGSE can help separate broad background responses from signals that more directly drive the model's pathway involvement decision. Overall, these external

applications suggest that SaintGSE scores and drivers should be interpreted as summaries of the input DEG signature, which can shift with replicate composition and biological heterogeneity. Building on these generalizability analyses, we next interpret SaintGSE predictions in the OA perturbation setting and connect them to experimentally testable mechanisms.

Biological implications of Senna obtusifolia in osteoarthritis

The progression of OA is known to be exacerbated by the activation of pro-inflammatory cytokines like IL-1β, which triggers Erk, NF-κB, P38, and JNK signaling pathways that play well-established roles in OA regulation and are widely studied as therapeutic targets [13,14,48–50]. Previous studies have represented the synergistic effects from various active compounds in natural products [21,51] and the therapeutic role of *Senna obtusifolia,* as a traditional medicine in Asia and a natural product with anti-inflammatory, anti-hepatotoxic, anti-cancer, and anti-neurodegenerative activities [17–20]. Furthermore, our previous research revealed that obtusifolin, a single compound in *S. obtusifolia,* effectively regulates OA by modulating the NF-κB signaling pathway [27].

We hypothesized that *S. obtusifolia* would have broader effects on gene expression in OA beyond those of obtusifolin alone. SaintGSE predicted that treatment with either *S. obtusifolia* extract or obtusifolin in OA-induced mouse chondrocytes were associate with attenuation of OA-related phenotypes, primarily through modulation of ECM-related genes, and that *S. obtusifolia* had a broad impact compared to obtusifolin alone. These predictions were supported by downstream molecular assays and IPA-based drug-perturbation interpretation, linking the predicted signatures to actionable regulatory programs. Together, these results support SaintGSE as a practical framework for linking perturbation-specific DEG signatures to testable pathway mechanisms and candidate targets, and suggest that the multi-compound *S. obtusifolia* extract modulates broader OA-relevant programs than obtusifolin alone.

Author contributions

MJ&JN: Investigation, Formal analysis, Data Curation, Writing – Original draft, Writing – Review & editing, Visualization. **ML&CC:** Formal analysis, Data curation. **SY&SE:** Conceptualization, Writing – Original draft, Writing – Review & editing, Supervision, Project administration.

Funding

This work was supported by the National Research Foundation of Korea [RS-2025–02216962] and the Korea Institute of Marine Science & Technology Promotion [RS-2025–02215227 and RS-2026–25543836] funded by the Ministry of Oceans and Fisheries.

Data availability

The RNA-Seq data from mouse chondrocytes generated in this study have been deposited in the BioProject NCBI database under accession code PRJNA1178545 [http://www.ncbi.nlm.nih.gov/bioproject/1178545]. The SaintGSE code is available at https://github.com/Msjeon27/SaintGSE/ and has also been archived on Zenodo at https://doi.org/10.5281/zenodo.14219500.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors declare that no generative AI tools were used in the scientific writing.

Acknowledgments

We thank members of the Eyun lab for their valuable discussions.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.joca.2026.04.009.

References

[1] C.W.T. Koh, J.S.G. Ooi, E.Z. Ong, K.R. Chan, STAGEs: A web-based tool that integrates data visualization and pathway enrichment analysis for gene expression studies, Sci. Rep. 13 (2023) 7135, https://doi.org/10.1038/s41598-023-34163-2
[2] Y. Yuan, Z. Bar-Joseph, Deep learning of gene relationships from single cell time-course expression data, Brief. Bioinform. 22 (2021) bbab142, https://doi.org/10.1093/bib/bbab142
[3] Q. Li, Y. Yu, P. Kossinna, T. Lun, W. Liao, Q. Zhang, XA4C: eXplainable representation learning via Autoencoders revealing Critical genes, PLoS Comput. Biol. 19 (2023) e1011476, https://doi.org/10.1371/journal.pcbi.1011476
[4] H. Chen, F.J. King, B. Zhou, Y. Wang, C.J. Canedy, J. Hayashi, Y. Zhong, M.W. Chang, L. Pache, J.L. Wong, Y. Jia, J. Joslin, T. Jiang, C. Benner, S.K. Chanda, Y. Zhou, Drug target prediction through deep learning functional representation of gene signatures, Nat. Commun. 15 (2024) 1853, https://doi.org/10.1038/s41467-024-46089-y
[5] J.G. Burkhart, G. Wu, X. Song, F. Raimondi, S. McWeeney, M.H. Wong, Y. Deng, Biology-inspired graph neural network encodes reactome and reveals biochemical reactions of disease, Patterns 4 (2023), https://doi.org/10.1016/j.patter.2023.100758
[6] J. Zrimec, C.S. Börlin, F. Buric, A.S. Muhammad, R. Chen, V. Siewers, V. Verendel, J. Nielsen, M. Töpel, A. Zelezniak, Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure, Nat. Commun. 11 (2020) 6141, https://doi.org/10.1038/s41467-020-19921-4
[7] M. Sprang, M.A. Andrade-Navarro, J.-F. Fontaine, Batch effect detection and correction in RNA-seq data using machine-learning-based automated assessment of quality, BMC Bioinforma. 23 (2022) 279, https://doi.org/10.1186/s12859-022-04775-y
[8] G. Somepalli, M. Goldblum, A. Schwarzschild, C.B. Bruss, T. Goldstein, 2021, SAINT: Improved neural networks for tabular data via row attention and contrastive pre-training, Preprint at https://arxiv.org/abs/2106.01342.
[9] V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, G. Kasneci, Deep neural networks and tabular data: a survey, IEEE Trans. Neural Netw. Learn. Syst. 35 (2024) 7499–7519, https://doi.org/10.1109/TNNLS.2022.3229161
[10] M. Sundararajan, A. Taly, Q. Yan, Axiomatic attribution for deep networks, Int. Conf. Mach. Learn. ICML 70 (2017) 3319–3328, https://doi.org/10.5555/3305890.3306024
[11] E. Charlier, C. Deroyer, F. Ciregia, O. Malaise, S. Neuville, Z. Plener, M. Malaise, D. de Seny, Chondrocyte dedifferentiation and osteoarthritis (OA), Biochem. Pharm. 165 (2019) 49–65, https://doi.org/10.1016/j.bcp.2019.02.036
[12] M. Rahmati, G. Nalesso, A. Mobasheri, M. Mozafari, Aging and osteoarthritis: central role of the extracellular matrix, Ageing Res. Rev. 40 (2017) 20–30, https://doi.org/10.1016/j.arr.2017.07.004
[13] P. Wojdasiewicz, A. Poniatowski, D. Szukiewicz, The role of inflammatory and anti-inflammatory cytokines in the pathogenesis of osteoarthritis, Mediat. Inflamm. 2014 (2014) 561459, https://doi.org/10.1155/2014/561459
[14] S. Yang, J. Kim, J.-H. Ryu, H. Oh, C.-H. Chun, B.J. Kim, B.H. Min, J.-S. Chun, Hypoxia-inducible factor-2α is a catabolic regulator of osteoarthritic cartilage destruction, Nat. Med. 16 (2010) 687–693, https://doi.org/10.1038/nm.2153
[15] H.A. Wieland, M. Michaelis, B.J. Kirschbaum, K.A. Rudolphi, Osteoarthritis — an untreatable disease? Nat. Rev. Drug Discov. 4 (2005) 331–344, https://doi.org/10.1038/nrd1693
[16] A.G. Atanasov, S.B. Zotchev, V.M. Dirsch, C.T. Supuran, Natural products in drug discovery: advances and opportunities, Nat. Rev. Drug Discov. 20 (2021) 200–216, https://doi.org/10.1038/s41573-020-00114-z
[17] M.Y. Ali, S. Park, M. Chang, Phytochemistry, Ethnopharmacological Uses, Biological Activities, and Therapeutic Applications of Cassia obtusifolia L.: A Comprehensive Review, Molecules 26 (2021) 6252, https://doi.org/10.3390/molecules26206252
[18] C. Li, M. Li, W. Chang, B. Guo, Purification and characterization of a novel cholesterol-lowering protein from the seeds of *Senna obtusifolia*, Sci. China C. Life Sci. 51 (2008) 1020–1024, https://doi.org/10.1007/s11427-008-0129-2
[19] B.D. Drever, W.G.L. Anderson, G. Riedel, D.H. Kim, J.H. Ryu, D.-Y. Choi, B. Platt, The seed extract of *Cassia obtusifolia* offers neuroprotection to mouse

hippocampal cultures, J. Pharmacol. Sci. 107 (2008) 380–392, https://doi.org/10.1254/jphs.08034FP
[20] T. Kowalczyk, P. Sitarek, M. Toma, L. Picot, M. Wielanek, E. Skała, T. Śliwiński, An extract of transgenic *Senna obtusifolia* L. hairy roots with overexpression of PgSS1 gene in combination with chemotherapeutic agent induces apoptosis in the leukemia cell line, Biomolecules 10 (2020) 510, https://doi.org/10.3390/biom10040510
[21] H.-O. Re, P.-W. Kl, B. M, Medicinal components recoverable from sicklepod (*Senna obtusifolia*) seed: analysis of components by HPLC-MSn, J. Chromatogr. Sep. Tech. 2 (2012) 2, https://doi.org/10.4172/2157-7064.S1-001
[22] H.A. Jung, M.Y. Ali, J.S. Choi, Promising inhibitory effects of anthraquinones, naphthopyrone, and naphthalene glycosides, from cassia obtusifolia on α-glucosidase and human protein tyrosine phosphatases 1B, Molecules 22 (2017) 28, https://doi.org/10.3390/molecules22010028
[23] E.M. Malik, C.E. Müller, Anthraquinones as pharmacological tools and drugs, Med. Res. Rev. 36 (2016) 705–748, https://doi.org/10.1002/med.21391
[24] D. Shrimali, M.K. Shanmugam, A.P. Kumar, J. Zhang, B.K.H. Tan, K.S. Ahn, G. Sethi, Targeted abrogation of diverse signal transduction cascades by emodin for the treatment of inflammatory disorders and cancer, Cancer Lett. 341 (2013) 139–149, https://doi.org/10.1016/j.canlet.2013.08.023
[25] X. Zhu, K. Zeng, Y. Qiu, F. Yan, C. Lin, Therapeutic effect of emodin on collagen-induced arthritis in mice, Inflammation 36 (2013) 1253–1259, https://doi.org/10.1007/s10753-013-9663-6
[26] L. Savarino, A. Fioravanti, G. Leo, R. Aloisi, M. Mian, Anthraquinone-2,6-disulfonic acid as a disease-modifying osteoarthritis drug: an in vitro and in vivo study, Clin. Orthop. 461 (2007) 231–237.
[27] J. Nam, D.-W. Seol, C.-G. Lee, G. Wee, S. Yang, C.-H. Pan, Obtusifolin, an anthraquinone extracted from *Senna obtusifolia* (L.) H.S.Irwin & Barneby, reduces inflammation in a mouse osteoarthritis model, Pharmaceuticals 14 (2021) 249, https://doi.org/10.3390/ph14030249
[28] B. Muzellec, M. Teleńczuk, V. Cabeli, M. Andreux, PyDESeq2: a python package for bulk RNA-seq differential expression analysis, Bioinformatics 39 (2023) btad547, https://doi.org/10.1093/bioinformatics/btad547
[29] J. Ranstam, Sampling uncertainty in medical research, Osteoarthr. Cartil. 17 (2009) 1416–1419, https://doi.org/10.1016/j.joca.2009.04.007
[30] J. Ranstam, Reporting laboratory experiments, Osteoarthr. Cartil. 18 (2010) 3–4, https://doi.org/10.1016/j.joca.2009.07.006
[31] A. Turkiewicz, G. Luta, H.V. Hughes, J. Ranstam, Statistical mistakes and how to avoid them – lessons learned from the reproducibility crisis, Osteoarthr. Cartil. 26 (11) (2018) 1409, https://doi.org/10.1016/j.joca.2018.07.017
[32] M.V. Kuleshov, M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S.L. Jenkins, K.M. Jagodnik, A. Lachmann, M.G. McDermott, C.D. Monteiro, G.W. Gundersen, A. Ma'ayan, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic Acids Res 44 (2016) W90–W97, https://doi.org/10.1093/nar/gkw377
[33] Z. Fang, X. Liu, G. Peltz, GSEApy: a comprehensive package for performing gene set enrichment analysis in Python, Bioinformatics 39 (2023) btac757, https://doi.org/10.1093/bioinformatics/btac757
[34] J. Ranstam, L.S. Lohmander, What's in a number or in a picture? Osteoarthr. Cartil. 18 (2010) 1003–1005, https://doi.org/10.1016/j.joca.2010.05.001
[35] J. Ranstam, J.A. Cook, Considerations for the design, analysis and presentation of in vivo studies, Osteoarthr. Cartil. 25 (2017) 364–368, https://doi.org/10.1016/j.joca.2016.06.023
[36] Y. Gorishniy, I. Rubachev, V. Khrulkov, A. Babenko, Revisiting deep learning models for tabular data, in: M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, J.W. Vaughan (Eds.), Adv. Neural Inf. Process. Syst. 34 2021, pp. 18932–18943, https://doi.org/10.5555/3540261.3541708
[37] S.O. Arik, T. Pfister, 2020, TabNet: Attentive Interpretable Tabular Learning, Preprint at https://doi.org/10.48550/arXiv.1908.07442.
[38] S. Popov, S. Morozov, A. Babenko, 2019Neural Oblivious Decision Ensembles for Deep Learning on Tabular Data, Preprint at https://doi.org/10.48550/arXiv.1909.06312.
[39] C. Gao, H. Pu, Q. Zhou, T. Tao, H. Liu, X. Sun, X. He, J. Xiao, Two reactive behaviors of chondrocytes in an IL-1β-induced inflammatory environment revealed by the single-cell RNA sequencing, Aging 13 (2021) 11646–11664, https://doi.org/10.18632/aging.202857
[40] Z. Lu, C. Zheng, P. Ren, J. Gao, C. Zhang, J. Vijg, S. Sun, Comprehensive bioinformatics analysis reveals associations between the DNA damage response and osteoarthritis, Gerontology 71 (2025) 885–898, https://doi.org/10.1159/000547422
[41] G. Pintarelli, A. Dassano, C.E. Cotroneo, A. Galvan, S. Noci, R. Piazza, A. Pirola, R. Spinelli, M. Incarbone, A. Palleschi, L. Rosso, L. Santambrogio, T.A. Dragani, F. Colombo, Read-through transcripts in normal human lung parenchyma are down-regulated in lung adenocarcinoma, Oncotarget 7 (2016) 27889–27898, https://doi.org/10.18632/oncotarget.8556
[42] N.C. Butterfield, K.F. Curry, J. Steinberg, H. Dewhurst, D. Komla-Ebri, N.S. Mannan, A.-T. Adoum, V.D. Leitch, J.G. Logan, J.A. Waung, E. Ghirardello, L. Southam, S.E. Youlten, J.M. Wilkinson, E.A. McAninch, V.E. Vancollie, F. Kussy, J.K. White, C.J. Lelliott, D.J. Adams, R. Jacques, A.C. Bianco, A. Boyde, E. Zeggini, P.I. Croucher, G.R. Williams, J.H.D. Bassett, Accelerating functional gene discovery in osteoarthritis, Nat. Commun. 12 (2021) 467, https://doi.org/10.1038/s41467-020-20761-5
[43] M. Menschikowski, A. Hagelgans, S. Fuessel, O.A. Mareninova, L. Asatryan, M.P. Wirth, G. Siegert, Serum amyloid A, phospholipase A2-IIA and C-reactive protein as inflammatory biomarkers for prostate diseases, Inflamm. Res. 62 (2013) 1063–1072, https://doi.org/10.1007/s00011-013-0665-5
[44] K.B. Heidelberg, W.C. Nelson, J.B. Holm, N. Eisenkolb, K. Andrade, J.B. Emerson, Characterization of eukaryotic microbial diversity in hypersaline Lake Tyrrell, Australia, Front. Microbiol. 4 (2013) 115, https://doi.org/10.3389/fmicb.2013.00115
[45] Z. Wang, J. Petricca, M. Liu, S. Zhang, S. Chen, M. Li, A. Besschetnova, S. Patalano, K. Venkataramani, K.R. Siegfried, J.A. Macoska, D. Han, S. Gao, M. Vedadi, C.H. Arrowsmith, H.H. He, C. Cai, SETD7 functions as a transcription repressor in prostate cancer via methylating FOXA1, Proc. Natl. Acad. Sci. 120 (2023) e2220472120, https://doi.org/10.1073/pnas.2220472120
[46] Y. Fu, J. Chen, X. Zhu, M. Ding, H. Wang, S. Fu, Roles and therapeutic potential of the SLC family in prostate cancer—literature review, BMC Urol. 25 (2025) 32, https://doi.org/10.1186/s12894-025-01714-w
[47] J. Nam, H. Woo, J. Yang, S.J. Kim, K.P. Lee, J.H. Yu, T.J. Park, S. Eyun, S. Yang, Blockade of ZMIZ1–GATA4 axis regulation restores youthfulness to aged cartilage, Adv. Sci. 12 (2025) 2404311, https://doi.org/10.1002/advs.202404311
[48] Z. Li, A. Dai, M. Yang, S. Chen, Z. Deng, L. Li, p38MAPK signaling pathway in osteoarthritis: pathological and therapeutic aspects, J. Inflamm. Res. 15 (2022) 723–734, https://doi.org/10.2147/JIR.S348491
[49] Y. Xu, Y. Gu, W. Ji, Q. Dong, Activation of the extracellular-signal-regulated kinase (ERK)/c-Jun N-terminal kinase (JNK) signal pathway and osteogenic factors in subchondral bone of patients with knee osteoarthritis, 663–663, Ann. Transl. Med. 9 (2021), https://doi.org/10.21037/atm-21-1215
[50] S. Rigoglou, A.G. Papavassiliou, The NF-κB signalling pathway in osteoarthritis, Int. J. Biochem. Cell Biol. 45 (2013) 2580–2584, https://doi.org/10.1016/j.biocel.2013.08.018
[51] L.K. Caesar, N.B. Cech, Synergy and antagonism in natural product extracts: when 1 + 1 does not equal 2, Nat. Prod. Rep. 36 (2019) 869–888, https://doi.org/10.1039/C9NP00011A