Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Functionally decoupled cis-regulatory signals at PCOLCE2 link fibroblast-lineage collagen maturation to pelvic organ prolapse

Suting Li
Renmin Hospital of Wuhan University
Hongming Li
Renmin Hospital of Wuhan University
Bingshu Li
Renmin Hospital of Wuhan University
Zhi Wang
Renmin Hospital of Wuhan University
Jie Min
Renmin Hospital of Wuhan University
Jinhua Tian
Renmin Hospital of Wuhan University
Li Hong
Renmin Hospital of Wuhan University
Songming He
francishsm@qq.com
Renmin Hospital of Wuhan University

Research Article

Keywords: pelvic organ prolapse, PCOLCE2, fine-mapping, SuSiE, colocalisation, cis-eQTL, plasma pQTL, Mendelian randomisation, cell-type-specific regulation, connective tissue genetics

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11780387/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Functionally decoupled cis-regulatory signals at PCOLCE2 link fibroblast-lineage collagen maturation to pelvic organ prolapse

Running title: Cis-regulatory decoupling at PCOLCE2 in prolapse

Authors: Suting Li[1,#], Hongming Li[1,#], Bingshu Li[1], Zhi Wang[1], Jie Min[1], Jinhua Tian[1], Li Hong[1,*], Songming He[1,*]

Affiliations:

[1] Department of Gynecology and Obstetrics, Renmin Hospital of Wuhan University, Wuhan 430060, China

[#] Suting Li and Hongming Li contributed equally to this work.

* Correspondence: Songming He (francishsm@qq.com; ORCID 0009-0000-0104-4433) and Li Hong (drhongli77@163.com; ORCID 0000-0002-3841-5557)

Word count: abstract 417; main text 9452 (Background to Conclusions, excluding Methods); Methods 3933

Abstract

Background. Pelvic organ prolapse (POP) is a common failure of fibrocollagenous pelvic support. Genome-wide association studies (GWAS) report a sub-genome-wide suggestive association at PCOLCE2 (regional minimum p = 2.5×10^{-7}), which encodes a secreted regulator of procollagen C-terminal processing, but they do not establish which cell type, which molecular species or which regulatory signal carries the risk.

Methods. This was a pre-specified study of a five-gene collagen-assembly panel and of one GWAS-suggestive locus, not an unbiased genome-wide screen, and it used public summary-level and single-cell data only; no functional assay was performed. The analysis combined cis-Mendelian randomisation (MR) with fibroblast expression quantitative trait loci (eQTL) instruments, colocalisation, the regional causal-inference method CAUSE, stratified fine-mapping across four molecular phenotypes, plasma protein QTL (pQTL) deconvolution, Steiger filtering, a fifteen-endpoint specificity panel, single-cell and pseudo-bulk transcriptomics, and exploratory regulatory annotation.

Results. Conditional on the sub-genome-wide index signal set out under Background, genetic evidence supports a triangulated statistical causal inference for a fibroblast-lineage cis-regulatory effect at PCOLCE2 that raises POP risk (OR 1.282 per SD, 95% CI 1.112–

1.477), with specificity of this cis-regulatory effect to pelvic-floor endpoints, colocalisation (PP4 = 0.954) and a causal rather than shared-confounding architecture (CAUSE ΔELPD = −7.99 ± 1.64). Within a single haplotype block the locus resolves into two functionally decoupled, statistically inferred cis-regulatory profiles (r^2 = 0.066); the blood-restricted profile controls circulating PCPE-2 but is unrelated to POP, and every variant that raises plasma PCPE-2 falls significantly below the POP effect predicted by a circulating-dosage model. The fibroblast signal also accounts for the locus's bone-density association, which we read as parallel pleiotropy rather than skeletal mediation. PCOLCE2 was expressed selectively in pelvic-floor fibroblasts, and exploratory, under-powered pseudo-bulk analysis (16 versus 5) placed an observed disease-associated expression increment within ECM-remodelling fibroblasts; there the collagen-assembly module was positively but non-significantly displaced (Hedges' g = +0.45, label-shuffling p = 0.38) and only the inhibitor module reached significance, so assembly-biased matrix decay remains a working mechanistic hypothesis.

Conclusions. Conditional on independent POP GWAS replication — the index signal is sub-genome-wide significant and every inference here depends on it — the disease-relevant signal at this locus is a fibroblast-restricted cis-regulatory effect on PCOLCE2 rather than circulating protein dosage. More generally, the locus shows that apparent gene-level pleiotropy at matrix genes can resolve into functionally decoupled, tissue-restricted cis-regulatory profiles within a single haplotype block — a principle that changes how such GWAS signals should be read and followed up.

Keywords: pelvic organ prolapse; PCOLCE2; fine-mapping; SuSiE; colocalisation; cis-eQTL; plasma pQTL; Mendelian randomisation; cell-type-specific regulation; connective tissue genetics

Background

Pelvic organ prolapse is a common consequence of failure of the fibrocollagenous support of the pelvic floor: the anterior vaginal wall, the uterus or the apex descends to or beyond the introitus, and roughly 40% of women are affected to some degree after the menopause [1]. The condition is a leading indication for gynaecological surgery, and recurrence is common among those operated on. Its molecular antecedents, however, remain poorly defined. Its established risk factors — vaginal delivery, parity, ageing, obesity and raised intra-abdominal pressure — all impinge on collagen synthesis, processing and turnover. Failed matrix assembly or repair, rather than a single anatomical defect, is therefore suggested as a final common pathway.

Family and twin analyses indicate a substantial heritable component, and genome-wide association studies have begun to localise it: a meta-analysis of 28,086 European-ancestry cases and 546,291 controls identified 26 loci and 30 independent signals, with connective-tissue, urogenital and cardiometabolic gene sets among the prioritised mechanisms, and reported the PCOLCE2 locus among the POP associations [1]. The only East Asian study published to date (771 Japanese cases and 76,625 controls) identified WT1 and, in cross-ancestry meta-analysis, FGFR2, with concordant effect directions for 21 of 24 European-derived loci [2]. POP genetic architecture is therefore largely shared across ancestries, and matrix genes recur among the associated loci. What these results do not establish is which cell type in the pelvic floor expresses the causal gene, nor which molecular species — transcript, secreted protein or a downstream matrix property — carries the risk signal. Those questions determine whether a locus is a target for mechanistic work at all.

Expression quantitative trait locus (eQTL) resources now allow these questions to be addressed with summary statistics alone. Uniformly processed cis-eQTLs across tissues and cell types are available from the eQTL Catalogue [3] and GTEx [4], and the combination of Mendelian randomisation (MR) [5], colocalisation [6] and fine-mapping [7] supports causal inference at a locus even when the GWAS signal itself is only suggestive; the CAUSE framework further separates expression-mediated causality from a shared latent factor [8]. What this framework cannot do is resolve two regulatory signals within one linkage-disequilibrium (LD) block: where every candidate instrument is a near-perfect proxy of the same lead variant, the classical multi-instrument MR estimators cease to be identifiable and credible sets from single-trait fine-mapping overlap by construction. Dense matrix-gene regions are exactly where this arises, and it must be addressed explicitly rather than assumed away. The distinction at stake is between gene-level pleiotropy and tissue-restricted regulatory pleiotropy, and only the second can be resolved by comparing cell-type-specific eQTL profiles.

PCOLCE2 encodes procollagen C-endopeptidase enhancer 2 (PCPE-2), one of two secreted regulators of the procollagen C-proteinases BMP-1 and the tolloid-like proteinases that excise the C-propeptide of fibrillar procollagens — the step that licenses fibril assembly. PCPE-2 is not functionally interchangeable with PCPE-1. Depending on biological context and protein concentration, PCPE-2 may either enhance procollagen C-terminal maturation or bind and inhibit BMP-1/tolloid-family proteinases. Its net matrix-regulatory effect is therefore stoichiometry-dependent and does not simply increase collagen-fibril assembly activity. This biology makes the gene an attractive candidate for a disorder in which the mechanical competence of collagenous support fails while total collagen content may be preserved or even increased. The present work uses genetic and omics inference rather than wet-

laboratory experiments, so functional validation of the downstream biochemical consequences remains outstanding.

We therefore designed a pre-specified, hypothesis-driven study of five ECM-assembly genes on different chromosomes — PCOLCE2, LOX, LOXL1, TNXB and FBLN5 — each tested with its own cis-eQTL wherever it lies in the genome, asking four questions. Is the genetically predicted expression of any of them associated with POP, and is that association pelvic-floor-specific? Is it causal at the level of a shared variant, and separable from a correlated pleiotropic factor? Which cis-regulatory profile carries the risk — the fibroblast-lineage one, the haematopoietic one controlling circulating PCPE-2, or both? And does patient tissue show the transcriptional correlate of the hypothesised assembly bias, in which fibroblast state? Throughout, we treat the fine-mapping, colocalisation and MR results as statistical inference about regulatory architecture rather than as demonstration of a molecular mechanism, and we report the identification constraints the local LD structure imposes rather than presenting estimators the design cannot support. Two objectives, of deliberately different scope, were fixed in advance. **(i) The primary, biological objective** is to dissect the sub-genome-wide PCOLCE2 locus. **(ii) The secondary, methodological objective** is to assemble a transferable workflow (Box 1) for decomposing regulatory heterogeneity inside matrix-gene loci that dense haplotype blocks make intractable; it is derived from the biological analysis and stress-tested locus-wide, not a substitute for it.

Methods

Study design and data sources

This is a secondary analysis of publicly available summary-level and single-cell data; no new sequencing or primary clinical sample was collected. POP outcome data were taken from FinnGen release R13 (endpoint N14_FEMGENPROL, ICD-10 N81; n = 146,682) [9], and fifteen FinnGen R13 endpoints spanning connective tissue, spinal supportive tissue, the urinary-system endpoint N14_URINOTH and negative controls provided the specificity panel. Fibroblast cis-eQTLs were from the eQTL Catalogue release 8 (dataset QTD000216, GTEx v10 fibroblast reprocessing, 649 donors) [3]; blood cis-eQTLs from eQTLGen phase 1 (31,684 individuals) [10]; independent tissue-level replication from GTEx v8 [4]; plasma pQTLs from three GWAS Catalog studies (deCODE SomaScan GCST90089278, n = 5,361; GCST90426574, n = 3,506; GCST90242374, n = 3,301) [11, 12]; estimated heel bone mineral density (eBMD) from UK Biobank (GCST90726625, n = 448,010); and locus-level reported associations from the GWAS Catalog [11]. Genotype reference data were the 1000 Genomes Project phase 3 panels [13]. Single-cell and bulk pelvic-floor transcriptomes were

GSE331133, GSE151192, GSE151188, GSE53868 and GSE208261. The analysis plan, gene set and endpoint panel were fixed before any MR or colocalisation estimate was computed. All summary statistics were harmonised identically across datasets: variants matched on chromosome, position and alleles; effect alleles aligned to ALT with signs flipped where required; palindromic A/T and C/G variants with minor-allele frequency 0.42–0.58 discarded; and variants absent from the LD reference panel dropped rather than imputed (per-dataset exceptions in Supplementary Note S1). The design has two pre-specified tiers. The first is a **candidate-gene panel**: five ECM-assembly genes spanning three sequential steps — procollagen C-terminal processing (*PCOLCE2*, 3q24), lysyl-oxidase crosslinking (*LOX*, 5q23; *LOXL1*, 15q24) and matricellular glycoprotein assembly (*TNXB*, 6p21; *FBLN5*, 14q32) — each tested with its own cis-eQTL wherever it lies, all five reported whether or not an instrument existed (Supplementary Table S1, Group A; Supplementary Note S1, item 3). The second is a **candidate-region analysis at the PCOLCE2 locus**, which carries a suggestive rather than genome-wide-significant association and is resolved by stratified fine-mapping, colocalisation and CAUSE (Supplementary Table S1, Group B). The panel is therefore not a regional instrument set, and the regional analysis is not a gene-level screen. Neither tier was expanded after the results were seen.

The error rate that matters is therefore the number of tests we declared in advance, not the number a genome-wide study would perform. Nominal p-values are reported uncorrected, with Benjamini–Hochberg adjustment wherever a test family is explicitly defined (the pseudo-bulk modules, the cross-phenotype panel and the five-gene panel), and none should be read as genome-wide-corrected evidence. All sensitivity analyses reported below were specified as robustness checks on those primary analyses rather than as exploratory searches.

Analysis status. Subsections below are marked by status. *Pre-specified primary analyses* were fixed before any estimate was computed: the five-gene panel, the PCOLCE2 regional chain (cis-MR, colocalisation, CAUSE, stratified fine-mapping, conditional analyses, the pQTL dose–response test), the fifteen-endpoint specificity panel, the cross-tissue sensitivity analyses and the locus-wide matrisome scan. *Exploratory post hoc analyses* were added while interpreting those results and are reported as such: the TEAD motif and chromatin annotation, the spatial transcriptomics, the circulating-metabolite screen and the pseudo-bulk module analysis. The distinction matters for multiple-testing interpretation, and no exploratory analysis enters the primary inference chain.

Statistical analyses: cis-MR and instrument quality control

The overall design and analytical flow are summarised in Fig. 1. Genetically predicted gene expression was the exposure and FinnGen R13 disease endpoints [9] the outcomes. Where

a single cis-eQTL survived instrument selection, effects were estimated by the Wald ratio with a delta-method standard error; where multiple independent instruments were available, fixed-effect inverse-variance weighting was applied [5] (Supplementary Fig. S1; Supplementary Table S1). The Wald ratio assumes that the cis-variant influences the outcome only through the target transcript. With a single instrument that exclusion-restriction assumption cannot be tested internally, so it is addressed by triangulation — colocalisation, CAUSE, the fifteen-endpoint specificity panel and the locus-wide scan — rather than by pleiotropy-robust estimators, none of which is identified in this LD geometry. As a power complement, every panel gene was additionally tested with whole-blood cis-eQTL instruments from eQTLGen under the same rules (Supplementary Fig. S2C); the TNXB blood instrument was excluded for the MHC-region reason given in Results. Panel conclusions were additionally tested for dependence on the tissue supplying the instrument (ten GTEx v10 tissues plus eQTLGen blood, with Cochran's Q, descriptive random-effects pooling and leave-one-tissue-out; Supplementary Fig. S18; Supplementary Table S31; Supplementary Note S4). Instrument strength was quantified with the first-stage F statistic ($F = 2z^2$); the fibroblast lead instrument for PCOLCE2 (rs55959862) yielded F = 23.8 and the blood lead (rs4683702) F = 63.3, both above the conventional threshold of 10. As a specificity complement, the identical workflow was applied locus-wide: every core-matrisome gene (GO:0031012 or GO:0005201, plus the five panel genes) within ±1 Mb of the 36 POP lead variants and of the PCOLCE2 lead was taken through instrument construction, identifiability assessment, MR, credible-set comparison, colocalisation and Steiger filtering, with Benjamini–Hochberg correction across testable pairs (Supplementary Fig. S19; Supplementary Table S32; Supplementary Note S5). The fine-mapped credible variant rs73234745 carries an indistinguishable fibroblast effect (β −0.267 versus −0.262 per allele) and is used interchangeably with rs55959862. Steiger directionality filtering was applied to every instrument–outcome pair (559/559 passed), and for the PCOLCE2 locus the variance explained in expression ($r^2 = 0.0353$) exceeded that attributable to POP liability by a factor of approximately 115.

Colocalisation and regional causal inference

We performed colocalisation between the eQTL Catalogue [3] fibroblast cis-eQTLs and FinnGen POP summary statistics over ±1 Mb windows using SuSiE [7] fine-mapping and coloc [6] with default priors, reporting the posterior probability of a single shared causal variant (PP4); prior sensitivity was assessed by varying p_{12} from 1×10^{-6} to 1×10^{-4} (gene-set and prior grids in Supplementary Tables S7 and S9). Because the plasma pQTL and eBMD datasets each contain more than one independent signal in this region, the single-causal-variant assumption of coloc-ABF is violated for those comparisons; their posteriors are

therefore corroborative only, and the stratified SuSiE credible sets carry the primary inference. To distinguish expression-mediated causality from a shared latent factor driving both the eQTL and the GWAS signals, we applied CAUSE [8] to the PCOLCE2 and LOXL1 regions, pruning cis-eQTLs within ±300 kb to $r^2 < 0.1$ (1000 Genomes EUR [13]; 71 and 95 independent variants) and comparing the causal and sharing models by in-sample expected log predictive density (ELPD) with the loo package (parameters and sensitivity analyses in Supplementary Table S2). CAUSE distinguishes expression-mediated causality from shared-factor confounding under its own modelling assumptions — a Gaussian liability model in which a single latent factor carries correlated pleiotropy — and is sensitive to variant-set choice and to LD architecture. It is therefore interpreted together with MR and colocalisation rather than in isolation, and cannot by itself exclude local multi-variant confounding. Support for the inferred relationship relies on triangulation across all three analytical approaches.

To examine whether the credible variant sets are phenotype-specific within the shared haplotype block, we performed stratified SuSiE fine-mapping: susie_rss was run separately on four trait-level regional summary-statistic sets — FinnGen POP (n = 146,682; genomic-control-corrected z reported alongside the raw z), the fibroblast cis-eQTL (n = 966), the deCODE plasma pQTL (n = 5,361) and UK Biobank heel eBMD (n = 448,010) — over a common chr3:142.4–143.4 Mb window, with the same 1000 Genomes EUR panel for LD and each dataset harmonised to the ALT allele. Purity, size, prior variance and membership of all 95% credible sets were compared across traits (Supplementary Fig. S9; Supplementary Table S17). Credible sets are statistical sets of variants: they express posterior support under the assumed priors and LD, and are not equivalent to biologically independent causal variants — a distinction that matters here, where the two sets are non-overlapping but a single unassayed variant could in principle underlie both. The result was stress-tested across ten settings (Supplementary Table S25): separation was preserved in every one (Jaccard = 0 throughout), and rs55940906 and rs147652399 remained inside their sets; only under the most extreme LD shrinkage did rs73234745 lose top rank, where all credible sets degrade together as expected. Size and purity therefore depend on LD-panel quality, whereas lead-variant identity and cross-trait separation do not. Recomputing LD from the 1000 Genomes phase 3 v5b full call set (EUR, n = 503) left lead-variant PIPs consistent (rs73234745, 0.975; rs55940906, 1.000) and cross-signal PIPs negligible (0.006 and 0.008). Conditional SuSiE and conditional colocalisation then fitted each trait's credible-set lead as a fixed effect in the other trait's model (Supplementary Table S28).

Cross-tissue and cross-phenotype MR and haplotype dissection

The same fibroblast instrument was tested against fifteen FinnGen R13 endpoints spanning abdominal-wall and vascular connective tissue, spinal supportive tissue and negative-control

traits. To disentangle the two functionally decoupled cis-regulatory profiles of PCOLCE2, pairwise LD between the fibroblast lead (rs73234745/rs55959862) and the blood lead (rs4683702) was computed from the 1000 Genomes EUR quality-controlled haplotype panel (n = 489), and independently rechecked through the Ensembl LD REST interface across five phase-3 European panels (r^2 = 0.098 EUR, 0.129 FIN, 0.076 CEU, 0.101 TSI, 0.169 GBR; $|D'| \geq 0.999$ in every panel; Supplementary Table S3). FIN was included because FinnGen R13 is a Finnish cohort, making it the most outcome-matched reference panel available; the weak-LD conclusion is therefore not an artefact of panel choice.

Because a multivariable MR is not identified with one instrument per haplotype, a locus-level pleiotropy scan was performed instead: all 30 GWAS Catalog [11] associations mapped to PCOLCE2 were collected, and the LD (r^2) of each lead variant against both haplotype tools was computed from the same panel. The blood haplotype accounts for the locus's reported associations with circulating proteins (PCOLCE2 and the neighbouring ECE1), lipids, total cholesterol and pulse pressure, whereas the fibroblast haplotype shares its signal with heel bone mineral density (r^2 = 1.00), bone density (r^2 = 0.98) and, in a separate signal, ECE-1 protein levels (r^2 = 0.57); a further, stronger bone-density signal at the same locus ($p = 1\times10^{-300}$) is not in LD with either tool. No association with body-mass index, parity or oestrogen-related traits has been reported at this locus (Supplementary Fig. S5; Supplementary Table S14).

Because the fibroblast haplotype shares the locus's heel bone mineral density signal, that association was followed up directly: regional summary statistics for eBMD (UK Biobank, GWAS Catalog GCST90726625, n = 448,010, GRCh38, harmonised and tabix-indexed) were retrieved for the PCOLCE2 interval, allele-harmonised to the fibroblast and POP reference panels, and colocalised against both with the same implementation in a ±200 kb window around the fibroblast lead (window sensitivity ±10–200 kb; prior sensitivity p_{12} = 10^{-6}–10^{-4}). Cross-tissue eQTL effect profiles were assembled from the eQTL Catalogue (fibroblast, uterus, vagina, skin, skeletal muscle) and eQTLGen whole blood. To exclude the possibility that the cis signal regulates a neighbouring gene rather than PCOLCE2, all 20 genes within chr3:142.30–143.20 Mb were tested for fibroblast cis-eQTL inside the lead block (rs73234745 ± 150 kb; 9,692 gene × variant tests, Bonferroni-corrected), and the block lead of every gene was LD-checked against rs73234745 and rs55940906 (Supplementary Fig. S13; Supplementary Table S26).

Definition of the two profiles. We refer to two *functionally decoupled* cis-regulatory **profiles**, each carried by a distinct haplotype. Throughout, “haplotype” denotes the genotype only and “signal” denotes a statistical association in a GWAS or eQTL dataset; “profile” denotes the inferred pattern of cell-type-restricted cis-regulation that a set of such signals

supports. The evidence for decoupling is the tissue-exclusive eQTL profiles and the opposing haplotype-level disease associations, not a direct demonstration of two independent cis-regulatory elements, and not a claim that two independent causal variants have been identified. The weak LD (r^2 = 0.066 in the primary panel; r^2 = 0.076–0.169 across phase-3 populations) and the tissue-exclusive expression signals support the interpretation that the two variants mark distinct, tissue-specific regulatory effects rather than serving as redundant tags of one signal; however, we cannot exclude that a single, unassayed causal variant underlies both profiles. For the MR framework the two variants were analysed as separate single-instrument models (one per haplotype) rather than as a joint two-instrument model, a choice justified by their weak linkage and tissue-exclusive expression effects.

Single-cell and pseudo-bulk transcriptomics

Single-cell RNA-seq data of anterior vaginal wall from postmenopausal women with uterovaginal prolapse, with or without local oestrogen treatment (GSE331133; 10x Genomics, 40,412 annotated cells; sample metadata in Supplementary Table S5), were used to map PCOLCE2 expression across lineages. GSE331133 lineage labels are marker-based rather than reference-mapped (fibroblasts 1,756/40,412 cells; per-sample fractions 0.5–11.2%; elevated mitochondrial fractions), so no composition-level statement is inferred from this dataset; its contribution is restricted to expression localisation. For patient-tissue pathway evidence we reprocessed the GSE151192 single-cell prolapsed anterior vaginal wall cohort (as deposited, “human vaginal wall”; 16 POP versus 5 controls): cells were assigned to the fibroblast compartment and counts aggregated to pseudo-bulk samples (317–6,313 fibroblasts per profile, median 2,726; per-sample values in Supplementary Table S5). Counts were normalised to counts per million and log-transformed; module scores were the mean gene-wise z-score across four pre-defined modules — collagen fibril **assembly** (31 genes curated, 28 detected), matrix **degradation** (31 and 24), **inhibitors** (TIMPs, SERPINE1) and **housekeeping** — curated from UniProt and REACTOME collagen-fibril-organisation components and prior pelvic-floor ECM literature (gene lists in Supplementary Table S6). Case–control differences were tested with Welch's t-test and a label-shuffling permutation test (10,000 permutations of donor labels); the paired within-patient comparison used a within-pair sign-flip permutation (10,000 sign flips). Effect sizes are Hedges' g with Benjamini–Hochberg adjustment across modules within each dataset. The two null distributions are never pooled: every p-value quoted below is labelled with the test that produced it. An analogous paired analysis of bulk anterior vaginal wall biopsies from prolapsed versus non-prolapsed (precervical) sites within the same patients (GSE53868, 12 paired) was performed for comparison. For this two-colour Agilent cohort with a balanced dye-swap design (opposite dyes within each patient, direction swapped across six patients),

each log2 intensity was regressed per feature on site and dye direction, and module analyses were run on both uncorrected and dye-corrected intensities; because the swap is balanced, dye effects bias neither the paired mean nor the module direction — they only inflate the variance of within-pair differences, so uncorrected results are conservative.

Statistical boundaries of the pseudo-bulk analysis. This pseudo-bulk analysis is a subcompartment-level pathway-score permutation test based on 21 aggregated samples (16 POP versus 5 controls; 10,000 label permutations), and is not a conventional bulk RNA-seq differential-expression analysis of independent tissue samples; donor-level aggregation and label permutations were used to mitigate pseudo-replication arising from cell-level autocorrelation. The contrasts are not driven by the smallest fibroblast compartment (donor N4, 317 fibroblasts): leave-one-donor-out refitting (21 fits) left the direction of every module contrast unchanged (assembly effect-size range +0.13 to +1.06, always positive), and excluding N4 specifically gave an assembly Hedges' g of +0.25 (Supplementary Table S11). Stability across refits shows that no single donor drives these estimates; it is not evidence that any effect is non-zero. The paired bulk analysis of a separate anterior vaginal wall cohort (GSE53868; 12 patients, each contributing a prolapsed and a non-prolapsed precervical site) is provided as an independent observational comparison, not as replication.

Precision and robustness. Because the permutation test evaluates module specificity against random gene sets rather than sampling variance, we additionally quantified the precision of each effect: (i) donor-level bootstrap confidence intervals (10,000 stratified resamples within group); (ii) leave-one-donor-out refitting (21 fits) to assess directional stability; (iii) three scoring definitions — gene-wise z-score mean (primary), within-sample rank mean and unstandardised log2-CPM mean — to test dependence on score construction; and (iv) the minimum detectable standardised effect at 16 versus 5 samples ($\alpha = 0.05$ two-sided, 80% power; $|g| = 1.38$). Results are reported in Supplementary Table S11.

Fibroblast subtypes were resolved by unsupervised clustering of the same GSE151192 fibroblast compartment (61,962 quality-controlled fibroblasts). A stratified subsample (500 cells per donor; 10,318 cells) was log-normalised, reduced to 1,200 highly variable genes and clustered by principal components (30 components) and Leiden; all remaining cells were assigned by nearest-centroid projection in the same principal-component space. Clusters were annotated by cross-cluster z-scores of four marker modules into four canonical stromal states — quiescent/resting, ECM-remodelling, senescent/SASP-like and inflammatory — and the disease-related PCOLCE2 increment was decomposed as $\Delta(\text{mean}) = \Sigma_k \Delta f_k \cdot \tilde{m}_k$ (composition) + $\Sigma_k \tilde{f}_k \cdot \Delta m_k$ (within-subtype expression), with donor-level Welch tests (16 POP versus 5 controls). GSE151192 carries no genotype, so no subtype-level (sc-eQTL) association between the risk allele and expression could be estimated (see Limitations).

Exploratory regulatory annotation of the lead variant

As an exploratory analysis, and not as a mechanism claim, the PCOLCE2 lead region was annotated with ENCODE candidate cis-regulatory elements (cCREs), Roadmap Epigenomics 25-state ChromHMM segmentations of five epigenomes and 100-way vertebrate conservation. Allele-specific transcription-factor motif effects were assessed by scanning ±30 bp around each variant on both strands with TEAD1–TEAD4 position-weight matrices (JASPAR CORE), with empirical p-values from the distribution of PWM-score changes across 469 regional variants. The pipeline was calibrated against experimentally verified TEAD4 binding sites: 264 pooled IDR-thresholded TEAD4 ChIP-seq peaks (ENCODE ENCSR285HHZ) on chromosome 3 carried higher best-strand TEAD PWM scores than GC-matched background (AUC 0.53–0.58), confirming that the scan recovers known TEAD-occupied sites while illustrating its limited discriminative power. No other transcription-factor family was scanned (epigenome identifiers in Supplementary Table S21; full results in Supplementary Note S2).

The annotation was extended along two axes. First, chromatin accessibility: ATAC-seq or DNase-seq IDR peak sets for six cell types — four fibroblast-lineage (IMR-90, BJ, primary dermal fibroblast, abdominal skin fibroblast) and two non-fibroblast references (GM12878, HUVEC) — were retrieved from the ENCODE portal (GRCh38) and intersected with seven key variants (the two fibroblast eQTL leads, the eBMD fine-mapped variant rs147652399, the pQTL top variant rs55940906 and the three blood tags), reporting peak overlap, distance to the nearest peak and ±5 kb peak density. Second, chromatin state: the 25-state Roadmap segmentations of the five epigenomes above were intersected with the same variants. The ENCODE cCRE registry itself was not reachable from our environment at execution time; per-cell-type accessibility peaks and chromatin states provide the same cell-type-specific regulatory evidence.

Exploratory spatial transcriptomics (supplementary only)

As a descriptive localisation analysis, kept outside the causal inference chain, the Visium HD spatial transcriptomes of anterior vaginal wall (GSE331133) were analysed for the two available sections — one untreated (POP7-2) and one locally oestrogen-treated (E2_6) donor — at 16 µm bin resolution. The HAS1-positive/ESR1-positive stromal niche was defined on 64 µm super-bins with ±160 µm neighbourhood smoothing, using a ninetieth-percentile expression threshold among expressing bins only, and ACTA2-positive and RGS5-positive bins by the same rule. Enrichment of PCOLCE2-positive bins was tested as an odds ratio with within-section label permutations, and proximity as median bin-to-niche distance against random placements. With one section per condition the analysis is purely

descriptive and no between-group inference is attempted (bin depths, permutation counts and sensitivity analyses in Supplementary Note S3; Supplementary Fig. S16; Supplementary Table S30).

Plasma protein colocalisation and protein-level Mendelian randomisation

To test whether the POP association is mediated by circulating PCOLCE2 protein rather than by fibroblast-lineage expression, we used three public plasma pQTL studies of PCOLCE2 identified in the GWAS Catalog [11]: GCST90089278 (deCODE SomaScan, 5,361 Icelandic individuals; full summary statistics downloaded as a regional slice of chr3:141.6–143.4 Mb) [12], GCST90426574 (3,506 European individuals) and GCST90242374 (3,301 European individuals; SomaScan aptamer PCOLCE2.6081.52.3). The deCODE slice (3,026 variants, GRCh37) was lifted to GRCh38 with the UCSC hg38ToHg19 over-chain file, validated by exact coordinate recovery of the pQTL lead variant rs11716897. deCODE serum protein levels were Box-Cox transformed before GWAS [12]; the reported effect sizes are expressed per standard deviation of the transformed protein level, which we verified empirically rather than assumed: the standard errors of all 3,026 regional variants match the unit-variance expectation (median observed/expected ratio 0.97, 100% within 10%), the small sub-unity offset being consistent with covariate adjustment. All protein-effect estimates are therefore interpretable as per-SD effects on this scale. Pairwise LD was computed from the 1000 Genomes EUR phase-3 panel (n = 489) [13]. Protein-level MR used per-variant Wald ratios (β_POP/β_pQTL) harmonised on alleles with delta-method standard errors; because all regional variants are in strong mutual LD, ratios are reported descriptively for the haplotype-defining variants and are not pooled into an inverse-variance estimate (single-locus instrument; MR-PRESSO and leave-one-out are not applicable, consistent with the single-instrument design used throughout). Colocalisation between the plasma pQTL and the fibroblast cis-eQTL used the same coloc-ABF implementation as above; window sensitivity was assessed by contracting the shared region to ±200, ±100, ±50, ±25 and ±10 kb around the fibroblast lead, and prior sensitivity (p_{12} = 1×10^{-6} to 1×10^{-4}) left PP3 between 0.88 and 0.91.

Variant-level dose–response discordance test

The circulating-dosage model was tested directly by a dose–response concordance test. The POP log OR predicted for each pQTL-defining variant is (β_POP/β_pQTL)_reference × β_pQTL, where the reference ratio is the per-SD plasma-protein effect estimated from the fibroblast lead; the deviation (observed minus predicted) was tested with a delta-method standard error carrying the uncertainty of both stages. Under the circulating-dosage hypothesis a variant that raises circulating PCOLCE2 while leaving POP risk unchanged falls

significantly below the predicted line. The test's null is proportional mediation, and a significant negative deviation rejects that model rather than quantifying the effect of the protein.

Exploratory screen for circulating collagen-metabolite mediation

To test whether the association could act through a systemic route rather than within the stroma, we screened four collagen-related circulating metabolites — proline, citrulline, glycine and *trans*-4-hydroxyproline — against the same 32 near-perfect regional proxies, in public metabolite GWAS panels, applying the study-wide Bonferroni threshold and quantifying each panel's minimum detectable effect before interpretation. The screen was pre-specified as a negative-control check for systemic mediation, not as a discovery scan; because no association survived correction it is interpreted qualitatively in Discussion rather than tabulated.

Statistical identification constraints imposed by the haplotype-block LD structure

As a robustness check on the single-instrument design, a robust adjusted profile score estimator (MR-RAPS [8]) was fitted to the 32 near-perfect regional proxies ($r^2 \geq 0.96$ with the lead) in both its standard and LD-aware forms; because the proxies form a rank-1 LD block the correlated version is not identifiable, and the standard version serves only as a concordance check of the primary Wald ratio (Supplementary Table S24). Conditional SuSiE and conditional colocalisation verified that no second independent shared signal is hidden within the block (Supplementary Table S28).

The adequacy of every MR sensitivity estimator was made explicit rather than assumed (Supplementary Table S24): weighted median, MR-Egger, simple and weighted mode require at least three near-independent instruments and MR-PRESSO at least four, so none is identifiable in a single-instrument design, and LD-aware generalised-least-squares estimators are numerically ill-conditioned inside a single haplotype block. All 32 fibroblast cis-eQTL variants with $p < 1\times10^{-4}$ that are present in the 1000 Genomes EUR panel are in near-perfect LD with the lead ($r^2 \geq 0.96$) and cannot serve as independent instruments. Three analyses that are identified in this design were performed instead — a proxy-concordance panel across all 32 near-perfect proxies, inverse-variance weighting across the two near-independent haplotype tags (rs55959862 and rs73234745, $r^2 = 0.10$) and leave-one-tag-out — with estimates reported in Results and in Supplementary Table S24. The roles otherwise played by MR-Egger and MR-PRESSO are filled by CAUSE (correlated pleiotropy), Steiger directionality, the fifteen-endpoint specificity panel with negative controls, and the dose–response discordance test.

This failure is a property of the LD geometry rather than of the data, and is therefore stated as a design constraint: where every candidate instrument is a near-perfect proxy of the lead, the classical robust MR estimators are not estimable at any sample size, and reporting them would require imposing arbitrary external assumptions. Supplementary Fig. S12 overlays the eQTL, pQTL and eBMD credible-set memberships on the regional LD heatmap and shows the two lead variants (rs73234745, rs55940906; pairwise $r^2 = 0.068$) sitting in locally distinguishable signal clusters of one and the same $D' \approx 1$ block — the geometric basis on which the two-instrument inverse-variance analysis and the stratified SuSiE analyses are identified.

Software, package versions and data provenance

Colocalisation used a custom coloc-ABF implementation validated numerically against the published algorithm; because no C toolchain was available, the Rcpp likelihood kernels of cause were re-expressed in pure R, with behavioural equivalence verified by simulation (5/5 architectures assigned correctly). Analysis scripts are archived on Zenodo (see Data availability). Software and package versions, release versions and download dates for every dataset, together with the repository, the tissue or cell model and the use made of each resource, are itemised in Supplementary Table S27; the dataset screening log is Supplementary Table S10 and the pipeline structure Supplementary Fig. S14.

Results

Fig. 1 summarises the design and analytical flow; Fig. 2 the colocalisation and CAUSE results, Fig. 3 the tissue-restricted cis-eQTL architecture, Fig. 4 the single-cell localisation, Fig. 5 the pseudo-bulk module analysis and Fig. 6 a working mechanistic hypothesis. The integrated two-signal model behind Fig. 6 is Supplementary Fig. S17.

A candidate-gene cis-MR screen identifies PCOLCE2 as a POP risk factor

One caveat governs everything that follows: the POP signal at this locus is sub-genome-wide suggestive (regional minimum $p = 2.5\times10^{-7}$), not a genome-wide significant GWAS hit, and every estimate reported below is conditional on that index signal (Table 1, first row; Supplementary Table S1). Subject to that condition, of the five predefined ECM-assembly genes only PCOLCE2 yielded a fibroblast cis-instrument of adequate strength with a concordant POP association. Genetically predicted fibroblast PCOLCE2 expression was associated with higher POP risk (OR 1.282 per SD, 95% CI 1.112–1.477, $p = 5.97\times10^{-4}$; $F = 23.8$) — a triangulated statistical causal inference rather than a demonstrated mechanism; Steiger filtering was consistent with the exposure-to-outcome direction (559/559 pairs

passed; quantitative detail in Methods). Steiger filtering excludes only extreme reverse causation and has limited power against weak reverse effects, so directionality here rests on the triangulation described below rather than on Steiger alone. LOXL1, by contrast, showed the opposite pattern in a different connective tissue: its fibroblast instrument (F = 90, a genome-wide significant instrument) associated with abdominal aortic aneurysm (OR 0.81, FDR 0.037) but was null for POP, indicating that a single matrix gene can carry endpoint-specific effects in different supportive tissues rather than a uniform connective-tissue liability.

Because the PCOLCE2 instrument is a single variant, the robustness of the estimate was assessed with the three analyses that remain identified in this design rather than with the standard multi-instrument suite (Methods). All 32 near-perfect proxies ($r^2 \geq 0.96$ with the lead) gave Wald ratios within 0.246–0.288 (median OR 1.299, 95% CI 1.108–1.523); inverse-variance weighting across the two near-independent haplotype tags ($r^2 = 0.10$) reproduced the primary estimate (OR 1.283, 95% CI 1.160–1.419) with no evidence of heterogeneity ($Q = 5\times10^{-4}$, $p = 0.98$, $I^2 = 0\%$); and leave-one-tag-out left the two estimates within 0.4% of each other (Supplementary Fig. S11; Supplementary Table S24). MR-RAPS [8] on the same proxies returned a concordant point estimate ($\beta = 0.264$, OR 1.302); because the proxies are perfectly correlated its nominal p-value is invalid and is not used anywhere, and the LD-aware version is not identifiable in this rank-1 block (Supplementary Table S24). Weighted median, MR-Egger, mode-based estimators and MR-PRESSO are not identifiable with one independent instrument (Methods); the questions they would address are covered by CAUSE, Steiger filtering, the fifteen-endpoint panel and the dose–response test below.

The remaining panel members are reported in full in Supplementary Table S1, Group A; instrument availability rather than effect size determined what could be tested. TNXB lies inside the MHC region (6p21): it has no fibroblast instrument (regional minimum $p = 2.4\times10^{-3}$), and although a strong two-variant whole-blood cis-instrument exists (eQTLGen), it was excluded from causal interpretation because the same instrument associates at least as strongly with type 2 diabetes ($p = 1.1\times10^{-8}$) and other non-connective endpoints — the pattern expected when a blood-eQTL signal inside the MHC is dominated by pleiotropy rather than by TNXB biology (Supplementary Fig. S2C). LOX and FBLN5 had no fibroblast instrument either (minimum $p = 1.9\times10^{-5}$ and 7.6×10^{-4}) and were analysed with whatever instrument existed — skin and vagina in the connective-tissue panel, whole blood as the power complement. LOX was protective for POP in the skin analysis (IVW OR 0.79, 95% CI 0.74–0.85, $p = 2.5\times10^{-10}$), but the same instrument pointed the same way for hernia (OR 0.89, $p = 6.2\times10^{-9}$), abdominal hernia (0.88, $p = 1.6\times10^{-4}$) and constipation (0.94, $p = 0.006$), and colocalisation between the fibroblast LOX eQTL and POP was unsupportive (PP4 = 0.05), although the weak fibroblast cis-eQTL limits that comparison. The whole-blood

instruments point the same way: LOX was protective for POP (OR 0.43, $p = 3.6\times10^{-4}$) and for hernia (OR 0.55, $p = 1.7\times10^{-5}$; single-variant Wald ratios), the FBLN5 blood instrument associated with abdominal aortic aneurysm (OR 1.17, $p = 1.2\times10^{-3}$) but not with POP (p = 0.13), and FBLN5 was null in the vaginal analysis (OR 1.02, 95% CI 0.89–1.18, p = 0.75). Across skin, vagina and blood these are the profiles of generalised connective-tissue axes rather than pelvic-floor-specific ones, and none of these estimates is interpreted causally (Supplementary Fig. S2C). A cross-tissue re-analysis of the panel (ten GTEx v10 tissues plus eQTLGen blood) excluded a tissue-of-instrument artefact: LOX was protective in every source with an instrument but with heterogeneous magnitude, FBLN5 was null and homogeneous, LOXL1 was null apart from one nominally non-null uterine estimate that does not survive the twelve-test family threshold (0.0042), and PCOLCE2 has no non-fibroblast GTEx instrument (eQTLGen blood OR 1.10, 95% CI 0.79–1.52, imprecise) and is excluded from Supplementary Fig. S18 (Supplementary Table S31; Supplementary Note S4).

Applying the Box 1 workflow locus-wide returns no second matrix-gene candidate

To test whether this chain is specific to PCOLCE2 rather than a property of any matrix gene at a reported POP locus, we applied the same pre-specified workflow (Box 1) to all 24 locus–gene pairs formed by the core matrisome (GO:0031012 or GO:0005201, plus the five panel genes) within ±1 Mb of the 36 genome-wide-significant POP lead variants and of the PCOLCE2 lead itself (Supplementary Table S32; Supplementary Fig. S19; Supplementary Note S5). Five genes are not measured as fibroblast eGenes, and twelve have fibroblast cis-eQTLs but no instrument at either the genome-wide or the suggestive tier (including TNXB, concordant with the panel result above); the remaining seven pairs are all single-instrument ($r^2 = 1.00$ among candidate instruments, so no multi-instrument estimator is identifiable). Of these seven, only PCOLCE2 completes the chain: OR 1.28 per SD (95% CI 1.11–1.48, $p = 5.97\times10^{-4}$, BH q = 0.0042) with PP4 = 0.95 and complete overlap between the eQTL and POP credible sets (30/30 variants). The next-ranked estimate, ALPL at the WNT4 locus, is nominally non-null (OR 1.07, 95% CI 1.01–1.14, p = 0.025) but does not survive correction (q = 0.087) and shows no colocalisation ($PP4 = 1.1\times10^{-6}$; credible-set overlap 0/4) — a point estimate without variant-level support. The remaining five estimates are null (all $p \geq 0.11$) with $PP4 \leq 0.037$, and all seven estimates are Steiger-consistent (expression variance exceeds POP liability variance at the instrument). The scan therefore returns exactly one matrix gene with joint MR and colocalisation support — the panel's own primary hypothesis. The screen is conditioned on a high-quality fibroblast cis-eQTL instrument: it cannot rule out matrix-gene effects mediated through other cell types, and bounds the fibroblast-lineage claim rather than every possible route from these loci to POP.

The POP association colocalises and is consistent with causality rather than shared confounding

The fibroblast PCOLCE2 cis-eQTL and FinnGen POP summary statistics were best explained by a single shared causal variant (PP4 = 0.954; Fig. 2B), and the posterior was stable when the shared-prior probability was varied (PP4 = 0.95 at $p_{12} = 1\times10^{-5}$; Supplementary Tables S7 and S9). Tissue context was informative: in uterus and vagina, the same analysis returned PP4 ≈ 0.044, so the colocalisation is a property of the fibroblast compartment rather than of pelvic-floor tissue generically. The pQTL and eBMD layers analysed below each carry more than one regional signal and therefore violate the single-causal-variant assumption of coloc-ABF; their posteriors are corroborative, and the stratified credible sets carry that claim.

CAUSE separated expression-mediated causality from a shared latent factor: the causal model outperformed the sharing model (ΔELPD = −7.99 ± 1.64; z = −4.88, one-sided p = 5.2×10^{-7}; Fig. 2A), with a small estimated expression-mediated effect (γ = 0.276, 95% CI 0.250–0.306) and a sharing proportion q = 0.05; an equivalent fit would require 80% of variants to act through a shared factor (Fig. 2C; Supplementary Table S2). LOXL1 provided an internal negative control: at that region CAUSE favoured sharing (ΔELPD = +0.52), consistent with its colocalisation result (PP3 = 0.957, i.e. distinct causal variants), and LOX could not be tested in this framework because it has no fibroblast instrument (regional minimum p = 1.9×10^{-5}). The CAUSE conclusion was insensitive to the variant set used for posterior comparison (all SNPs, z = −7.29; p < 0.05 subset, z = −4.40). CAUSE separates expression-mediated causality from correlated pleiotropy only under its own Gaussian-liability, single-latent-factor modelling assumptions, and is sensitive to variant-set choice and to LD architecture. Model preference from CAUSE alone does not constitute biological proof; causal-inference conclusions rely on triangulation across MR, colocalisation and fine-mapping rather than on CAUSE output in isolation (Methods).

Two functionally decoupled cis-regulatory profiles within one haplotype block underlie a tissue-restricted genetic association

Throughout this section, “decoupled cis-regulatory profiles” denotes distinct tissue-restricted eQTL–disease association profiles inferred from summary statistics. The analysis does not constitute experimental proof of two physically separate cis-regulatory DNA elements; that demonstration would require reporter-assay or epigenome-editing experiments. The two lead variants lie 2.36 kb apart and are only weakly correlated (r^2 = 0.066, |D′| = 0.818 in the primary panel), a pattern reproduced across phase-3 European panels (r^2 = 0.076–0.169, |D′| ≈ 1 throughout). The combination of weak r^2 with high D′ arises from the difference in minor-

allele frequency between the two variants, and means that the two are not interchangeable tags: the informative evidence is the decoupling of their effects, not the recombination distance between them.

That decoupling is visible across tissues. The fibroblast lead rs73234745 is a strong cis-eQTL in cultured fibroblasts (β = −0.267 per C allele, z = −4.87, p = 1.1×10^{-6}) and in skin (β = −0.206, p = 0.0012), but lies outside the significant set in whole blood; conversely, the blood lead rs4683702 is a strong blood cis-eQTL (z = 7.95, p = 1.8×10^{-15}) and is also active in uterus (β = +0.229) while null in fibroblasts (Fig. 3). The mutual exclusivity of the two profiles was reproduced independently in GTEx v8 (fibroblast normalised effect size −0.175 and −0.158 for rs73234745 and rs55959862, significant, with whole blood not significant; rs4683702 significant in whole blood, +0.123, and not significant in fibroblasts), with the same sign as the eQTL Catalogue estimate in cultured fibroblasts (Supplementary Fig. S6).

The disease side follows the same partition. Using the fibroblast instrument (F = 23.8), the strongest association outside POP was with the broad FinnGen urinary-system endpoint N14_URINOTH (OR 1.115, 95% CI 1.062–1.169, p = 9.0×10^{-6}), with a borderline concordant effect on lumbar disc prolapse (OR 1.073, p = 0.076) and null effects on every hernia, aneurysm, diverticular, venous and negative-control endpoint (all |z| < 1.3). The blood-active haplotype, which drives the same gene in a lineage irrelevant to pelvic-floor integrity, was associated with neither POP nor the urinary-system endpoint (OR 1.078, 95% CI 0.924–1.256, p = 0.34). Two qualifications matter. First, rs4683702 also regulates the neighbouring genes PAQR9 and TRPC1, so its null result is an observational contrast rather than a causal attribution to PCOLCE2 alone. Second, N14_URINOTH denotes “other diseases of urinary system” (107,917 cases of 500,186), which includes but is not restricted to urinary incontinence; FinnGen R13 has no incontinence-specific endpoint (Supplementary Table S8). PCOLCE2 cis-expression is therefore a pelvic-floor-restricted susceptibility factor acting through fibroblast-lineage expression, not a systemic connective-tissue weakness allele (Fig. 3; Supplementary Figs. S2 and S6; Supplementary Table S4). These data define two functionally decoupled cis-regulatory signals within one haplotype block, whose decoupling rests on tissue-exclusive expression profiles and opposing disease associations; direct demonstration of two independent cis-regulatory elements awaits functional assays.

Stratified fine-mapping resolves the two signals to non-overlapping credible sets. Stratified SuSiE fine-mapping of the same region across four traits showed that the credible variant sets are phenotype-specific and mutually exclusive. The fibroblast cis-eQTL resolved to a single-variant 95% credible set at rs73234745 (PIP = 0.975; three credible sets, all with purity ≥ 0.98); the plasma pQTL resolved to a single-variant set at rs55940906 (PIP = 1.00) with seven further high-purity sets, none of which contains any fibroblast-eQTL credible

variant; and heel eBMD produced one single-variant set (PIP = 0.9998) together with a 1,112-variant set of purity 0 and two further sets of 23 and 25 variants, the 1,112-variant set being the expected signature of an unresolved D′ ≈ 1 block on the GWAS side. All pairwise credible-set overlaps had Jaccard ≤ 0.011. Because this evidence rests on each trait's own summary statistics rather than on LD estimation, it constitutes an LD-independent confirmation that the two cis signals are carried by different variants within the haplotype block (Supplementary Fig. S9; Supplementary Table S17). The POP signal itself produced no credible set (maximum PIP 0.04), consistent with its sub-genome-wide association (see Limitations).

Conditional analyses confirm one shared causal variant and no hidden second signal. Conditioning each trait's regional model on the other trait's credible-set lead variant preserved the architecture rather than eroding it. The fibroblast eQTL credible set retained its lead variant when conditioned on the pQTL fine-mapped lead (rs73234745 PIP 0.975 → 0.963), and the eQTL × pQTL colocalisation separation was stable under conditioning in both directions (PP3 0.90–0.97, PP4 ≤ 0.003 throughout). Conversely, the POP × eQTL colocalisation collapsed when either side was conditioned on the other's lead variant (PP4 0.962 → 0.006 conditioning the GWAS on rs73234745; PP4 0.962 → 0.024 conditioning the eQTL on the POP regional top variant rs35783198), and the eQTL × eBMD colocalisation collapsed identically (PP4 0.966 → 0.007) — the signature expected when POP and eBMD each share exactly one causal variant with the eQTL and no second independent shared variant is present anywhere in the block (Supplementary Table S28).

The separation is not an artefact of settings, and the target gene is unique. Across ten SuSiE configurations the eQTL and pQTL credible sets never shared a variant (Jaccard = 0 throughout; Supplementary Table S25), and the separation survived recomputation of LD from the independent 1000G v5b call set. Nor is the target-gene assignment ambiguous: within the lead block (rs73234745 ± 150 kb), the only fibroblast cis-eQTL surviving Bonferroni correction across regional genes with fibroblast summary statistics (9,692 gene–variant tests) is PCOLCE2 itself (block minimum p = 1.4×10^{-6}, corrected p = 0.014); the nearest competitor, the lncRNA ATP1B3-AS1 (block lead r^2 = 0.58 with rs73234745), has nominal p = 0.002 (corrected p = 20.1), and no other gene in the block carries a fibroblast eQTL signal at all (Supplementary Fig. S13; Supplementary Table S26). Allele-level summaries of the key variants across the four traits are consolidated in Supplementary Table S18.

Cross-trait colocalisation gives the same answer without LD estimates. Between the fibroblast cis-eQTL and the plasma pQTL, the posterior favoured two distinct causal variants (PP3 = 0.92) over a shared variant (PP4 = 0.002); the slightly larger PP3 than in the protein-layer colocalisation of Methods (0.903) reflects the different harmonised variant set, not a

different conclusion. This was stable across ±100–600 kb windows (PP4 = 0.0016 in all four) and across priors from 1×10^{-6} to 1×10^{-4} (PP4 ≤ 0.016). The plasma pQTL and heel eBMD shared no signal (PP3 = 1.00, PP4 < 0.001), whereas the fibroblast eQTL and eBMD converged on a shared variant when the window matched the single-causal assumption (PP4 = 0.96 within ±200 kb), falling to 0.04 at ±600 kb as the eBMD D′ ≈ 1 block entered the region — an instability that marks the limit of the single-causal approximation rather than contradicting the narrow-window estimate. Because the pQTL (eight credible sets) and eBMD (one unresolved 1,112-variant block) both violate the single-causal assumption of coloc.abf, the SuSiE separation remains the primary evidence and these posteriors are corroborative (Supplementary Table S19).

PCOLCE2 is selectively expressed in pelvic-floor fibroblasts

Mapping PCOLCE2 across lineages in GSE331133 (40,412 cells from the anterior vaginal wall of postmenopausal women with prolapse) showed selective expression in the fibroblast compartment: mean normalised expression was 0.986 in fibroblasts versus 0.02–0.07 in all other lineages, a 14- to 40-fold difference, with 42% of fibroblasts positive (Fig. 4). The same compartment bias was present for LOX and COL1A1, indicating that the assembly programme as a whole is a fibroblast property in this tissue rather than a PCOLCE2-specific anomaly. This dataset is not a case–control contrast: it compares untreated women with prolapse against locally oestrogen-treated women, and it is used solely as expression localisation. This attribution rests on lineage correspondence, not on identity of regulatory architecture between dermal and fascial fibroblasts (see Limitations).

Exploratory pseudo-bulk transcriptomics of patient tissue

All transcriptomic observations from this small cohort are hypothesis-generating correlates and are not treated as confirmatory mechanistic evidence. To ask whether the genetic susceptibility has a transcriptional correlate in patient tissue, we generated pseudo-bulk profiles of the fibroblast compartment from single-cell prolapsed anterior vaginal wall data (GSE151192; 16 POP versus 5 controls; Fig. 5C). With five controls the design can exclude only large effects, so these comparisons are descriptive. The collagen-assembly module showed a positive but non-significant displacement (Hedges' g = +0.45, label-shuffling p = 0.38; 28 genes), and the matrix-degradation module was centred on zero (g = +0.04, p = 0.94; 24 genes). The assembly-minus-degradation differential was positive but did not reach significance under either null (label-shuffling p = 0.16; gene-set permutation p = 0.14). The only module that moved significantly was the inhibitor set (TIMP1–4 and SERPINE1; g = +1.55, p = 0.005, BH-adjusted p = 0.026); the housekeeping module was null (p = 0.29). The direction of the contrast was stable — positive in all 21 leave-one-donor-out fits (+0.42 to

+1.15) and under three alternative scoring definitions (+0.72, +0.80, +0.72) — but directional stability is not significance and is not treated as one: it shows only that the estimate is not driven by a single outlier donor and cannot constitute statistical evidence for a biological effect at 16 versus 5 samples. Donor-level bootstrap intervals are wide (assembly 95% CI −0.67 to +1.78; minimum detectable |g| = 1.38 at 80% power) and age adjustment left the contrast unchanged (Supplementary Table S11). Because GSE151192 carries no individual genotype data, this increment cannot be attributed to the PCOLCE2 risk allele: it is a correlate of disease state, not of genotype.

PCOLCE2 itself was elevated in POP tissue (log2 fold change +2.13 within fibroblasts, p = 0.097; +2.74 across all cells, p = 0.027), concordant in direction with the MR estimate. In paired bulk analysis of prolapsed versus non-prolapsed anterior vaginal wall sites (GSE53868; 12 patients; within-pair sign-flip permutation), the matrix-degradation module was up-regulated (g = +0.86, sign-flip p = 0.0003, BH-adjusted p = 0.0015) and the inhibitor module was also up-regulated (g = +0.87, sign-flip p = 0.0067, BH-adjusted p = 0.017). The collagen-assembly module was not displaced (g = −0.03, sign-flip p = 0.90), and the assembly-minus-degradation differential was negative (g = −0.59, sign-flip p = 0.038 before dye correction; g = −0.54, p = 0.059 after per-feature dye-effect correction, Methods), consistent with substantially altered matrix turnover and cellular composition in overt prolapse. The design is balanced, so dye correction is variance-only and conservative: every module direction is unchanged, and the differential is directional rather than nominally significant. The contrast is also not a mirror of cellular composition (marker z = −0.18, p = 0.19; marker-adjusted 624 of 2,276 features significant; COL1A1 R^2 = 0.02). Neither design can be chained onto the genetic estimate: the paired comparison contrasts two anatomical sites within the same patients and is not a population case–control effect, and the under-powered 16-versus-5 comparison neither establishes nor excludes an assembly-stage perturbation (Supplementary Table S11). Neither design therefore serves as independent validation of the MR-colocalisation associations.

The increment is within-subtype expression, not a composition shift. Clustering resolved four canonical stromal states — ECM-remodelling (65.7% of POP fibroblasts), quiescent/resting, senescent/SASP-like and inflammatory. The disease-related PCOLCE2 increment was almost entirely a within-subtype expression change rather than a composition shift: of the total POP-versus-control difference in mean PCOLCE2 expression (Δ = 0.64 on the log1p-CPM10k scale), 0.62 (96%) was attributable to increased expression within subtypes, −0.005 to shifts in subtype proportions, and 0.035 to their interaction. The increment was largest in the ECM-remodelling subtype (Δ = +0.88, Welch p = 0.010; fraction of PCOLCE2-positive cells +37 percentage points, 0.39 to 0.76, p = 0.034), with a smaller

concordant shift in the quiescent subtype (Δ = +0.28, p = 0.012); the only significant composition change was expansion of the senescent/SASP subtype (+7.4 percentage points, p = 0.040), which itself did not carry the PCOLCE2 increment (Δ = +0.15, p = 0.086). The target cell implied by the genetic signal is therefore best modelled as an ECM-remodelling fibroblast whose per-cell PCOLCE2 output rises, not as an expansion of PCOLCE2-positive cells. Because GSE151192 carries no genotype, this is cell-localisation evidence for the effect compartment and not an sc-eQTL (Supplementary Fig. S10; Supplementary Table S22).

PCOLCE2 elevation is a discrete, gene-specific event rather than programme co-activation. Within the ECM-remodelling subtype, per-cell PCOLCE2 did not co-vary with the collagen-assembly module (donor-level median Spearman ρ = +0.02 in POP versus −0.07 in controls; Mann–Whitney p = 0.84), whereas a moderate correlation with the quiescent-matrix identity module was preserved (ρ = 0.41 versus 0.29). PCOLCE2 elevation is therefore a discrete, gene-specific shift rather than proportional co-activation of the remodelling programme — consistent with a cis-regulatory origin acting on this gene, and against interpreting PCOLCE2 as a passive passenger of a global ECM programme. Cell-level correlations are reported for completeness only; they are dominated by donor batch structure and are not interpretable (Supplementary Table S20).

Circulating PCPE-2 does not mediate the POP association

We then asked whether circulating PCOLCE2 protein, rather than fibroblast-lineage expression, carries the association. The fibroblast-active lead rs73234745 is itself a strong plasma pQTL (β = −0.298 SD per C allele, $p = 2.6\times10^{-31}$ in the deCODE SomaScan study of 5,361 individuals), and a Wald-ratio MR using this variant yielded OR 1.249 (95% CI 1.133–1.377, $p = 8\times10^{-6}$) per SD higher plasma PCOLCE2, closely matching the expression-level estimate. Because the same instrument underlies both, the protein-MR estimate should not be interpreted as independent confirmatory evidence for the disease-relevant mechanism. It shares its genetic instrument with the expression-MR and serves only as an internal consistency check, showing that the variant's effect is reproducible at the circulating-protein level and is therefore not a transcript-only artefact. The dominant plasma protein signal, however, is genetically distinct: stratified fine-mapping had already placed the two layers in non-overlapping credible sets (fibroblast eQTL rs73234745, pQTL rs55940906), and coloc-ABF — corroborative only, because the pQTL violates its single-causal-variant assumption — gave the same answer (PP3 = 0.903, PP4 = 0.003; 2,825 variants; PP3 0.86–0.90 when contracted to ±10 kb around the fibroblast lead). At the pQTL lead rs11716897, which drives plasma PCOLCE2 far more strongly (β = +0.39 versus −0.30 SD), fibroblast expression

shows no effect (z = 0.26), and the blood-haplotype tags (rs4683702 and rs11716897, r^2 = 0.97) showed no association with POP (observed per-allele OR 1.006 and 1.008).

This pattern was quantified with the dose–response concordance test. If circulating dosage mediated the effect, each pQTL-defining variant should reproduce the per-SD plasma-protein effect estimated from the fibroblast lead (log OR 0.222 per SD; estimates from deCODE, SD scaling verified in Methods). Every non-fibroblast pQTL variant fell significantly below that expectation: the blood-haplotype tags were predicted to a per-allele OR of 1.090 but observed OR ≈ 1.01 (z = −3.5, p = 4.0–5.1×10^{-4}), and the independent third pQTL lead rs2707975 was predicted to OR 1.037 but observed OR 0.973 (z = −3.7, p = 2.1×10^{-4}). The discordance was directionally uniform, with all three variants falling on the same side of the expectation, and a parallel contrast extends to two further pQTL cohorts (rs34516933, r^2 = 0.94 with rs4683702, n = 3,301; rs2707975, n = 3,506). Collectively, variants that raise circulating PCOLCE2 do not raise POP risk by the amount the circulating-dosage model requires: the risk signal follows the fibroblast-lineage expression arm, and circulating protein dosage is not the operative mediator — the compartmentalised pattern expected for a locally acting ECM processing factor (Supplementary Figs. S3 and S4; Supplementary Table S13).

A bone-density signal shares the same causal variant as the fibroblast eQTL and POP

Colocalisation between the fibroblast-active PCOLCE2 signal and heel bone mineral density (eBMD) reflects parallel pleiotropy rather than mediation of POP risk through skeletal properties. The locus-level pleiotropy scan had shown that the fibroblast haplotype accounts for the region's reported eBMD associations, and the fine-mapping agrees: eBMD produced one single-variant credible set at rs147652399 (PIP ≈ 1.00; r^2 ≈ 0.98 with the fibroblast tools), alongside an unresolved 1,112-variant set that marks the limits of fine-mapping on the GWAS side. Colocalisation against eBMD (UK Biobank, n = 448,010) is consistent with sharing at the causal-variant level — eBMD and the fibroblast cis-eQTL were best explained by a single shared causal variant (PP4 = 0.961, PP3 = 0.036; 960 variants), as were eBMD and POP (PP4 = 0.963; both stable across ±10–200 kb windows and p_{12} = 10^{-6}–10^{-4} priors) — but because the eBMD block violates the single-causal assumption these posteriors are corroborative only. All three signals therefore trace to the fibroblast-active haplotype, and the partition repeats in eBMD: rs73234745 and rs147652399 lower eBMD per C allele (β = −0.0146 and −0.0149 SD, p = 6.7×10^{-9} and 2.9×10^{-9}), whereas the blood tags rs4683702 and rs11716897 are eBMD-null (p = 0.56 and 0.58) — the same tissue-restricted pattern seen for POP. The direction of the shared effect is informative: the expression-raising, POP-risk allele (T) raises eBMD, opposite to the naive route in which POP risk travels through

skeletal weakness. Colocalisation supports a shared causal variant for both signals, but it does not by itself establish a mediation chain: whether the POP effect transits fibroblast expression, a parallel matrix effect, or both cannot be resolved genetically (Supplementary Fig. S7; Supplementary Table S15).

The cis-regulatory mechanism of the lead variant remains unresolved

Finally, we performed an exploratory annotation of the three lead variants. None overlapped an ENCODE cCRE (nearest distal enhancer-like signatures 450–962 bp away), all fell within quiescent chromatin across five Roadmap epigenomes, and allele-specific TEAD1–TEAD4 motif scanning showed no evidence of motif creation or disruption (largest allele-specific change −3.2 bits at TEAD4, rs4683702 T>G; empirical p = 0.43–0.93 against the regional null of 469 variants; Supplementary Fig. S1, chromatin panel). A direct TEAD/YAP mechanism at the lead variant is therefore not supported, and the cis-regulatory mechanism of the fibroblast-active haplotype remains unresolved.

In the expanded annotation, none of the seven key variants — including the pQTL top variant rs55940906 — overlapped an accessible peak in any of six ENCODE cell types (nearest peak 160 bp to 36 kb away), and all are quiescent in the three fibroblast Roadmap epigenomes (Supplementary Fig. S8; Supplementary Table S21). The flanking ±5 kb windows, however, were peak-dense only in fibroblast-lineage cells — 3–5 peaks per variant in primary dermal fibroblasts and 0–3 in the other three fibroblast preparations — against 0 in HUVEC endothelial cells and 0–1 in GM12878 cells. The lead variants therefore tag rather than coincide with accessible regulatory chromatin, and the restriction of flanking accessibility to fibroblast-lineage cells is concordant with the tissue-restricted expression architecture. Regulatory activity must nonetheless be assessed in pelvic-floor-derived fibroblasts; this annotation motivates, and supplies candidate fragments for, the reporter assays proposed in the Discussion.

Discussion

In this study we resolve two decoupled, cell-type-specific cis-regulatory profiles within a single LD block at the *PCOLCE2* locus, and show that only the fibroblast-lineage signal is associated with pelvic organ prolapse. Stratified SuSiE fine-mapping assigns the fibroblast eQTL (rs73234745) and the plasma pQTL (rs55940906) to non-overlapping credible sets — a separation preserved across ten settings, an independent LD estimate and every window and prior perturbation — although the two variants remain weakly correlated ($r^2 = 0.068$) inside one $D' \approx 1$ block. Fibroblast *PCOLCE2* instruments associate with prolapse but with no systemic connective-tissue endpoint; variants that raise circulating PCPE-2 do not raise POP

risk by the amount a circulating-dosage model requires; single-cell mapping places *PCOLCE2* in fibroblast lineages of patient tissue; and within the ECM-remodelling subtype the increment is subtype-internal rather than compositional. We read this constellation as evidence that the prolapse signal acts through fibroblast-lineage cis-regulation of the collagen-processing machinery, with the circulating-protein arm genetically distinct from it — a statistical inference, not a demonstration of validated regulatory elements.

A compartmentalised regulatory architecture rather than a gene-level pleiotropy

The standard description of a locus such as this one is gene-level pleiotropy: one gene, several traits. What the data show is more specific. The two haplotypes carry different credible-set memberships, different tissue expression profiles and different trait associations, and the partition repeats in an independent molecular phenotype (eBMD). The circulating-protein arm is genetically real — the pQTL lead drives plasma PCPE-2 far more strongly than the fibroblast lead — but it is not the arm that carries POP risk, which is the pattern expected for a locally acting matrix-processing factor whose relevant concentration is at the fascial interface rather than in plasma. The protein-MR estimate should therefore not be read as independent support: it shares its instrument with the expression-MR and shows only that the variant's effect is reproducible at the protein layer. CAUSE likewise separates causal from correlated pleiotropy only under its own assumptions, so the causal language used here is shorthand for a triangulated statistical inference, not for a demonstrated mechanism.

Cross-tissue sensitivity MR for the panel (Supplementary Fig. S18; Supplementary Table S31) showed LOX protective across the three stromal sources with heterogeneous magnitude (Cochran Q = 6.5, p = 0.038, I^2 = 69%). LOX also achieved no supportive colocalisation for POP (PP4 = 0.05). We therefore read that heterogeneity as an example of compartment-restricted regulatory behaviour — the same architecture seen at PCOLCE2 — and not as evidence that LOX is causal for POP. PCOLCE2, although nominally cis-regulated in bulk vagina and uterus (regional minimum p = 3.5×10^{-4} and 1.3×10^{-4}), reached no variant at the 1×10^{-5} instrument tier in any non-fibroblast source — the attenuation expected when a fibroblast-minority organ is assayed in bulk. Applied locus-wide, the same workflow returned no second candidate (seven testable locus–gene pairs; Supplementary Fig. S19; Supplementary Table S32). That bounds the claim — within public fibroblast eQTL resolution the effect at 3q24 is not a generic property of POP matrix loci, and the nominal ALPL estimate without colocalisation illustrates the failure mode the workflow is designed to filter out.

Stoichiometric perturbation rather than more collagen

Up-regulation of an assembly programme does not imply “more collagen”, and certainly not a mechanically stronger fascia. PCOLCE2 encodes PCPE-2, a secreted regulator of procollagen C-proteinase (BMP-1/tolloid) activity; because PCPE-2 can both enhance C-terminal maturation and bind and inhibit BMP-1/tolloid, its net effect is concentration- and stoichiometry-dependent, and up-regulation is best understood as a perturbation of the PCPE-1/PCPE-2/BMP-1 axis rather than a gain of assembly activity. Which direction that perturbation takes in pelvic-floor fibroblasts cannot be resolved from genetic data alone; the working hypothesis draws on prior in-vitro biochemistry rather than establishing the reaction here. Such a perturbation would be expected to unbalance fibril nucleation against lateral growth and enzymatic cross-linking, broadening fibril-diameter distribution at unchanged total collagen content — a fascia that is collagen-rich yet mechanically incompetent, the “assembly-biased matrix-decay” state we propose. The patient-tissue signatures are consistent with that picture but are exploratory, under-powered descriptions; titration of PCPE-2 against BMP-1 activity in primary pelvic-floor fibroblasts, with fibril morphology, cross-link maturity and tensile readouts, would test it directly.

PCOLCE2 is a cis-regulated target, not a passenger of the remodelling programme

The single-cell data also constrain what kind of molecule PCOLCE2 is in this system. Within the ECM-remodelling subtype, where the disease-associated increment is concentrated, per-cell PCOLCE2 did not track the collagen-assembly module (donor-level median $\rho = +0.02$ in POP versus -0.07 in controls, Mann–Whitney $p = 0.84$), although a moderate correlation with the quiescent-matrix identity module persisted ($\rho = 0.41$ versus 0.29). PCOLCE2 up-regulation is therefore discrete and gene-specific rather than a proportional read-out of the programme, which supports a cis-regulatory origin and argues against reading PCOLCE2 as a downstream passenger. The permitted causal statement is correspondingly narrow — *a cis-regulatory variant specifically up-regulates PCOLCE2 within ECM-remodelling fibroblasts, perturbing collagen maturation; this effect is independent of the broader ECM transcriptional programme and decoupled from the haplotype's separate haematopoietic signal controlling circulating PCPE-2; the shared association with eBMD represents parallel pleiotropy rather than mediation* — and we do not claim that PCOLCE2 up-regulation coordinates the programme as a whole.

Where in the tissue would this act? Exploratory spatial transcriptomics of anterior vaginal wall (Visium HD, 16 µm bins; one untreated and one oestrogen-treated section; Supplementary Note S3) localised PCOLCE2-positive bins within the HAS1-positive stromal niche in both sections (enrichment odds ratios 2.18 and 2.24, permutation $p \leq 0.002$), close to but not within ACTA2-positive vascular smooth-muscle bins (odds ratios 0.65 and 0.58), consistent with perivascular stromal rather than vascular-wall action. These observations are

descriptive, rest on single sections with imbalanced depth, and carry no mechanistic weight (Supplementary Fig. S16; Supplementary Table S30). The circulating collagen-metabolite screen likewise gave no robust evidence of a systemic mediator, consistent with a local stromal action model (Methods).

Relationship to previous genetic and proteomic studies

The PCOLCE2 locus was reported as a POP association in the European meta-analysis of 28,086 cases and 546,291 controls [1], which established the locus-level association but did not decompose it molecularly. The present study extends the same locus in three ways: it resolves the two decoupled profiles with tissue-exclusive expression; it separates the local expression pathway from the circulating-protein pathway by dose–response discordance; and it identifies parallel pleiotropy with bone mineral density as a shared-variant collagen effect rather than skeletal mediation. The locus association itself is unchanged; what changes is its interpretation. The East Asian GWAS is consistent at the level of genetic architecture, with 21 of 24 European-derived loci showing concordant directions [2], and we extended that check to this locus using the released Japanese summary statistics (hum0467 / NHA000194; 771 cases, 76,625 controls): the fibroblast lead rs73234745 is directionally concordant (per-T OR 1.263, $p = 1.7\times10^{-3}$) and is the strongest variant in the proxy block, while the blood tag rs4683702 is null (OR 0.92, p = 0.38) — the same partition seen in Europe. At the observed East Asian allele frequencies the study has 4–7% power, so this is a consistency check, not a replication (Supplementary Table S29).

Independent support for collagen-axis involvement comes from tissue proteomics. In prolapsed versus control vaginal wall 1,957 of 5,713 detected proteins differed, with elastin and collagen VI the strongest diagnostic biomarkers [14]; in uterosacral-ligament proteomics of four POP versus four control women, fibromodulin and collagen XIV were down-regulated (log2 fold changes −0.57 and −0.49; Welch p = 0.022 and 0.015) [15]. That 1,789-protein panel detected neither PCOLCE2 nor BMP1, the LOX family or other core processing machinery, so direct protein-level validation of PCOLCE2 is outstanding rather than contradicted. Elevated assembly transcription alongside reduced deposited matrix glycoproteins is consistent with dysregulated turnover rather than greater matrix output (Supplementary Table S23).

A concurrent preprint adds context from the nearest clinical relative of POP on this axis. A genome-wide meta-analysis of urinary incontinence (UI) over 1,045,436 individuals (HUNT, UK Biobank, FinnGen, Michigan Genomics Initiative) identified 54 loci (52 novel) [16]; fibroblast-lineage PCOLCE2 was among the loci with strong colocalisation for female UI (PP4 > 0.75; fibroblast eQTLs from the eQTL Catalogue) in both ICD-coded and self-

reported definitions, POP emerged as a genetic risk factor for UI (12–24% increase), and rg(POP, SUI) = 0.58. This converges with the fibroblast-restricted architecture described here and with the fibroblast haplotype's urinary-system endpoint association. It is not an independent replication: the UI meta-analysis includes FinnGen participants who also contribute to the POP endpoint analysed here, and both analyses draw on the eQTL Catalogue. The agreement is phenotype-level corroboration, not replication; the preprint is non-peer-reviewed and its lead variants were not re-derived here.

A testable generalisation for matrix-gene pleiotropy

The architecture we describe — two decoupled profiles within one gene, each on its own haplotype with its own tissue-restricted expression, disease association and credible-set membership — may not be idiosyncratic. Collagen-maturation and remodelling genes (LOX, LOXL1, ELN, FBN1, ADAMTS16 and related) commonly show multi-tissue disease associations that are read as gene-level pleiotropy. The PCOLCE2 case suggests a sharper, testable alternative: that different regulatory haplotypes within one cis block can be specialised for different collagenous tissues, so that apparent gene-level pleiotropy is an aggregate of tissue-specialised haplotype-level signals. The hypothesis predicts that stratified fine-mapping across the relevant tissue eQTLs and diseases should yield mutually exclusive credible sets, as it does here, and it is testable with summary statistics alone. We therefore propose that gene-level pleiotropy at matrix loci be treated as a hypothesis to be tested rather than as a default reading. The generalisation remains testable rather than established, and evaluating it requires locus-level empirical work in further matrix-gene GWAS loci. Pelvic organ prolapse serves here as the connective-tissue exemplar; the full two-signal architecture of the locus is summarised in Supplementary Fig. S17.

A transferable procedure. The five steps are worth stating once as a recipe, because each answers a question the others cannot; Box 1 sets them out, and the sequence applies to any matrix locus whose GWAS signal has resisted molecular interpretation. Applied elsewhere it should either split a composite signal into tissue-restricted components or show that it does not split, and both outcomes are informative.

Box 1 | A transferable analytical workflow to disentangle matrix-gene pleiotropy within single haplotype blocks

Step 1 — Pre-specify the panel and the endpoints. *Fix the genes and outcomes before estimation and report every panel member, including those for which no valid eQTL instrument exists, so that instrument availability rather than effect size determines what is testable.*

Step 2 — Diagnose estimator identifiability. *Quantify LD among the candidate regional instruments. Where all candidates are near-perfect proxies of the same lead variant, classical multi-instrument MR estimators are not identifiable (that is, they cannot yield a stable, unique point estimate under the local LD structure); reporting them would be uninformative, and designs that remain identified should be substituted.*

Step 3 — Fine-map each molecular phenotype separately. *Run stratified fine-mapping on the eQTL, pQTL and GWAS trait sets and compare credible sets, because colocalisation assumes one causal variant per trait and fails precisely in the dense blocks where matrix genes sit.*

Step 4 — Test the mediator at the variant level. *Use a dose–response discordance test against the mediator-model prediction rather than simple mediator MR, which inherits the same LD constraints.*

Step 5 — Treat the eQTL tissue as part of the hypothesis. *Analyse the tissue of origin as a component of the question rather than as a nuisance variable, and add cross-tissue sensitivity MR to show whether direction, magnitude or both depend on the source.*

Scope. *Every genetic inference generated by this workflow is conditional on the authenticity of the input GWAS index association, which in our study is sub-genome-wide suggestive. This workflow organises and prioritises genetic evidence for regulatory-architecture hypotheses; it cannot establish molecular-level causal mechanisms and is no substitute for targeted functional experimental validation.*

The broader “assembly-biased matrix-decay” framing is likewise a working hypothesis, not a disease classification. LD score regression across fourteen pelvic-floor and connective-tissue endpoints shows positive genetic correlations between POP and hernia (rg = 0.38), incisional hernia (0.42), diverticular disease (0.28), constipation (0.31), lumbar disc disorders (0.23) and varicose veins (0.15) (all $|z| \geq 2$; Supplementary Fig. S15). Correlation alone does not locate a shared mechanism: locus-level evidence at PCOLCE2 exists only for POP and the eBMD arm. Extending the framing requires the same tissue-matched colocalisation, conditional fine-mapping and directionality chain applied trait by trait, and until that is done it should be read as a hypothesis-generating summary of the correlation structure rather than as an established nosology.

Figure 6 presents a working mechanistic hypothesis that combines our genetic inference with public biochemical literature; every dashed-arrow step downstream of PCOLCE2 expression remains experimentally unvalidated here. The TEAD/YAP route that motivated our annotation has substantial precedent in pelvic-floor tissue, which is why it was worth testing even though our data do not support it at this variant: YAP/TAZ are canonical mechanosensitive co-activators of TEAD-dependent transcription [17], and in prolapse,

matrix stiffness drives autophagy-mediated Kindlin-2 translocation, MOB1 degradation and nuclear YAP accumulation in fibroblasts [18]. What our data add is a negative allele-level result: the lead variants do not alter predicted TEAD binding, so the cis-regulatory mechanism of the fibroblast haplotype lies elsewhere — a distant element, variants not tagged by the lead, or a non-transcriptional route, none of which is tested here. Exploratory transcriptomics likewise showed no stable correlation between PCOLCE2 expression and canonical YAP/TAZ or TGF-β signatures, which may simply reflect that YAP/TAZ activation is a dynamic mechanical consequence rather than a steady-state transcriptional correlate in static biopsies. The cis-regulatory effect on PCOLCE2 is therefore the only solid link in Fig. 6, and the downstream cascade is drawn dashed.

Implications

If the fibroblast-lineage signal is the relevant arm, then the circulating protein is a poor proxy for both mechanism and therapeutic target: interventions aimed at plasma PCPE-2 would be predicted to have little effect on the fascial compartment, whereas anything that normalises procollagen processing within ECM-remodelling fibroblasts — including restoration of the PCPE-1/PCPE-2/BMP-1 balance — would engage the proposed mechanism. What this study provides is genetic prioritisation, not a therapeutic programme: it identifies the arm worth engaging, but cannot specify a modality, a target product profile or a dosing strategy, and it provides no evidence that any existing drug acts through this route.

Two hypotheses are testable in cohorts that already exist or could be assembled without new molecular work. **(i) Phenotype-stratified association.** If the risk allele acts through local procollagen processing, its association should be stronger for prolapse dominated by fascial failure than for prolapse driven mainly by levator injury; testing this needs cohorts with imaging or surgical phenotyping, because registry endpoints cannot make the distinction. **(ii) Genotype-stratified recurrence after native-tissue repair.** If the mechanism is local matrix assembly, carriers of the fibroblast-haplotype allele should show higher recurrence after fascial repair but not after mesh augmentation, where the load is carried by the implant. Both are falsifiable with existing surgical series and genotype data.

A structural limit on clinical translation is that no POP GWAS with anatomical or surgical subtype phenotypes exists. Published and registry endpoints pool anterior, posterior and apical prolapse and do not separate fascial from levator injury. The clinical value of this locus is therefore for now confined to genetic prioritisation and to the design of the cohort studies above. The co-localisation of the POP and bone-density signals at a single regulatory variant likewise cautions against reading the epidemiological overlap between osteoporosis and

prolapse as one tissue causing the other; the shared signal is better explained by a common regulatory architecture in collagenous tissues.

Table 1 summarises the inferential status of each claim, so that the strength of the evidence and its limits can be read together.

Table 1. Summary of claims, supporting evidence, inferential strength, key limitations and source locations. Claims marked "Hypothesis-level" are derived from prior literature or exploratory data and are not empirically proven in the present study. **All inferences in this table are conditional on the index signal being a sub-genome-wide-threshold POP association (regional minimum p = 2.5×10^{-7}); independent replication in a larger POP GWAS is mandatory before any claim is treated as established. "Causal-inference" here denotes statistical triangulation from MR, colocalisation and CAUSE, not experimentally proven molecular causation.**

Claim	Evidence	Strength	Main limitation	Source
Two cis-regulatory profiles decouple within one haplotype block	Stratified SuSiE (credible-set Jaccard = 0 across all ten settings) + cross-trait coloc (eQTL × pQTL PP3 = 0.92)	Strong	coloc's single-causal-variant assumption is violated by the pQTL and eBMD traits; SuSiE carries the claim	Supplementary Tables S17, S19, S25
The signal regulates PCOLCE2 (target-gene assignment unique)	Block-wide neighbouring-gene eQTL scan: only PCOLCE2 survives Bonferroni (9,692 tests)	Strong	Fibroblast eQTL n = 966; no pelvic-floor primary cells	Supplementary Table S26
The variant changes transcriptional output	cis-eQTL (z = −4.87), 32/32 proxy concordance, two-tag inverse-variance concordance	Strong	Single independent instrument; no allele-specific expression data exist for this variant	Supplementary Tables S1, S24
The POP effect is consistent with causality rather than confounding	MR OR 1.282 + coloc PP4 0.954 + CAUSE (causal over sharing, ΔELPD z = −4.88)	Intermediate	Wald ratio; no multi-instrument robust estimator is identifiable	Supplementary Tables S1, S16, S17
Cell source: ECM-remodelling fibroblasts, within-subtype expression	Pseudo-bulk decomposition, 96% within-subtype component	Strong (direction) / intermediate (magnitude)	No sc-eQTL; subtype call is unsupervised	Supplementary Table S22
PCOLCE2 is not a passenger of the ECM programme	Subtype-internal donor-level ρ ≈ 0 (PCOLCE2 × assembly module)	Intermediate	Negative result; 21 donors	Supplementary Table S20
Circulating PCPE-2 does not mediate POP	Dosage-discordance z = −3.5 to −3.7, directionally uniform; coloc PP3 = 0.90	Strong	pQTL scaling verified for deCODE only	Supplementary Tables S13, S16
Parallel pleiotropy with eBMD	eQTL × eBMD coloc PP4 = 0.96 (±200 kb); shared lead variant	Intermediate	D′ ≈ 1 block violates the single-causal assumption; window-dependent	Supplementary Tables S15, S19
Two physically distinct cis-regulatory DNA elements underlie the two expression signals	No direct experimental evidence; chromatin annotation is compatible with, but non-informative for, this claim	Hypothesis-level (unproven)	Requires reporter-assay or genome-editing experiments in primary pelvic-floor fibroblasts; surrogate-cell epigenomic data cannot resolve the question	Supplementary Tables S8, S21

Assembly-biased matrix decay contributes to POP pathogenesis	Prior PCPE-2 biochemistry plus exploratory pseudo-bulk transcriptomics (assembly versus degradation, 16 versus 5)	Hypothesis-level	Small donor sample (16 versus 5; minimum detectable effect size 1.38); no functional test of fibril assembly or matrix mechanics	Supplementary Table S11; Fig. 5
In-situ protein, sc-eQTL or allele-specific function	Not available in public data	Absent	Requires primary tissue experiments	—

An experimental roadmap

The inferential chain stops at statistical evidence, and four experiments would test it directly. **(1) Allele-specific reporter assays.** Clone the fibroblast-haplotype and blood-haplotype sequences into reporter constructs and transfect primary pelvic-floor vaginal fibroblasts obtained at surgery, with blood-derived cells as the contrast compartment, testing whether the haplotypes differ in cell-type-specific activity and whether two physically distinct regulatory elements exist at all. **(2) PCPE-2 biochemical and mechanical assays.** Manipulate PCOLCE2 expression in the same primary cells and measure BMP-1/tolloid activity, processing products, fibril diameter distribution and matrix tensile properties, addressing whether PCOLCE2 up-regulation enhances or restrains maturation. **(3) Single-cell eQTL in target tissue.** Test whether rs73234745 regulates PCOLCE2 in the true target cell, removing the surrogate-cell limitation. **(4) Prospective cohort with phenotype stratification.** Genotype POP cases with characterised fascial versus levator injury and test the interaction predicted under (i) above. A fifth, computational layer would extend rather than substitute for these experiments. Cell–cell spatial reconstruction from single-cell transcriptomes (for example CSOmap), in larger cohorts with healthy controls and validated against measured spatial transcriptomics, would allow fibroblast–immune communication in POP to be examined at a resolution the present two sections cannot support. We would not prioritise germline Pcolce2 knockout models: pelvic anatomy and the prolapse phenotype differ substantially between mice and women, and human primary-cell systems test the human-specific regulatory question more directly.

Limitations

The most important limitation is stated first: the POP association at this locus is suggestive rather than genome-wide significant (regional minimum $p = 2.5\times10^{-7}$). Every claim in this paper therefore rests on an eQTL-anchored MR–colocalisation–CAUSE chain built around a sub-threshold index signal, and replication in an independent POP GWAS is required before the locus is treated as established. Because the entry point is a pre-specified panel and region rather than an agnostic scan, the p-values quoted here are not genome-wide corrected: they are error rates for the tests we declared in advance, and the claim they

support is a hypothesis rather than a discovery. Beyond that, our conclusions are bounded in three layers — genetic, cell-model and phenotypic; the full sixteen-item account is in Supplementary Note S1, and Supplementary Table S27 maps each item to its resource.

Genetic-level limitations

The single-cis-instrument design imposes boundaries that no sensitivity analysis can remove: the estimate is one linear combination of linked variants, so it cannot separate the causal transcript from a fully linked co-expressed transcript, and it leaves the classical multi-instrument estimators unidentifiable — a constraint of the LD geometry rather than of the data (Methods). All LD computations used the 1000 Genomes EUR panel; in other ancestries different LD structure could alter both credible-set geometry and the degree of decoupling, so the architecture described here is established for European-ancestry populations and its transferability is a hypothesis. Directionality was addressed within these constraints rather than by reverse-MR, which is not identifiable here: no variant reaches genome-wide significance for POP in FinnGen R13 and could serve as a reverse instrument, and re-using the fibroblast lead in reverse would only invert the forward Wald ratio. Direction was therefore assessed positively — all 559 instrument–outcome pairs passed Steiger filtering, and CAUSE preferred the causal to the sharing model (ΔELPD = −7.99 ± 1.64, z = −4.88, one-sided $p = 5.2\times10^{-7}$), the sharing model requiring that roughly 80% of regional variants act through a shared factor against a posterior estimate of 0.05. Genetic support that PCOLCE2 expression increases POP risk is not evidence that the protein alters collagen assembly in vivo; that link rests on prior biochemistry and on the patient-tissue correlate, and awaits perturbation experiments. All GWAS and eQTL resources are of European ancestry, so the East Asian statement is a directional concordance rather than a replication: the three lead variants are common in Korean and Japanese panels (MAF 0.17–0.21, 0.24–0.30 and 0.06–0.10; Supplementary Table S12), 21 of 24 European-derived loci show concordant directions [2], and the released Japanese summary statistics show the fibroblast lead concordant with no genome-wide-significant signal, accounted for by power (4–7%; Supplementary Table S29). This is a directional consistency check only, not a replication: with 771 Japanese cases the study has at most 7% power, so the conclusion is that no direction discrepancy was detected. Testing conservation of the two-signal decoupling requires much larger East Asian case sets and ancestry-specific fine-mapping. Finally, colocalisation assumes one causal variant per trait, an assumption violated by the pQTL (eight credible sets) and by eBMD (a purity-0 block), which is why the SuSiE separation rather than coloc carries the fine-mapping claim; the conditional analyses (Supplementary Table S28) remove each signal in turn.

Cell-model-level limitations

The eQTL, chromatin-accessibility and chromatin-state annotations rely on cultured dermal fibroblasts and on surrogate dermal, foreskin and lung fibroblast epigenomes. Whole-blood cis-eQTLs were analysed not as a proxy for the target tissue but as a contrast tissue: the dissociation between the blood and fibroblast instrument sets is itself the observation that supports tissue restriction. No chromatin-accessibility or genotype-linked data from primary pelvic-floor fibroblasts are public (Supplementary Table S27), so the functional regulatory element remains unmapped and the two “haplotypes” are described as decoupled cis-regulatory signals rather than demonstrated elements. Public pelvic-floor proteomes do not cover PCOLCE2 or the core processing enzymes, so the transcriptional finding has no protein-level validation in the target tissue. Mechanistically, we did not measure PCPE-1, PCPE-2 or BMP-1 abundance, proteinase activity, fibril morphology or matrix mechanics, so the chain from an expression change to a collagen-rich but mechanically incompetent fascia remains a hypothesis.

Phenotype-level limitations

POP was ascertained from national health registries (FinnGen R13 endpoint N14_FEMGENPROL; ICD-10 N81), that is, clinically documented prolapse rather than self-report: specific but incompletely sensitive, and because such misclassification is non-differential with respect to genotype it biases estimates toward the null rather than inflating them. FinnGen R13 provides neither prolapse subtypes nor an incontinence-specific endpoint, so subtype-specific and incontinence-specific statements cannot be made, and the analysis cannot address severity, anatomical compartment or sub-clinical prolapse. The single-cell data are a cross-sectional snapshot from a small donor cohort (16 versus 5), cannot distinguish a susceptibility-stage state from secondary remodelling, and carry no genotype, so no per-subtype regulatory effect could be estimated. Parity, body-mass index and oestrogen exposure were not modelled: with one cis-instrument per haplotype and no individual-level data, multivariable MR is not identified, and gene–environment interaction cannot be tested in a summary-data design at all.

Scope

Scope. All analyses are a secondary analysis of public summary-level and single-cell data; no primary clinical sample, prospective collection or wet-laboratory experiment was performed. The limitations above are resource boundaries of public in silico research rather than remediable analytical gaps: no pelvic-floor fibroblast eQTL, tissue pQTL or proteomic GWAS exists publicly, the only East Asian POP GWAS offers at most 7% power at this effect size (Supplementary Table S29), and functional validation of the stoichiometric model requires primary tissue experiments that this design cannot substitute for.

Conclusions

Two conclusions stand independently of each other. First, at the level of the disease: independent genetic, colocalisation and regional-causal evidence converges on fibroblast-restricted PCOLCE2 cis-expression as a pelvic-floor-specific susceptibility factor for POP, with the circulating-protein arm genetically distinct from it and not mediating risk. The assembly-biased matrix-decay framing of pathogenesis remains a working mechanistic hypothesis grounded in genetic correlation and in exploratory, under-powered patient-tissue transcriptomics rather than a validated disease classification. Second, at the level of method: within a single haplotype block at a matrix-gene locus, apparent gene-level pleiotropy can resolve into two tissue-restricted cis-regulatory signals, one disease-relevant and one not. Distinguishing them required stratified fine-mapping across molecular phenotypes, explicit treatment of the identification constraints imposed by LD, and a protein-level test designed to fail if circulating dosage were the mediator. Because that sequence is generic rather than locus-specific, it should apply to other matrix-gene loci whose GWAS signals have resisted molecular interpretation, and it nominates fibroblast-lineage cis-regulation of PCOLCE2 as a candidate target in pelvic-floor connective-tissue disorders.

Declarations

Ethics approval and consent to participate. This study is a secondary analysis of publicly available, de-identified summary-level and single-cell data. No new human participants were recruited, no identifiable personal data were accessed, and no ethics approval or informed consent was therefore required. Access to each dataset was in accordance with the terms of its repository (FinnGen, GWAS Catalog, eQTL Catalogue, eQTLGen, GTEx, GEO, ENCODE, Roadmap Epigenomics, JASPAR), and the original studies obtained their own ethical approvals and participant consents.

Consent for publication. Not applicable.

Availability of data and materials. All GWAS summary statistics used in this study were obtained from publicly available resources: FinnGen release R13 (https://r13.finngen.fi), the GWAS Catalog (POP meta-analysis GCST90102470; plasma pQTL GCST90089278, GCST90426574 and GCST90242374; heel eBMD GCST90726625; https://www.ebi.ac.uk/gwas), and the eQTL Catalogue release 8 (fibroblast cis-eQTL QTD000216 plus cross-tissue datasets; https://www.ebi.ac.uk/eqtl), with eQTLGen phase 1 (https://eqtlgen.org) and GTEx v8 (https://gtexportal.org) used for independent rechecks. Genotype reference data were from the 1000 Genomes Project phase 3 (v5b release). Single-cell RNA-seq data were accessed through GEO (GSE331133, GSE151192,

GSE151188); bulk pelvic-floor transcriptomic data through GEO (GSE53868, GSE208261); public proteomic summary tables from the supplementary material of the corresponding publications. Chromatin annotations were from ENCODE (cCREs and ATAC-seq; https://www.encodeproject.org) and Roadmap Epigenomics (E055, E062, E117, E126, E128); transcription-factor motifs from JASPAR CORE (https://jaspar.elixir.no); coordinate liftover used the UCSC hg38ToHg19 over-chain file. All primary analysis, plotting and sensitivity scripts are deposited on Zenodo (https://doi.org/10.5281/zenodo.24330648). The archive additionally contains every intermediate summary-statistic table underlying the supplementary tables and figures, the configuration files that record instrument thresholds ($p < 5\times10^{-8}$, otherwise $p < 1\times10^{-5}$), LD clumping parameters, fine-mapping priors and window definitions, the 522-gene core-matrisome list used for the locus-wide scan together with the MD5 checksums of the Gene Ontology and HGNC releases from which it was built, and a README mapping each script to the manuscript item it produces, so that the analysis can be re-run end to end from public inputs. Additional exploratory analyses that are not used for any claim in this manuscript are archived in the same record. No newly generated sequencing data were produced in this study; all results derive from secondary analysis of the public resources listed above.

Competing interests. The authors declare no competing interests.

Funding. This work was supported by the National Natural Science Foundation of China (grant numbers 82401896 and 82501958). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions. Suting Li and Hongming Li contributed equally to this work. Conceptualization: Suting Li and Hongming Li; methodology: Zhi Wang and Songming He; software: Jinhua Tian and Jie Min; validation: Hongming Li and Suting Li; formal analysis: Hongming Li, Suting Li and Jinhua Tian; investigation: Suting Li and Hongming Li; data curation: Zhi Wang and Suting Li; writing—original draft preparation: Bingshu Li and Suting Li; writing—review and editing: Suting Li and Hongming Li; visualization: Zhi Wang and Li Hong; supervision: Songming He and Li Hong; funding acquisition: Suting Li and Jie Min; project administration: Hongming Li and Bingshu Li. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgements. We thank the participants and investigators of FinnGen, the eQTL Catalogue, eQTLGen, GTEx, deCODE genetics, the UK Biobank, the GWAS Catalog, ENCODE, Roadmap Epigenomics and the GEO submitters whose data made this analysis possible.

References

1. Pujol-Gualdo, N., Läll, K., Lepamets, M., Rossi, H.-R., Arffman, R. K., Piltonen, T. T., Mägi, R., Laisk, T. & Estonian Biobank Research Team. Advancing our understanding of genetic risk factors and potential personalized strategies for pelvic organ prolapse. **Nat. Commun.** 13, 3584 (2022). doi:10.1038/s41467-022-31188-5

2. Matsunami, M., Imamura, M., Ashikari, A., Liu, X., Tomizuka, K., Hikino, K., Miwa, K., Kadekawa, K., Suda, T., Matsuda, K., Miyazato, M., Terao, C. & Maeda, S. Genome-wide association studies for pelvic organ prolapse in the Japanese population. **Commun. Biol.** 7, 1188 (2024). doi:10.1038/s42003-024-06875-2

3. Kerimov, N. *et al.* A compendium of uniformly processed human gene expression and splicing quantitative trait loci. **Nat. Genet.** 53, 1290–1299 (2021). doi:10.1038/s41588-021-00924-w

4. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. **Science** 369, 1318–1330 (2020). doi:10.1126/science.aaz1776

5. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. **Genet. Epidemiol.** 37, 658–665 (2013). doi:10.1002/gepi.21758

6. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. **PLoS Genet.** 10, e1004383 (2014). doi:10.1371/journal.pgen.1004383

7. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. **J. R. Stat. Soc. B** 82, 1273–1300 (2020). doi:10.1111/rssb.12388

8. Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M. & He, X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. **Nat. Genet.** 52, 740–747 (2020). doi:10.1038/s41588-020-0631-4

9. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. **Nature** 613, 508–518 (2023). doi:10.1038/s41586-022-05473-8

10. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. **Nat. Genet.** 53, 1300–1310 (2021). doi:10.1038/s41588-021-00913-z

11. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. **Nucleic Acids Res.** 51, D977–D985 (2023). doi:10.1093/nar/gkac1010

12. Gudjonsson, A. *et al.* A genome-wide association study of serum proteins reveals shared loci with common diseases. **Nat. Commun.** 13, 480 (2022). doi:10.1038/s41467-021-27850-z

13. 1000 Genomes Project Consortium, Auton, A. *et al.* A global reference for human genetic variation. **Nature** 526, 68–74 (2015). doi:10.1038/nature15393

14. Deng, W., Zhong, Z., Tong, Y., Liu, J., Wang, X., Xu, L., Li, Y., Chen, X., Wei, Q. & Rao, J. 4D DIA-PRM proteomic study identifying modulated pathways and biomarkers associated with pelvic organ prolapse. **J. Chromatogr. B** 1230, 123916 (2023). doi:10.1016/j.jchromb.2023.123916

15. Li, X.-J., Pan, H.-T., Chen, J.-J., Fu, Y.-B., Fang, M., He, G.-H., Zhang, T., Ding, H.-G., Yu, B., Cheng, Y., Tan, Y.-J., Zhao, F.-L., Morse, A. N. & Huang, H.-F. Proteomics of uterosacral ligament connective tissue from women with and without pelvic organ prolapse. **Proteomics Clin. Appl.** 13, e1800086 (2019). doi:10.1002/prca.201800086

16. Moreno, E., Arora, N., Bertucci-Richter, E., Eghtedarian, R., Andries, A., Thomas, L. F., Horn, J., Stafne, S. N., Pathak, S., Patil, S., Gulla, D., Ollila, H. M., Evans, D. M., Ippolito, G. M., Zhou, W., Coward, E., Surakka, I., Moksnes, M. R., Wolford, B. N. & Brumpton, B. M. Genome-wide association meta-analyses over one million individuals identify 54 loci associated with urinary incontinence and its subtypes. **medRxiv** 2026.03.12.26346952 (2026). doi:10.64898/2026.03.12.26346952

17. Dupont, S., Morsut, L., Aragona, M., Enzo, E., Giulitti, S., Cordenonsi, M., Zanconato, F., Le Digabel, J., Forcato, M., Bicciato, S., Elvassore, N. & Piccolo, S. Role of YAP/TAZ in mechanotransduction. **Nature** 474, 179–183 (2011). doi:10.1038/nature10137

18. Zhao, Z., Han, W., Huang, G., He, Y., Zuo, X. & Hong, L. Increased extracellular matrix stiffness regulates myofibroblast transformation through induction of autophagy-mediated Kindlin-2 cytoplasmic translocation. **Exp. Cell Res.** 436, 113974 (2024). doi:10.1016/j.yexcr.2024.113974

Figure legends

Fig. 1 | Study design and analytical flow. Four modules. **(1) Discovery.** FinnGen R13 summary statistics for 15 endpoints (POP; the broad urinary-system endpoint N14_URINOTH, which is not incontinence-specific; connective-tissue traits; five negative controls) were combined with fibroblast cis-eQTLs from the eQTL Catalogue (GTEx v10 reprocessing, n = 649 donors), rechecked in GTEx v8, to identify two functionally decoupled cis-regulatory signals of PCOLCE2 (fibroblast-active rs73234745/rs55959862 and blood-active rs4683702, $r^2 = 0.066$), with colocalisation (PP4 = 0.954), CAUSE (ΔELPD = −7.99, one-sided $p = 5.2\times10^{-7}$) and cross-tissue, cross-phenotype MR. **(2) Single-cell transcriptomics.** Fibroblast pseudo-bulk module analysis of GSE151192 (a subcompartment-level permutation test, not an independent-sample differential expression analysis) and lineage mapping of GSE331133 (expression localisation only). **(3) Bulk transcriptomics.** Paired-site analysis of GSE53868 and gene-list checks in GSE208261; observational comparisons, not replication. **(4) Exploratory regulatory annotation** (supplementary only): ENCODE cCRE, Roadmap ChromHMM and JASPAR TEAD1–TEAD4 scanning, returning a consistently negative result. Solid arrows denote the main analytical chain; the dashed grey arrow the exploratory branch.

Fig. 2 | Colocalisation and regional causal inference at the PCOLCE2 locus. A, CAUSE model comparison: the distribution of ΔELPD (sharing minus causal) favours the causal model (observed ΔELPD = −7.99 ± 1.64; z = −4.88, one-sided p = 5.2×10^{-7}). **B**, Regional z-scores for the fibroblast cis-eQTL against POP, with the lead variant annotated; the two traits were best explained by a single shared causal variant (PP4 = 0.954). **C**, Posterior estimate of the shared-factor proportion q of the sharing model (q = 0.05, 95% CI shown), with the hypothetical value of q that the sharing model would require for an equivalent fit to these data (≈ 0.80) marked for comparison. The two quantities denoted q are not the same thing: 0.05 is a fitted model parameter, 0.80 is the value that parameter would have to take for the sharing model to explain the data, and neither is an estimate of a causal effect. **CAUSE model comparison relies on Gaussian-liability, single-latent-factor assumptions; interpretation requires triangulation with the MR and colocalisation results. Every estimate shown is conditional on a sub-genome-wide-suggestive index POP association (regional minimum p = 2.5×10^{-7}).**

Fig. 3 | Tissue-restricted cis-eQTL architecture of the two PCOLCE2 haplotypes. cis-eQTL z-scores for PCOLCE2 in cultured fibroblasts, skin, skeletal muscle, uterus, vagina (eQTL Catalogue, GTEx v10 reprocessing) and whole blood (eQTLGen phase 1), shown for the fibroblast haplotype variants rs73234745 and rs55959862 (warm colours) and the blood haplotype variant rs4683702 (blue). Dotted vertical lines mark |z| = 2; crosses denote variants outside the significant set (FDR > 0.05). The two haplotypes are active in mutually exclusive tissues. Cross-phenotype MR results for these instruments are shown in Supplementary Fig. S2, and the independent GTEx v8 replication in Supplementary Fig. S6.

Fig. 4 | PCOLCE2 is selectively expressed in pelvic-floor fibroblasts (GSE331133). Top row, mean normalised expression (log1p CP10k) of PCOLCE2, LOX and COL1A1 across annotated cell types of the anterior vaginal wall, with the percentage of expressing cells above each bar; fibroblasts are highlighted. Bottom row, cell-type composition of the dataset; the difference in mean expression between fibroblasts and all other lineages for the assembly-pathway genes (positive values indicate fibroblast enrichment); and single-cell co-expression of PCOLCE2 and LOX, with fibroblasts in red and other lineages in grey. This dataset compares untreated with locally oestrogen-treated women with prolapse and is used for expression localisation only, not as a case–control contrast. Annotation is marker-based and fibroblasts comprise 4.3% of cells, so compositional statements are not inferred here.

Fig. 5 | Module-level transcriptomic profile of POP tissue across resolutions. A, Effect-size matrix (Hedges' g, POP minus control) for the collagen-assembly module, the matrix-degradation module and their difference, across four tissue resolutions: single-cell vaginal wall (GSE151188, 13 versus 2), fibroblast pseudo-bulk (GSE151192, 16 versus 5), all-cell

pseudo-bulk (GSE151192, 16 versus 5) and paired bulk sites (GSE53868, 12 paired). Asterisks: permutation p < 0.05 (*) and < 0.01 (), from label shuffling (case–control) or within-pair sign flips (paired); for GSE53868 the plotted values are uncorrected for the balanced dye-swap design, and the assembly-minus-degradation cell carries no star because its uncorrected nominal p = 0.038 falls to p = 0.059 after dye correction; the dye-corrected counterpart is given in the text (g = −0.54, p = 0.059). B, PCOLCE2 expression in POP versus control tissue across the same datasets (log2 fold change annotated). C, Fibroblast-compartment module scores in the pseudo-bulk profiles (n = 16 POP versus 5 controls); the assembly module is positively displaced but not significant (p = 0.38), the degradation module is centred on zero, and the only significant shift is the inhibitor module (p = 0.005, BH-adjusted p = 0.026). The pseudo-bulk comparison is a pathway-score permutation test on aggregated samples, not an independent-sample differential-expression analysis: it is exploratory and under-powered, can exclude only large effects, and generates hypotheses only — it is not mechanistic evidence and does not constitute formal replication.* These transcriptomic observations are correlative and cannot independently validate genetic-inference results from MR-colocalisation analyses.

Fig. 6 | Proposed feedforward mechanical feedback model linking PCOLCE2 to pelvic organ prolapse. The risk variant rs73234745 upregulates PCOLCE2 in perivascular HAS1+ vaginal fibroblasts; this relationship (solid arrow) is the inference supported by this study's genetic evidence (expression MR, colocalisation, CAUSE and conditional analyses), which is conditional on a sub-genome-wide-suggestive index association and is consistent with a statistically supported, fibroblast-restricted cis-regulatory effect. **Solid arrow: a link supported by this study's genetic inference. All dashed arrows: unvalidated mechanistic hypotheses derived from external literature and not demonstrated in-house.** The arrow key, together with an explicit statement that dashed arrows represent mechanistic hypotheses derived from external literature and are not experimentally validated in this work, is printed inside the panel. The downstream links — altered procollagen C-endopeptidase processing, abnormal collagen fibril assembly and aberrant ECM biomechanical properties (dashed arrows) — are inferred from published biochemical evidence rather than demonstrated here. Under recurrent pelvic mechanical loading (elevated intra-abdominal pressure), the altered matrix is sensed via integrin–cytoskeletal signalling, suppressing Hippo signalling and promoting YAP/TAZ nuclear translocation and TEAD-dependent transcription (CTGF, CYR61, COL1A1, LOX); this transcriptional programme drives further ECM remodelling, forming a self-amplifying feedforward loop that increases POP risk. Colour bars denote node groups: genetic predisposition (blue), ECM assembly and biomechanics (green), environmental trigger (yellow), mechanotransduction

(red), disease outcome (purple). **This is a working mechanistic hypothesis, not an established pathway.** The YAP/TAZ mechanotransduction cascade is inferred from published literature; the motif analyses in this study do not support direct disruption of TEAD binding by the risk variant, and no functional assay has tested any dashed link. The integrated two-signal genetic model on which this hypothesis rests is provided as Supplementary Fig. S17. No downstream mechanistic step represented by a dashed arrow is empirically validated in the present work.

Supplementary information

Supplementary Tables S1–S32 are provided in a single workbook (Supplementary_Tables_S1-S32.xlsx; S1 contains Group A, the pre-specified five-gene panel with instrument availability, cis-MR, colocalisation and CAUSE, and Group B, the PCOLCE2 locus-specific sensitivity analyses; S28 conditional fine-mapping and colocalisation, S29 East-Asian POP GWAS locus check and power, S30 exploratory spatial transcriptomic statistics, S31 cross-tissue instrument availability and per-tissue-source MR for the five-gene panel, S32 locus-wide matrisome scan at reported POP loci); Supplementary Figs. S1–S19 and Supplementary Notes S1–S5 are provided as separate files (S15: cross-endpoint LD score regression genetic-correlation matrix; S16: exploratory spatial maps; S17: integrated two-signal genetic model of the PCOLCE2 locus; S18: cross-tissue sensitivity of the panel-gene MR estimates; S19: locus-wide matrisome scan (A, MR estimates with colocalisation PP4; B, disposition of all 24 locus–gene pairs); Note S3: exploratory spatial transcriptomic analysis; Note S4: cross-tissue sensitivity protocol, full statistics and caveats; Note S5: locus-wide matrisome scan protocol, full statistics and caveats). Supplementary Table S27 maps each dataset to its tissue or cell model, its use in this study and its principal limitation, and corresponds item by item to the limitations above.

1 · DISCOVERY — public GWAS and eQTL resources

GWAS outcomes · FinnGen R13

15 endpoints:
- POP (N14_FEMGENPROL)
- Other urinary-system diseases (N14_URINOTH; broad proxy, not SUI-specific)
- Hernia · AAA · diverticular · lumbar disc prolapse · etc.
- 5 negative-control traits

eQTL resources

- eQTL Catalogue rel. 8 — fibroblast cis-eQTL (GTEx v10, n = 649)
- Independent recheck — GTEx v8 (43 associations, 2 tissues)
- Cross-tissue: uterus, vagina, skin, skeletal muscle; eQTLGen blood

Lead cis-variants

- Fibroblast-active haplotype rs73234745 / rs55959862
- Blood-active haplotype rs4683702 (2.36 kb apart)
- Weak LD: $r^2 = 0.066$ (0.076–0.169 in phase-3 recheck)

→ functionally decoupled

Instrument QC and causal inference

F = 23.8 (fibroblast) / 63.3 (blood) · Steiger 559/559 passed · coloc PP4 = 0.954
CAUSE ΔELPD = −7.99 ± 1.64 ($p = 5.2\times10^{-7}$) · cross-tissue and cross-phenotype MR (15 endpoints)

2 · SINGLE-CELL TRANSCRIPTOMICS — cell-type resolution

GSE151192 · uterosacral ligament (16 POP vs 5 controls)

Fibroblast-restricted pseudo-bulk (≥500 fibroblasts per sample)
Subcompartment-level pathway-score permutation test (10,000 label shuffles) — not an independent-sample bulk DE analysis

Assembly g = +0.45 (p = 0.001, 28 detected genes)
Degradation g = +0.04 (p = 0.87, 24 detected genes)
Assembly − Degradation g = +0.72 (p = 0.029)
PCOLCE2 log2FC = +2.13 (fibroblast compartment)

GSE331133 · anterior vaginal wall

Postmenopausal POP patients: untreated (n = 5) vs local oestrogen (n = 6)

Expression localisation only — NOT a case–control contrast

PCOLCE2 = 0.986 in fibroblasts vs 0.02–0.07 in other lineages

→ transcriptomic correlate of susceptibility in the fibroblast compartment (supporting tissue-level evidence, not functional validation)

3 · BULK TRANSCRIPTOMICS — whole-tissue context (observational comparison, not replication)

GSE53868 · paired sites (n = 12)

Prolapsed vs non-prolapsed site, same patient
Degradation g = +0.86 (p = 0.002)
Assembly − Degradation g = −0.62 (p = 0.002)
→ interpreted as end-stage remodelling

GSE208261 · RNA-seq (6 vs 6)

Published DE lists only (no full matrix) → partial use
Direction check: PCOLCE2/DCN/CCN2 ↓; MMP9/MMP19/ADAMTS4 ↑
→ qualitative gene-level consistency only

exploratory branch (supplementary only)

4 · EXPLORATORY REGULATORY ANNOTATION — supplementary only, not on the main chain

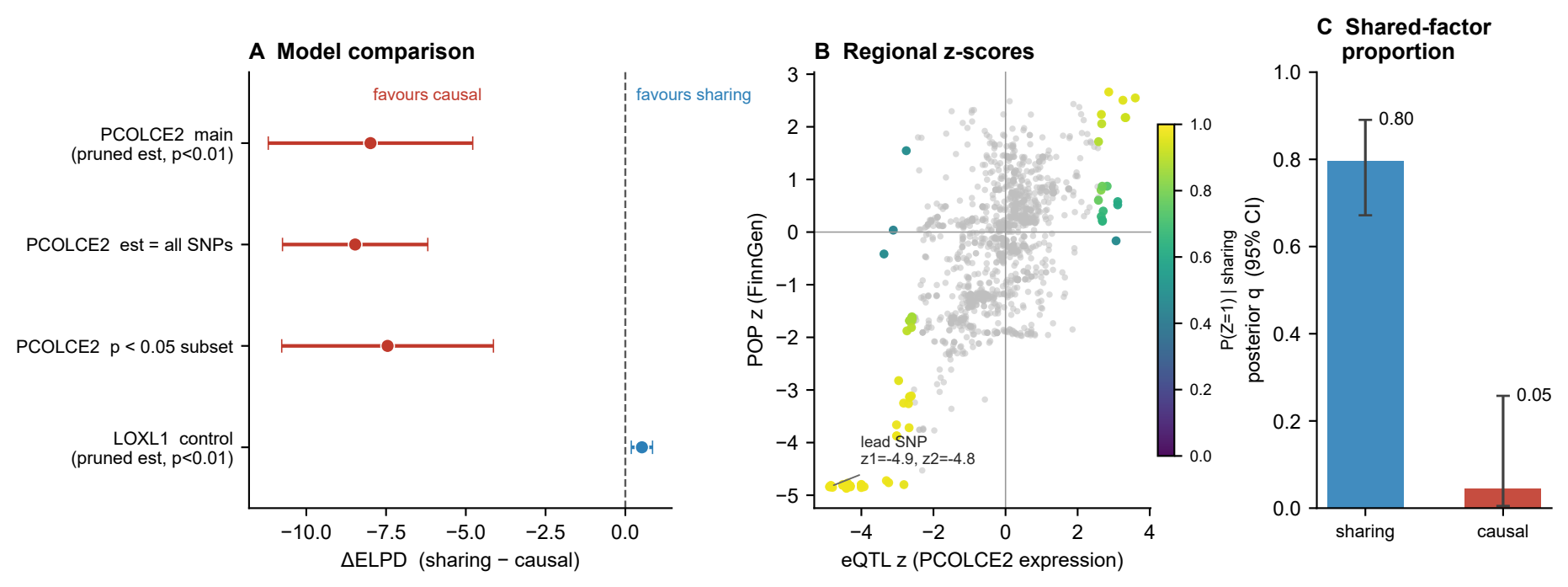
ENCODE cCRE · Roadmap ChromHMM (E055/E062/E117/E126/E128) · JASPAR TEAD1–TEAD4 motif scanning (empirical null, 469 variants)

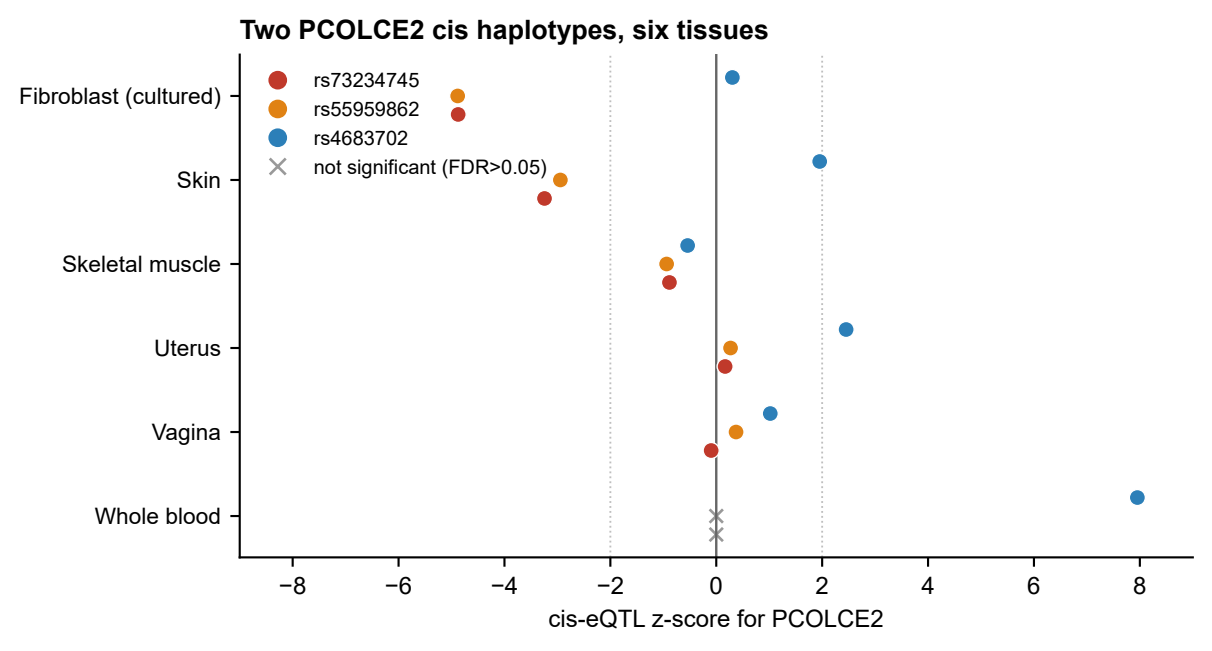
Result: no cCRE overlap · quiescent chromatin · no TEAD motif creation or disruption (p = 0.48–0.99)

→ direct TEAD/YAP mechanism not supported; cis-regulation unresolved (Discussion; Limitation 9)

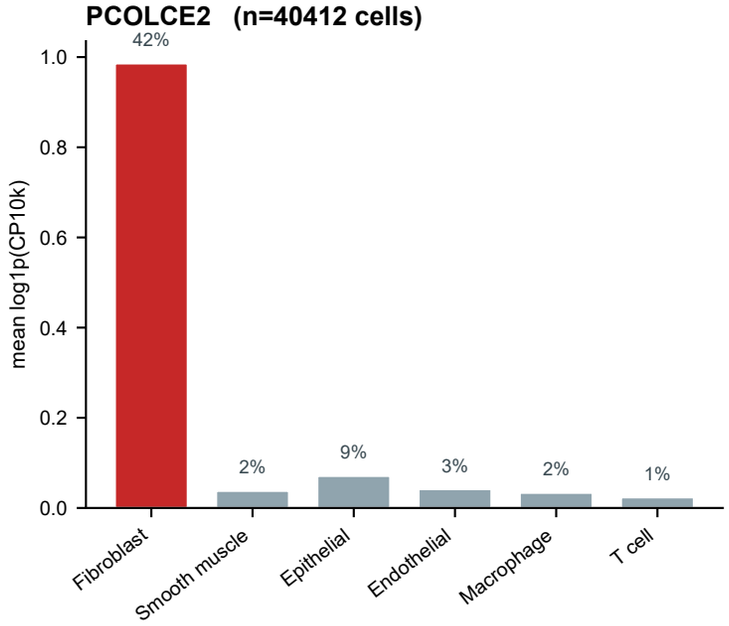
main analytical chain · exploratory branch · supplementary-only module

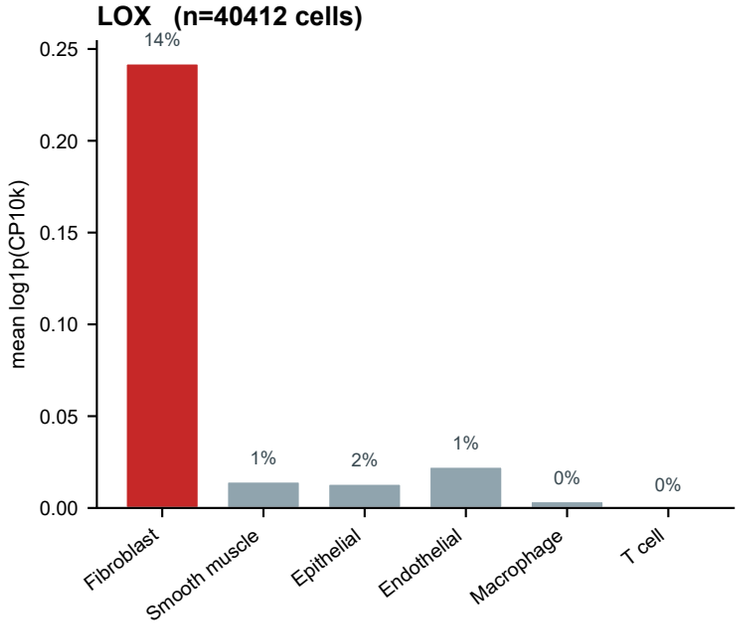
Genetic causal evidence (MR · coloc · CAUSE) is distinct from mechanistic proof. Pseudo-bulk module results are transcriptomic correlates, not functional validation; whole-tissue data are observational context. N14_URINOTH is a broad urinary-system endpoint, not an SUI-specific phenotype.

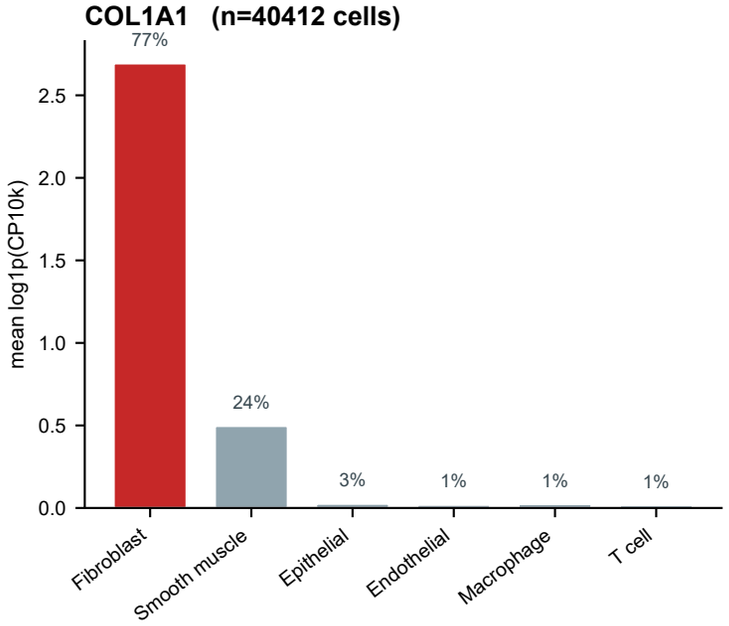

A Model comparison
favours causal
favours sharing
PCOLCE2 main
(pruned est, p<0.01)
PCOLCE2 est = all SNPs
PCOLCE2 p < 0.05 subset
LOXL1 control
(pruned est, p<0.01)
−10.0
−7.5
−5.0
−2.5
0.0
ΔELPD (sharing − causal)
B Regional z-scores
POP z (FinnGen)
eQTL z (PCOLCE2 expression)
lead SNP
z1=-4.9, z2=-4.8
P(Z=1) | sharing
posterior q (95% CI)
C Shared-factor
proportion
0.80
0.05
sharing
causal

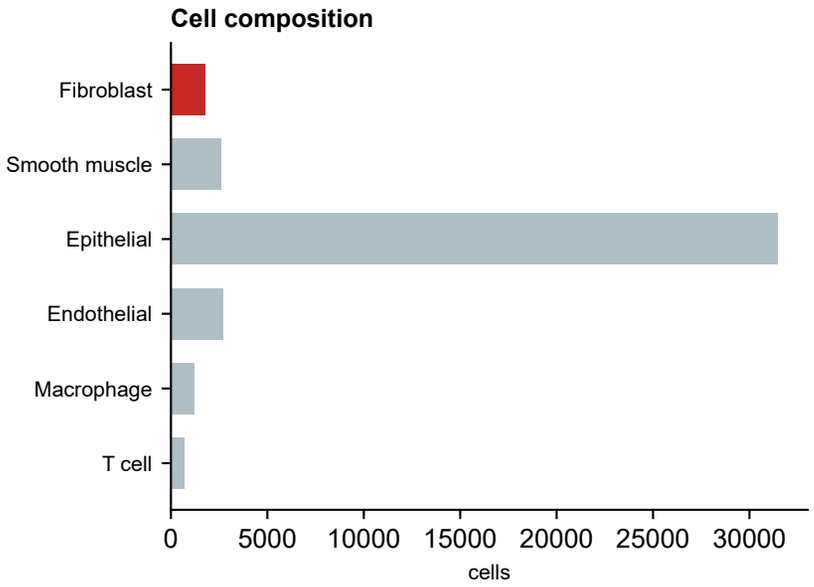

Two PCOLCE2 cis haplotypes, six tissues
rs73234745
rs55959862
rs4683702
not significant (FDR>0.05)
Fibroblast (cultured)
Skin
Skeletal muscle
Uterus
Vagina
Whole blood
−8
−6
−4
−2
0
2
4
6
8
cis-eQTL z-score for PCOLCE2

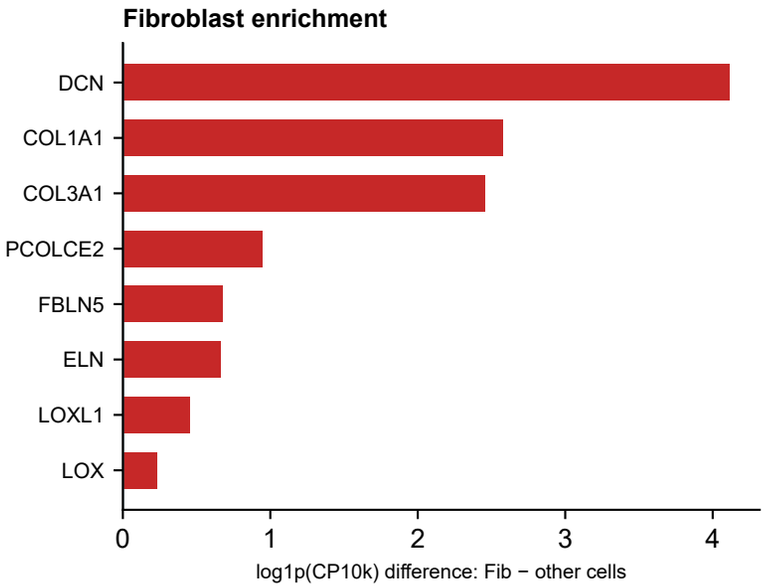
GSE331133 pelvic floor tissue: assembly genes are fibroblast-restricted

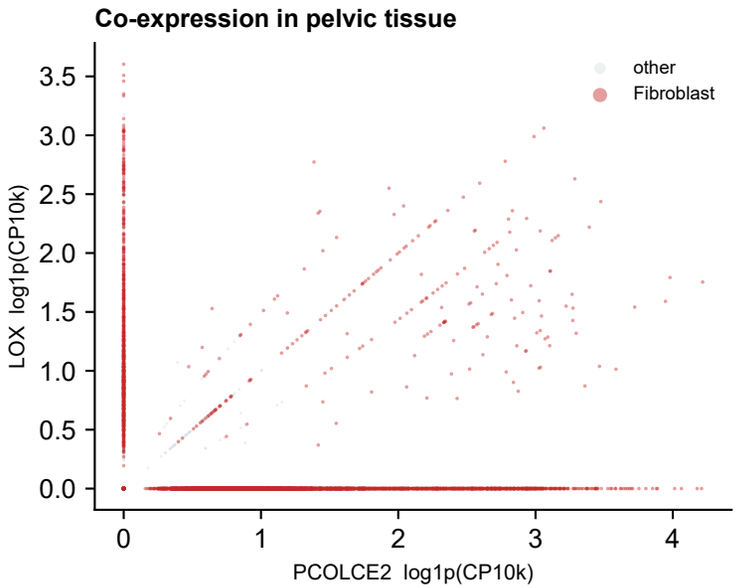


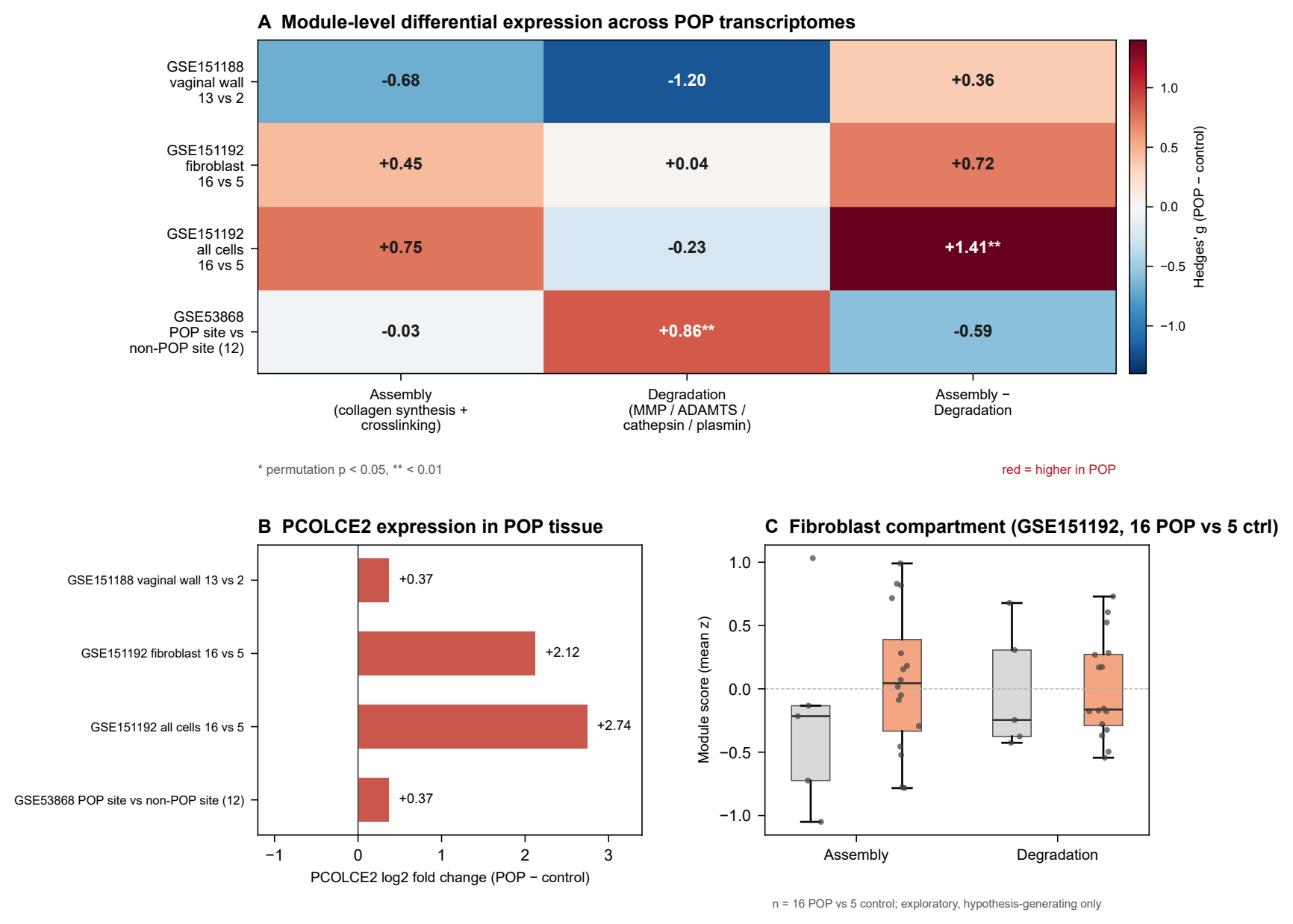




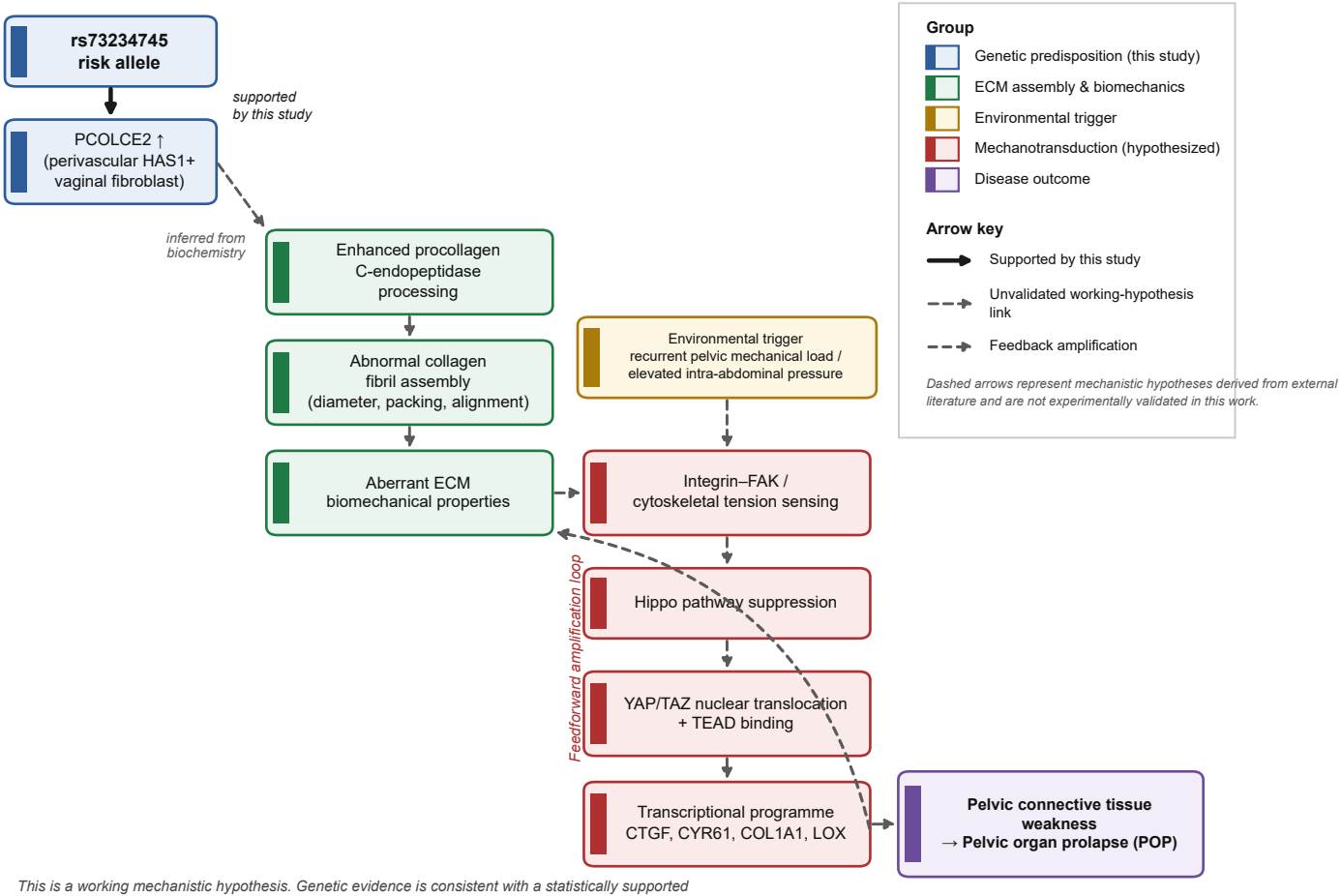







A Module-level differential expression across POP transcriptomes
GSE151188 vaginal wall 13 vs 2
GSE151192 fibroblast 16 vs 5
GSE151192 all cells 16 vs 5
GSE53868 POP site vs non-POP site (12)
-0.68
-1.20
+0.36
+0.45
+0.04
+0.72
+0.75
-0.23
+1.41**
-0.03
+0.86**
-0.59
Assembly (collagen synthesis + crosslinking)
Degradation (MMP / ADAMTS / cathepsin / plasmin)
Assembly − Degradation
Hedges' g (POP − control)
1.0
0.5
0.0
−0.5
−1.0
* permutation p < 0.05, ** < 0.01
red = higher in POP
B PCOLCE2 expression in POP tissue
GSE151188 vaginal wall 13 vs 2
GSE151192 fibroblast 16 vs 5
GSE151192 all cells 16 vs 5
GSE53868 POP site vs non-POP site (12)
+0.37
+2.12
+2.74
+0.37
−1
0
1
2
3
PCOLCE2 log2 fold change (POP − control)
C Fibroblast compartment (GSE151192, 16 POP vs 5 ctrl)
Module score (mean z)
1.0
0.5
0.0
−0.5
−1.0
Assembly
Degradation
n = 16 POP vs 5 control; exploratory, hypothesis-generating only

Proposed feedforward mechanical feedback model linking PCOLCE2 to pelvic organ prolapse



This is a working mechanistic hypothesis. Genetic evidence is consistent with a statistically supported fibroblast cis-regulatory effect of PCOLCE2 on POP; the YAP/TAZ mechanotransduction cascade is inferred from published literature. Motif analyses do not support direct disruption of TEAD binding by the risk variant.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryNoteS4CrossTissueSensitivity.docx
- SupplementaryNoteS5LocusWideMatrisomeScan.docx
- SupplementaryNoteS3SpatialExploratory.docx
- SupplementaryNoteS2ExploratoryRegulatoryAnnotation.docx
- SupplementaryNoteS1ExpandedLimitations.docx
- SupplementaryTablesS1S32.xlsx
- SupplementaryFigureS4doseresponse.pdf
- SupplementaryFigureS5locusPheWAS.pdf
- SupplementaryFigureS8epigenomicannotation.pdf
- SupplementaryFigureS6GTeXv8eQTL.pdf
- SupplementaryFigureS15rgheatmap.pdf
- SupplementaryFigureS13neighbourgenemanhattan.pdf
- SupplementaryFigureS18crosstissuesensitivity.pdf
- SupplementaryFigureS10fibroblastsubtypes.pdf
- SupplementaryFigureS7eBMDcolocalisation.pdf
- SupplementaryFigureS11MRsensitivity.pdf
- SupplementaryFigureS14analysispipeline.pdf
- SupplementaryFigureS1negativecontrolsandchromatin.pdf
- SupplementaryFigureS9stratifiedSuSiE.pdf
- SupplementaryFigureS19locuswidematrisome.pdf
- SupplementaryFigureS17decoupledcissignals.pdf
- SupplementaryFigureS2crossphenotypeMR.pdf
- SupplementaryFigureS3proteintriangulation.pdf
- SupplementaryFigureS12LDheatmap.pdf
- SupplementaryFigureS16spatialmaps.pdf