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.

Phenotype-resolved cis-eQTL Mendelian Randomization and Colocalization Prioritize MAPK3 and UQCC1 in Two Selected Osteoarthritis Candidate Regions

Jiahao Ma
Hefei First People's Hospital
Shiyu Yang
Anhui Medical University
Yibin Du
dybtxyx@163.com

Hefei First People's Hospital

Research Article

Keywords: Osteoarthritis, Mendelian randomization, Colocalization, Cis-eQTL, MAPK3, UQCC1

Posted Date: August 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10422413/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Osteoarthritis (OA) is a major cause of chronic pain and disability, yet no approved therapy modifies disease progression. Human genetic evidence can improve therapeutic target selection, but OA genome-wide association studies often identify loci without resolving the causal gene or direction of effect. We therefore evaluated genes within two OA-associated regions on chromosomes 16 and 20, selected from an earlier exploratory screen, to identify phenotype-specific candidate effector genes.

Methods

We analyzed 14 candidate genes using whole-blood cis-eQTL data from eQTLGen (N = 31,684) and four Genetics of Osteoarthritis consortium phenotypes: AllOA (N = 826,690), KneeHipOA (N = 490,345), KneeOA (N = 396,054), and HipOA (N = 353,388). Evidence was integrated using cis-eQTL Mendelian randomization (MR), Bayesian colocalization (coloc.abf and coloc.susie), and summary-data-based MR with the HEIDI heterogeneity test. Each phenotype was analyzed separately, and gene–phenotype associations were classified using a predefined evidence framework.

Results

MAPK3 was the leading candidate. Higher genetically predicted blood expression was associated with lower OA risk, with negative MR and SMR estimates across all four phenotypes. Concordant evidence was observed for KneeHipOA, KneeOA, and AllOA, with PP.H4 values of 0.991, 0.979, and 0.945, respectively; MR-IVW p values were 7.1×10^{-4}, 1.0×10^{-4}, and 7.5×10^{-4}, and HEIDI tests were non-significant. The HipOA MR-IVW association was not significant (p = 0.091). UQCC1 showed a risk-increasing direction and phenotype-dependent support, strongest for AllOA (PP.H4 = 0.919; SMR p = 7.6×10^{-12}). Evidence was weaker for KneeHipOA and HipOA, while KneeOA findings suggested linkage rather than a shared causal variant.

Conclusions

Integrated genetic evidence prioritizes MAPK3, and secondarily UQCC1, as phenotype-dependent candidate OA targets. Because the exposure was whole-blood eQTL and the UQCC1 signal could not be fully distinguished from the adjacent GDF5 locus, these findings support further experimental validation rather than definitive causal-gene assignment.

Introduction

Osteoarthritis (OA) is the most common joint disorder and a leading cause of chronic pain and disability, with rising prevalence as populations age and obesity become more widespread [1]. Current treatment remains limited to symptom control analgesics, anti-inflammatory agents, intra-articular injections, and arthroplasty for end-stage disease and no disease-modifying OA drug (DMOAD) has been approved [2].

A major barrier to DMOAD development is target selection: drug programs supported by human genetic evidence are substantially more likely to reach approval than those without [3, 4], which has shifted early target discovery toward genetically anchored approaches that identify genes whose perturbation in humans plausibly influences disease risk in a defined direction. This strategy is particularly suited to OA, where relevant tissues are difficult to sample and disease definitions vary across studies.

Genome-wide association studies (GWAS) have identified many OA-associated loci, but association signals alone rarely resolve an effector gene or its direction of effect. Because most trait-associated variants act through gene expression, integrating GWAS with expression quantitative trait loci (eQTL) offers a route from locus to candidate gene. Mendelian randomization (MR) uses genetic variants as instruments for gene expression to estimate its directional relationship with disease; colocalization tests whether the expression and disease signals share a causal variant, guarding against neighboring but distinct signals creating a spurious impression of shared mechanism; and summary-data-based MR (SMR), with its HEIDI test, provides an independent check for such heterogeneity. Agreement across these methods strengthens the case for a gene, while disagreement flags signals that require caution.

The chromosome 20 region containing GDF5 is among the most consistently replicated OA loci [5], and several neighboring genes including UQCC1, immediately adjacent to GDF5 map to the same association signal; because these genes share a block of linkage disequilibrium, an expression signal attributed to one may in fact reflect regulatory variation acting through another, and distinguishing among them is difficult when expression data are available for only some genes in the region. Separately, MAPK3 (encoding ERK1), on chromosome 16, participates in signaling pathways implicated in chondrocyte biology and joint homeostasis [6], but its directional relationship to OA risk has not been systematically examined across OA phenotypes using integrated genetic methods.

Two further considerations shaped this design. First, OA is not a single phenotype: knee OA, hip OA, their combination, and an overall OA definition differ in genetic architecture [7, 8], so a gene may be relevant to some joints but not others, and pooled analyses can obscure genuine, joint-specific effects. Second, the largest available cis-eQTL resources are derived from whole blood, which is not the primary disease tissue for OA, so expression-based inference here should be read as a genetically supported hypothesis rather than direct evidence in cartilage or synovium.

In this focused follow-up, we evaluated 14 genes within two OA associated regions on chromosomes 16 and 20 that .that had emerged from an earlier exploratory screen not discovered here using whole blood cis-eQTL data and four independently analyzed OA phenotypes. We combined cis-eQTL MR, Bayesian colocalization, and SMR with the HEIDI test, and graded each gene–phenotype signal using a structured evidence framework applied uniformly across all cells, while transparently reporting the phenotype-

specific inconsistencies and regional ambiguities most notably the inseparability of UQCC1 from the adjacent GDF5 locus in whole blood that constrain interpretation. This study asks: which genes at these two regions carry convergent, direction-resolved genetic evidence as candidate effector genes for osteoarthritis, and does resolving OA by phenotype change which genes are supported?

Methods

Study design

This was a focused follow-up genetic study of two OA-associated regions nominated by an earlier exploratory analysis, integrating cis-eQTL Mendelian randomization, Bayesian colocalization, and summary-data-based Mendelian randomization (SMR) with the HEIDI heterogeneity test. Fourteen candidate genes within two regions (chromosome 16 and chromosome 20) were evaluated against four OA phenotypes (AllOA, KneeHipOA, KneeOA, HipOA), yielding a 14 × 4 = 56 gene–phenotype grid. This is a targeted candidate-region analysis of pre-nominated loci, not a genome-wide discovery screen. The pipeline was deterministic (fixed random seed 20250715) and failed-close: a cell with fewer than 50 overlapping SNPs was skipped rather than analyzed on insufficient data. This study is reported in accordance with the STROBE-MR guideline for Mendelian randomization studies [27]. This Mendelian randomization relies on three core instrumental variable assumptions: (1) relevance — genetic instruments robustly predict the exposure; (2) independence — the instruments are unrelated to confounders of the expression–outcome association; and (3) exclusion restriction — the instruments affect the outcome only through the exposure. Assumption (1) was addressed by requiring genome-wide-significant ($p < 5\times10^{-8}$) cis-eQTL instruments; assumptions (2) and (3) cannot be directly verified but were assessed indirectly via the MR-Egger intercept test, colocalization, and the HEIDI test. A completed STROBE-MR checklist is provided as an additional file.

Candidate-region and gene selection

The two regions were not pre-registered; they were selected after inspection of an earlier exploratory cis-eQTL MR screen of the transcriptome across the four OA phenotypes, from which a chromosome 16 region containing MAPK3 and a chromosome 20 region spanning the GDF5/UQCC1 locus were carried forward. Within these regions, the gene set was defined data-drivenly: all genes with a testable whole-blood cis-eQTL probe in eQTLGen were included MAPK3 plus 13 genes in the chromosome 20 window (CEP250, UQCC1, ERGIC3, FER1L4, SPAG4, MMP24, MMP24-AS1, EDEM2, ROMO1, PROCR, RBM39, TRPC4AP, and PHF20), enumerated by proximity to the region lead SNP (rs143384). GDF5 itself has no cis-eQTL in eQTLGen whole blood and was therefore not testable. Because region and gene selection followed the exploratory results, all reported associations are focused follow-up analyses of pre-nominated loci, and the 56-cell grid, statistical thresholds, and evidence-grading framework were applied uniformly across all cells (Additional file 1: Section S.M.1, Table S1).

Data sources and instrument selection

Whole-blood cis-eQTL summary statistics were obtained from the eQTLGen Consortium (N = 31,684) [9]. OA outcome summary statistics were taken from the Genetics of Osteoarthritis (GO) consortium (Boer et al.) [7] for four independently analyzed phenotypes: AllOA (N = 826,690), KneeHipOA (N = 490,345), KneeOA (N = 396,054), and HipOA (N = 353,388), catalogued under GCST90061021-GCST90061039 (build hg19/GRCh37). Candidate instruments were cis-eQTL SNPs reaching $p < 5\times10^{-8}$ in eQTLGen, LD-clumped in PLINK (v1.90b7.2) [10] against a 1000 Genomes Phase 3 European reference (n = 503; --clump-r2 0.001 --clump-kb 1000). The eQTL and OA datasets were merged on a chromosome: position key and harmonized by allele identity; palindromic SNPs were not explicitly excluded, a limitation discussed below. Full data-processing, quality-filtering, and harmonization detail is provided in Additional file 1 (Sections S.M.2–S.M.4). In summary, the analyzed sample sizes were n = 31,684 for the whole-blood cis-eQTL exposure (eQTLGen) and n = 826,690 (AllOA; case proportion s = 0.2147), n = 490,345 (KneeHipOA; s = 0.1830), n = 396,054 (KneeOA; s = 0.1578), and n = 353,388 (HipOA; s = 0.1031) for the four OA outcome phenotypes.

Statistical methods

MR was performed with TwoSampleMR (v0.7.9) [11] on the harmonized data: the Wald ratio for a single instrument, inverse-variance weighting (IVW) only for two instruments [12], and IVW together with MR-Egger [13] and weighted median [14] estimators (plus the Egger intercept test for directional pleiotropy) for three or more. Single-causal-variant Bayesian colocalization was performed with coloc.abf (coloc v5.2.3) [15] on the full harmonized regional SNP set per cell, using default priors ($p1 = p2 = 1\times10^{-4}$, $p12 = 1\times10^{-5}$) and recording posterior probabilities PP.H0–PP.H4 (PP.H4, shared causal variant; PP.H3, distinct variants). As a fine-mapping sensitivity analysis, coloc.susie [16, 17] was run when both sides converged; because a substantial fraction of cells did not converge under the out-of-sample LD reference, a post hoc windowed coloc.susie analysis (± 100/±250/±500 kb, with staged $r^2 > 0.99$ pruning and matrix repair) was additionally performed for MAPK3 and UQCC1 (Additional file 1: Section S.M.7, Table S7). SMR and the HEIDI heterogeneity test [18] were run with SMR v1.3.1 using the eQTLGen BESD and the same LD reference (--peqtl-smr 5×10^{-8} --cis-wind 2000 --diff-freq-prop 0.3); a HEIDI $p \le 0.05$ was interpreted as evidence against a single shared variant. Two post hoc sensitivity analyses targeted the two most attackable signals: a window and leave-one-SNP-out analysis for MAPK3, and a region-specific GCTA-COJO (v1.94.1) [19, 20] conditional analysis with a multi-SNP SMR cross-check for UQCC1 × KneeOA, where coloc.abf favoured distinct variants and HEIDI rejected a shared variant (Additional file 1: Section S.M.9, Table S9).

Each gene–phenotype cell was graded (Level 1–3) using a structured, post-analysis evidence framework combining coloc.abf PP.H4, concordant and significant MR and SMR estimates, HEIDI, and robustness to sensitivity analyses; thresholds were not pre-registered. Direction of effect was taken from the sign of concordant MR and SMR estimates. The full 56-cell grid was executed twice under different random

seeds, in isolated output directories, to confirm determinism (Additional file 1: Section S.M.10, Table S10); MR was reported only for cells with at least one genome-wide-significant instrument. Analyses used R (v4.4.3; coloc 5.2.3, susieR 0.14.2, TwoSampleMR 0.7.9), PLINK (v1.90b7.2), GCTA (v1.94.1), and SMR (v1.3.1); no formal multiple-testing correction was applied across the 56 cells, and the evidence-grading framework rather than a single adjusted threshold was used to rank signals (Additional file 1: Section S.M.11).

Results

Overview of the 56-cell evidence grid

This focused follow-up evaluated a chromosome 16 region containing MAPK3 and a chromosome 20 region spanning the GDF5/UQCC1 locus, both nominated by an earlier exploratory screen; neither gene was newly discovered here. Every gene with a testable whole-blood cis-eQTL probe in these regions was evaluated, giving a 14 × 4 = 56 gene phenotype grid. Of the 56 cells, 48 produced an MR result; the eight without an MR result were all four phenotypes each for ROMO1 and SPAG4, which had no cis-eQTL reaching $p < 5\times10^{-8}$, whereas coloc.abf, coloc.susie, and SMR were generated for all 56 cells (Additional file 1: Table S2). No formal multiple-testing correction was applied within the grid, so nominal p-values should not be over-interpreted (Fig. 1).

MAPK3: a protective candidate across OA phenotypes

MAPK3 was the primary prioritized candidate. Across all four phenotypes, both MR and SMR estimates were negative, consistent with a protective direction in which higher predicted MAPK3 blood expression is associated with lower OA risk (Table 1; Fig. 2). Colocalization supported a shared causal variant for KneeHipOA, KneeOA, and AllOA (PP.H4 = 0.991, 0.979, and 0.945, respectively), with concordant, nominally significant MR-IVW estimates ($p = 7.1\times10^{-4}$, 1.0×10^{-4}, and 7.5×10^{-4}) and significant, protective SMR estimates with HEIDI not rejecting a shared variant (p = 0.525–0.765). In these three phenotypes the MR-Egger intercept showed no evidence of directional pleiotropy (intercept p = 0.61–0.89), and weighted-median estimates were concordant with IVW. For HipOA, colocalization (PP.H4 = 0.825) and SMR ($b = -0.0704$, $p = 1.7\times10^{-4}$; HEIDI p = 0.154) remained supportive, but the MR-IVW estimate was not significant (p = 0.091); we therefore do not claim MR support for MAPK3 in HipOA, and that signal rests on colocalization and SMR alone.

coloc.susie did not confirm colocalization for MAPK3 in any phenotype: all four main-grid cells returned NO_COLOC_PAIR, and in the windowed analysis SuSiE converged on both sides but produced zero OA credible sets. This reflects the sub-genome-wide-significant OA association at this locus rather than a contradiction of coloc.abf; MAPK3 is supported by coloc.abf and SMR, not by fine-mapping.

UQCC1: a phenotype-dependent risk candidate

UQCC1 was a phenotype-dependent secondary candidate with a risk-increasing direction (positive MR and SMR estimates); its support differed markedly across phenotypes (Table 1; Fig. 3). Evidence was strongest for AllOA (PP.H4 = 0.919, PP.H3 = 0.081; SMR p = 7.6×10^{-12}; HEIDI p = 0.192) and supportive but incomplete for KneeHipOA (PP.H4 = 0.902; SMR p = 8.2×10^{-19}; HEIDI p = 0.110), where fine mapping colocalization could not be evaluated because SuSiE did not converge on the eQTL side. For KneeOA, coloc.abf favored distinct causal variants (PP.H3 = 0.593 > PP.H4 = 0.407) and HEIDI rejected a shared variant (p = 1.84×10^{-3}) despite very strong marginal MR and SMR (p = 1.0×10^{-22} and 5.2×10^{-20}); we do not treat this cell as positive colocalization evidence. HipOA showed only weak evidence (PP.H4 = 0.250). Across all phenotypes, GDF5 immediately adjacent and lacking a whole-blood cis-eQTL in eQTLGen—cannot be tested or excluded as the true regional effector; the UQCC1 associations are consistent with, but do not establish, UQCC1 as the causal gene.

Other genes in the two regions

No other gene met criteria for a supported colocalized signal in any phenotype. ROMO1 and SPAG4 had no instrument and low coloc.abf posteriors (PP.H4 < 0.10). CEP250, ERGIC3, MMP24, MMP24-AS1, and RBM39 produced nominally significant MR and/or SMR estimates in one or more phenotypes but with coloc.abf PP.H3 approaching 1.0 and PP.H4 near zero the classic pattern of significant marginal MR unsupported by colocalization. EDEM2, PROCR, TRPC4AP, and PHF20 showed low PP.H4 and mostly non-significant or inconsistent MR. Full per-gene values for all 56 cells, including these null and PP.H3-dominated outcomes, are provided in Additional file 1: Table S2.

Sensitivity and conditional analyses

The MAPK3 coloc.abf signal was robust to window choice (full region and ± 100/±250/±500 kb; PP.H4 essentially unchanged) and to leave-one-SNP-out removal of the top-20 SNPs by OA or eQTL |Z| (no single-SNP dependence in any phenotype), showing the posterior is not an artifact of window choice or any individual variant, though this does not by itself prove a shared causal variant (Additional file 1: Table S8). For UQCC1 × KneeOA, a region-specific GCTA-COJO conditional analysis on chromosome 20 identified a single independent genome-wide-significant signal; conditioning on the UQCC1 eQTL lead SNP (rs4911178) left no SNP genome-wide significant (minimum conditional p = 1.5×10^{-4}), and a multi-SNP SMR cross-check confirmed a strong effect (p = 1.3×10^{-15}) while HEIDI remained rejecting (p = 1.84×10^{-3}). Under the disposition rule fixed before this analysis, a single independent signal means the cell is retained only as a heterogeneity-flagged secondary result rather than promoted to positive colocalization evidence; this does not resolve whether UQCC1 or GDF5 is the effector (Additional file 1: Table S9).

Fine mapping (coloc.susie)

Coloc.susie was applied as an exploratory fine-mapping sensitivity analysis and is reported honestly, including its high non-convergence rate; it functions as a bound on confidence rather than a source of independent confirmation. In the main grid, 10 of 56 cells returned a colocalizing credible-set pair (none

among MAPK3 or UQCC1), 13 converged with no colocalizing pair, and 33 did not converge, attributable to the out-of-sample 1000G EUR reference (n = 503), which is rank-deficient for these dense regional SNP sets. In the windowed analysis, all 12 MAPK3 windows converged on both sides but returned zero OA credible sets; for UQCC1, a single cell (UQCC1 × AllOA, ± 500 kb, PP.H4 = 0.993) converged after stage-2 $r^2 > 0.99$ pruning and is reported as an exploratory result, not independent confirmation of coloc.abf, while all other UQCC1 windows failed to converge on the eQTL side (Additional file 1: Table S7).

Evidence synthesis and reproducibility

Applying the structured evidence framework, MAPK3 reached the top level for KneeHipOA, KneeOA, and AllOA, and one level lower for HipOA because its MR-IVW estimate was not significant. UQCC1 reached the top level for AllOA and the second level for KneeHipOA, where SuSiE was not evaluable; UQCC1 × KneeOA was graded lowest (PP.H3-dominated with HEIDI rejection), as was UQCC1 × HipOA (Table 1). No gene other than MAPK3 and UQCC1 was assigned a supporting grade in any phenotype. The pipeline was deterministic and reproducible: an independent re-execution under a different seed (99999777), covering a representative 12-cell subset, reproduced all coloc.abf, MR, and SMR/HEIDI outputs exactly (maximum absolute between-run difference = 0), with the stochastic coloc.susie component seed-stable across the subset (Additional file 1: Table S10).

Table 1
Prioritization evidence matrix for the two lead-region candidate genes.

Gene / locus	Phenotype	OR (95% CI)	MR-IVW P	PP.H4	PP.H3	P_SMR	P_HEIDI	Tier
MAPK3 (chr16)	AllOA	0.965 (0.946–0.985)	7.46e-04	0.945	0.038	7.94e-06	0.525	Level 1
MAPK3 (chr16)	KneeHipOA	0.948 (0.919–0.978)	7.07e-04	0.991	0.008	3.39e-07	0.725	Level 1
MAPK3 (chr16)	KneeOA	0.927 (0.892–0.963)	1.04e-04	0.979	0.021	2.59e-07	0.765	Level 1
MAPK3 (chr16)	HipOA	0.962 (0.920–1.006)	9.15e-02	0.825	0.029	1.68e-04	0.154	Level 2
UQCC1/GDF5 (chr20)	AllOA	1.068 (1.043–1.094)	6.04e-08	0.919	0.081	7.58e-12	0.192	Level 1
UQCC1/GDF5 (chr20)	KneeHipOA	1.136 (1.107–1.166)	1.11e-21	0.902	0.098	8.23e-19	0.110	Level 2
UQCC1/GDF5 (chr20)	KneeOA	1.164 (1.129-1.200)	1.01e-22	0.407	0.593	5.16e-20	1.84e-03	Level 3
UQCC1/GDF5 (chr20)	HipOA	1.073 (1.023–1.125)	4.04e-03	0.250	0.162	2.70e-03	0.965	Level 3

Per gene-locus and OA phenotype: MR effect (OR with 95% CI), MR-IVW p-value, coloc.abf PP.H4 and PP.H3, SMR p-value, HEIDI p-value, and the structured post-analysis evidence tier. All p-values are nominal; no false-discovery-rate or Bonferroni correction was applied across the 56 gene-phenotype cells. Phenotype order: AllOA, KneeHipOA, KneeOA, HipOA. Genomic build GRCh37.

Discussion

Principal findings

In this focused follow-up of two osteoarthritis (OA)-associated regions a chromosome 16 region containing MAPK3 and a chromosome 20 region spanning the GDF5/UQCC1 locus both nominated by an earlier exploratory genome-wide screen, we re-examined 14 genes across four OA phenotypes using cis-eQTL Mendelian randomization (MR), colocalization (coloc.abf and coloc.susie), and summary-data-based MR (SMR) with the HEIDI test. MAPK3 emerged as the primary prioritized candidate, with a

consistent protective direction and colocalization support in three of four phenotypes, and UQCC1 emerged as a phenotype-dependent secondary candidate whose evidence was strongest for AllOA. Neither gene should be read as a discovery from an independent, unbiased screen: both entered this analysis because they had already been flagged, so these results characterize and grade evidence within pre-nominated regions rather than identifying new loci. We frame the findings as exploratory target prioritization supported by cross-method consistency, not as definitive statistical discovery.

Interpretation of MAPK3

MAPK3 (encoding ERK1) is a component of the RAS–RAF–MEK–ERK signaling cascade with established roles in chondrocyte proliferation, differentiation, and cartilage homeostasis [6]. Its identical protective direction across all four phenotypes for both MR and SMR, together with colocalization support in three phenotypes and cross-method agreement between MR, coloc.abf, and SMR, is what supports its prioritization. Two caveats bound this interpretation. First, the OA regional association at the MAPK3 locus was sub-genome-wide-significant, so coloc.susie produced no OA credible set in any phenotype; MAPK3 is therefore supported by coloc.abf and SMR rather than by fine-mapping, and these are not contradictory findings coloc.susie simply had no OA credible set to work with. Second, although the coloc.abf posterior was robust to window choice and leave-one-out removal of individual SNPs, this demonstrates stability of the result, not proof of a shared causal variant. For HipOA specifically, the non-significant MR-IVW estimate means the signal there rests on colocalization and SMR alone. Taken together, the evidence is consistent with a protective role for MAPK3 in OA and supports its prioritization as a candidate, while stopping short of establishing it as a causal or validated target.

Interpretation of UQCC1 and the adjacent GDF5 locus

UQCC1 lies immediately adjacent to GDF5, one of the most robustly replicated OA-associated loci [5], and this physical and linkage disequilibrium proximity is the central interpretive challenge for the UQCC1 signal. Its strongest evidence was for AllOA (strong coloc.abf, a strongly significant positive SMR estimate, and HEIDI not rejecting a shared variant), with KneeHipOA supportive but incomplete because fine mapping colocalization could not be evaluated. The windowed coloc.susie result for UQCC1 × AllOA (PP.H4 = 0.993) was obtained only at the ± 500 kb window after pruning of near-duplicate SNPs and is an exploratory result, not independent confirmation of the coloc.abf finding. UQCC1 × KneeOA was PP.H3-dominated with a HEIDI rejection of the single-shared-variant model, so despite very strong marginal MR and SMR it was not treated as positive colocalization evidence; UQCC1 × HipOA provided only weak evidence. Critically, the exposure here is whole-blood eQTL, and GDF5 has no cis-eQTL in eQTLGen whole blood, so it could not be instrumented or tested, and no analysis here can exclude GDF5 as the true regional effector or establish UQCC1 in its place. The region-specific conditional analysis identified a single independent OA signal, which locates the association but does not assign it to a gene. We do not claim UQCC1 is superior to GDF5 or that it is the causal gene; the UQCC1 associations are consistent with a genetically supported signal in this region while remaining regionally ambiguous with respect to GDF5.

Phenotype-specific heterogeneity

Analyzing four OA phenotypes separately revealed heterogeneity that a single pooled analysis would have masked. For MAPK3, colocalization and SMR support were consistent for KneeHipOA, KneeOA, and AllOA but attenuated for HipOA, where MR was non-significant, consistent with known epidemiological and genetic differences between knee and hip OA [21]. For UQCC1, the phenotype contrast was more pronounced: AllOA gave the strongest and internally consistent evidence, whereas KneeOA showed the paradoxical combination of extremely significant marginal MR/SMR estimates with colocalization pointing away from a shared variant and a HEIDI rejection among the most significant estimates in the entire grid, yet not constituting colocalization support. This illustrates why marginal association significance is insufficient on its own, and why the phenotype-resolved design both prevents over-generalization from AllOA to joint-specific phenotypes and surfaces heterogeneity that would otherwise be hidden.

Translational implications and future validation

If corroborated in disease-relevant systems, the prioritized candidates would have differing translational profiles. MAPK3/ERK1 is a pharmacologically tractable kinase within a pathway already targeted by MEK and ERK inhibitors in other indications [22], and the protective direction observed here where higher expression associates with lower OA risk would need careful directional interpretation before any therapeutic hypothesis, since it implies that augmenting rather than inhibiting the signal might be protective. The UQCC1/GDF5 region is biologically compelling given GDF5's established role in skeletal and joint development [5], but the inability to separate UQCC1 from GDF5 in blood means that target assignment must precede any intervention hypothesis. Priorities for validation are therefore replication in independent OA GWAS and ancestrally diverse cohorts; colocalization and eQTL/MR analysis using cartilage, synovial, and bone eQTL resources to test tissue relevance and to help resolve UQCC1 versus GDF5 [23]; and functional studies in chondrocyte and joint-tissue models to test the direction and mechanism of effect.

Strengths and limitations

Several design choices strengthen the reliability of this prioritization: results were reported phenotype-by-phenotype rather than pooled, exposing heterogeneity across AllOA, KneeHipOA, KneeOA, and HipOA; three methodologically distinct lines of genetic evidence MR, coloc.abf, and SMR/HEIDI were triangulated so that a candidate was prioritized only when independent approaches agreed; inconsistent, null, non-convergent, and heterogeneity flagged findings were reported transparently rather than suppressed; and targeted sensitivity analyses probed the stability of the leading signals, with the pipeline deterministic and reproducible under a two seed re-execution (Additional file 1: Table S10). This computational reproducibility demonstrates that the analysis is stable and faithfully reported; it is not biological validation and does not establish that the prioritized genes are causal. Several limitations remain. Because the two regions were selected after inspection of the earlier exploratory screen, the reported estimates are susceptible to selection bias and winner's curse [24], and effect sizes may be inflated relative to what independent data would show. The exposure is whole-blood eQTL rather than disease relevant cartilage, synovium, or bone [23]; regulatory effects can be tissue-specific, so a blood-

derived instrument may not reflect the causal tissue for OA. Possible sample overlap between the eQTL and OA GWAS datasets was not quantified and could bias MR and SMR estimates toward the observational association [25]. The LD reference was the 1000 Genomes EUR panel (n = 503), an out-of-sample reference that is rank-deficient for dense regional SNP sets; this is the most likely driver of the high coloc.susie non-convergence rate (33 of 56 cells) and may also affect HEIDI and conditional analyses that depend on accurate LD. UQCC1 was instrumented by only three cis-eQTLs, limiting the power and pleiotropy diagnostics of its MR analysis. No external replication cohort was analyzed, and no functional or experimental validation was performed, so all conclusions are genetic and in silico. The targeted sensitivity analyses were post hoc. Palindromic variants were handled during harmonization but not explicitly excluded, which could in principle allow residual strand ambiguity [11]. Finally, the analysis was restricted to European-ancestry data, limiting generalizability to other ancestries [26].

Conclusion

Within a focused follow up of two pre-nominated OA regions, and without correction for multiple testing across the 56 evaluated cells, integrated genetic evidence prioritizes MAPK3 as the more consistently supported candidate, showing a protective direction and colocalization support in three of four phenotypes, while UQCC1 remains a phenotype dependent secondary candidate whose signal is strongest for AllOA and regionally ambiguous with respect to the adjacent GDF5 locus. These findings are exploratory prioritizations grounded in cross method consistency rather than confirmed causal relationships, and both candidates require validation in disease-relevant tissues and experimental systems before any causal or therapeutic claim can be made.

Abbreviations

OA Osteoarthritis

GWAS Genome-wide association study

eQTL Expression quantitative trait locus

cis-eQTL Cis-acting expression quantitative trait locus

MR Mendelian randomization

IVW Inverse-variance weighted

SMR Summary-data-based Mendelian randomization

HEIDI Heterogeneity in dependent instruments

SNP Single-nucleotide polymorphism

LD Linkage disequilibrium

OR Odds ratio

CI Confidence interval

Declarations

Ethics approval and consent to participate

Not applicable. This study used publicly available, de-identified summary-level genetic data; no new human participant data were collected.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Role of the funding source

The funding sources had no role in the study design; collection, analysis, or interpretation of data; preparation of the manuscript; or the decision to submit the manuscript for publication.

Clinical trial number

Not applicable.

Funding

This work was supported by the Anhui Health Research (AHWJ2022b113), the Basic and Clinical Collaboration of Anhui Medical University (2022sfy004), the Anhui Traditional Chinese Medicine Inheritance and Innovation Research (2024CCCX287), and the Hefei Natural Science Foundation Project (HZR2548).

Author Contribution

Jiahao Ma: performed the Mendelian randomization and integrated statistical analyses. Shiyu Yang: contributed to data curation, interpretation of the genetic evidence, and manuscript preparation. Yibin Du: Project administration, Resources, Writing review & editing.

Acknowledgements

The authors gratefully acknowledge the eQTLGen Consortium and the Genetics of Osteoarthritis Consortium for making their summary-level genetic data publicly available. The authors also thank all investigators and participants who contributed to the original studies.

Data Availability

All data analyzed in this study were obtained from publicly available resources. Whole-blood cis-eQTL summary statistics were obtained from the eQTLGen Consortium, and osteoarthritis GWAS summary statistics were obtained from the Genetics of Osteoarthritis Consortium. The datasets and their access information are available through the corresponding consortium resources and the publications cited in the manuscript. Supplementary methods, tables (Table S1–S10), and figures (Figure S1–S6) are provided in Additional file 1. The analysis code and processed results supporting the findings of this study are available from the corresponding author upon reasonable request.

Additional files

Additional file 1: Supplementary Methods (S.M.1–S.M.11), Supplementary Tables S1–S10, and Supplementary Figures S1–S6, providing full provenance, data-processing, and sensitivity-analysis detail underlying the Methods and Results, including the complete 56-cell result grid (Table S2), instrument- and posterior-level detail (Tables S3–S6), coloc.susie window diagnostics (Table S7), MAPK3 window/leave-one-out sensitivity results (Table S8), the UQCC1 conditional analysis (Table S9), and the two-run reproducibility comparison (Table S10).

References

1. GBD 2021 Osteoarthritis Collaborators, Steinmetz JD, Culbreth GT, et al. Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Rheumatol. 2023;5(9):e508–22. 10.1016/S2665-9913(23)00163-7. . PMID:37675071.
2. Cho Y, Jeong S, Kim H, et al. Disease-modifying therapeutic strategies in osteoarthritis: current status and future directions. Exp Mol Med. 2021;53(11):1689–96. 10.1038/s12276-021-00710-y. . PMID:34799687.
3. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015;47(8):856–60. 10.1038/ng.3314. . PMID:26121088.
4. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLoS Genet. 2019;15(12):e1008489. 10.1371/journal.pgen.1008489. PMID:31830040.

5. Miyamoto Y, Mabuchi A, Shi D, et al. A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. Nat Genet. 2007;39(4):529–33. 10.1038/2005. . PMID:17384641.
6. Lu N, Malemud CJ. Extracellular signal-regulated kinase: a regulator of cell growth, inflammation, chondrocyte and bone cell receptor-mediated gene expression. Int J Mol Sci. 2019;20(15):3792. 10.3390/ijms20153792. . PMID:31382554.
7. Boer CG, Hatzikotoulas K, Southam L, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell. 2021;184(18):4784–e481817. 10.1016/j.cell.2021.07.038. . PMID:34450027.
8. Zengini E, Hatzikotoulas K, Tachmazidou I, et al. Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. Nat Genet. 2018;50(4):549–58. 10.1038/s41588-018-0079-y. . PMID:29559693.
9. Vosa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet. 2021;53(9):1300–10. 10.1038/s41588-021-00913-z. . PMID:34475573.
10. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7. 10.1186/s13742-015-0047-8. . PMID:25722852.
11. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408. 10.7554/eLife.34408. . PMID:29846171.
12. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658–65. 10.1002/gepi.21758. . PMID:24114802.
13. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512–25. 10.1093/ije/dyv080. . PMID:26050253.
14. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40(4):304–14. 10.1002/gepi.21965. . PMID:27061298.
15. Giambartolomei C, Vukcevic D, Schadt EE et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383. 10.1371/journal.pgen.1004383. PMID:24830394.
16. 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 Ser B Stat Methodol. 2020;82(5):1273–300. 10.1111/rssb.12388.
17. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the Sum of Single Effects model. PLoS Genet. 2022;18(7):e1010299. 10.1371/journal.pgen.1010299. PMID:35853082.
18. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48(5):481–7. 10.1038/ng.3538. . PMID:27019110.

19. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82. PMID:21167468.
20. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012;44(4):369–75. 10.1038/ng.2213. . PMID:22426310.
21. Henkel C, Styrkarsdottir U, Thorleifsson G, et al. Genome-wide association meta-analysis of knee and hip osteoarthritis uncovers genetic differences between patients treated with joint replacement and patients without joint replacement. Ann Rheum Dis. 2023;82(3):384–92. 10.1136/ard-2022-223199. . PMID:36600186.
22. Ram T, Singh AK, Kumar A, et al. MEK inhibitors in cancer treatment: structural insights, regulation, recent advances and future perspectives. RSC Med Chem. 2023;14(9):1704–23. 10.1039/d3md00145h. . PMID:37731690.
23. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318–30. 10.1126/science.aaz1776. . PMID:32913098.
24. Xiao R, Boehnke M. Quantifying and correcting for the winner's curse in genetic association studies. Genet Epidemiol. 2009;33(5):453–62. 10.1002/gepi.20398. . PMID:19140130.
25. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40(7):597–608. 10.1002/gepi.21998. . PMID:27625185.
26. Martin AR, Kanai M, Kamatani Y, et al. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584–91. 10.1038/s41588-019-0379-x. . PMID:30926966.
27. Skrivankova VW, Richmond RC, Woolf BAR et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. JAMA. 2021;326(16):1614–1621. 10.1001/jama.2021.18236. PMID:34698778.

Figures

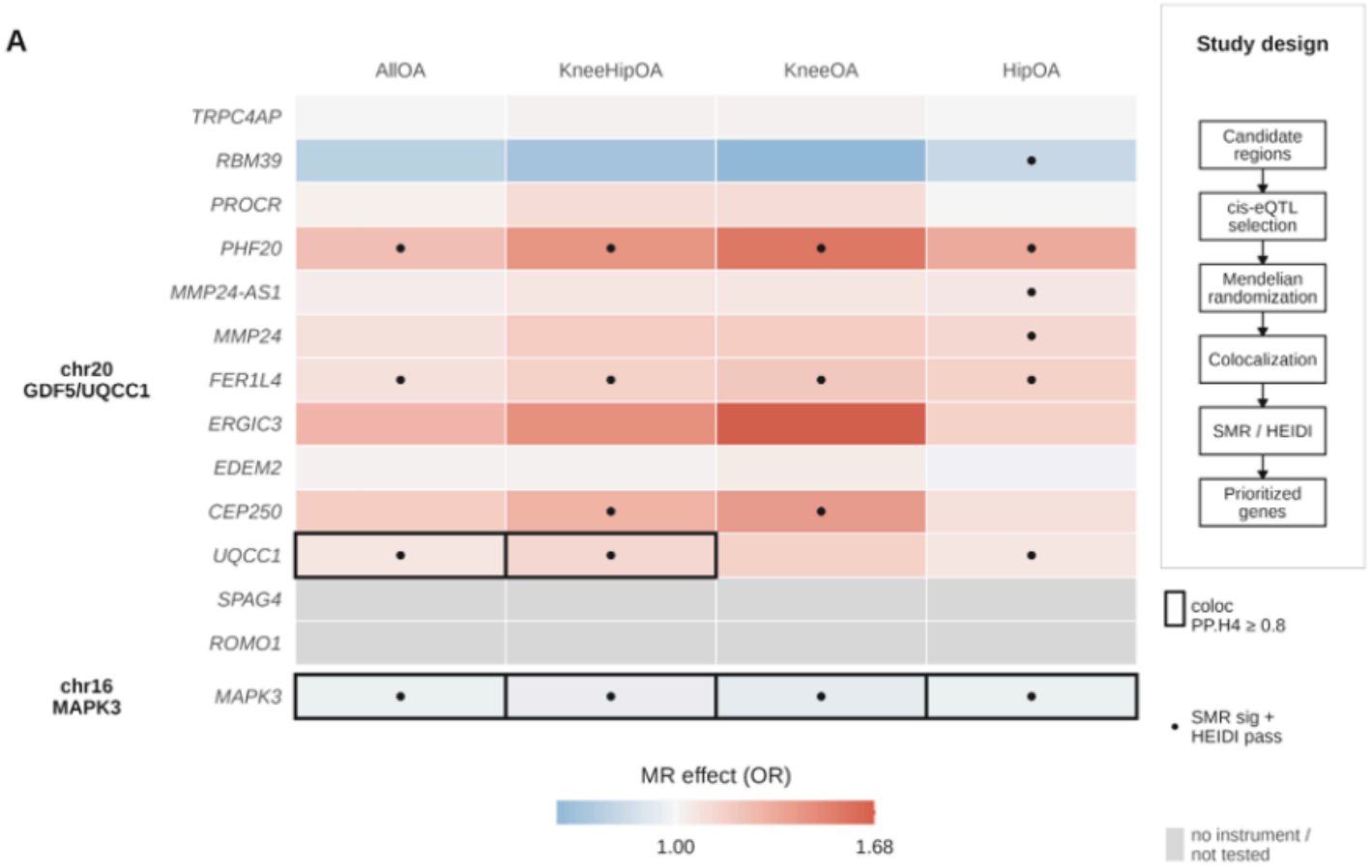


Figure 1

Phenotype-resolved evidence overview across 14 genes and four OA phenotypes, with analysis workflow inset. Merged evidence heatmap of all 56 gene-phenotype cells for the two study regions: the chromosome 16 MAPK3 locus (1 gene) and the chromosome 20 GDF5/UQCC1 locus (13 genes), evaluated against four independently analyzed OA phenotypes (AllOA, KneeHipOA, KneeOA, HipOA). Cell fill encodes the signed MR effect on a diverging scale centered at OR = 1 (protective, blue; risk, red); null-gray marks cells with no genome-wide-significant cis-eQTL instrument (ROMO1 and SPAG4). A bold cell border/dot flags strong colocalization (coloc.abf PP.H4 >= 0.8), and a separate glyph flag significant SMR with HEIDI not rejecting a shared variant (P_HEIDI > 0.05). An orientation insect shows the six-step analysis workflow: candidate regions -> cis-eQTL selection -> Mendelian randomization -> colocalization -> SMR/HEIDI -> prioritized genes. Strong colocalization (PP.H4 >= 0.8) is confined to MAPK3 (all four phenotypes) and UQCC1 (AllOA and KneeHipOA).

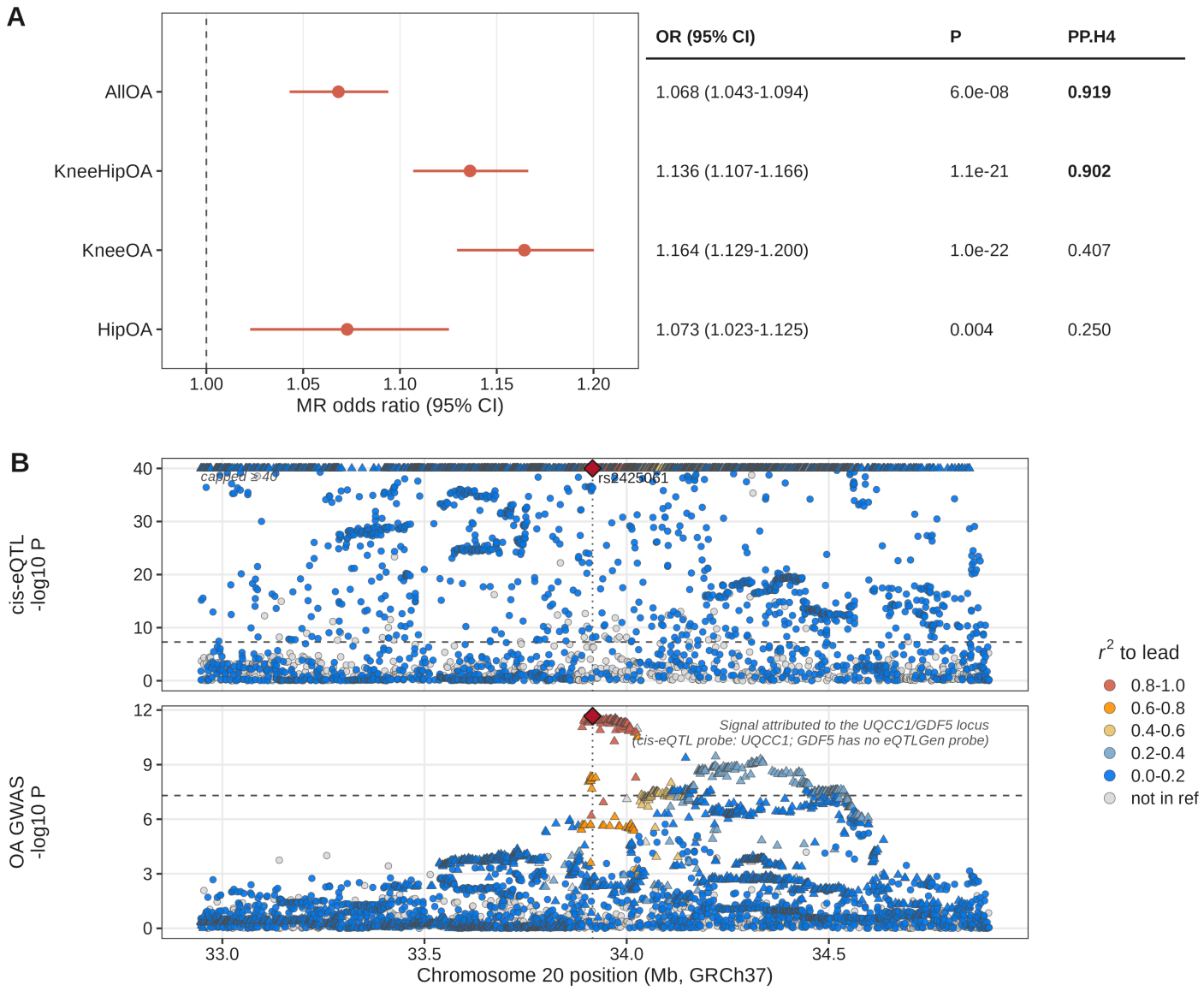


Figure 2

UQCC1/GDF5 (chromosome 20) secondary-locus spotlight, matched in layout to Figure 2. (A) Forest plot of MR effect estimates on OA across the four phenotypes, shown as OR (95% CI) per SD-standardized increase in UQCC1 cis-eQTL-predicted expression. All four UQCC1 estimates are significant and risk-increasing (OR > 1). (B) LD-colored regional colocalization plot for the representative phenotype (AllOA) across the locus on chromosome 20 (GRCh37); LD from the 1000 Genomes Phase 3 European reference (n = 503). Colocalization support is phenotype-dependent: strong for AllOA (PP.H4 = 0.919) and KneeHipOA (PP.H4 = 0.902) and not supported for KneeOA (PP.H3 = 0.593 > PP.H4 = 0.407, HEIDI rejecting, p = 1.84×10^{-3}) or HipOA (PP.H4 = 0.250). The signal is attributed to the UQCC1/GDF5 locus; GDF5 has no whole-blood cis-eQTL probe in eQTLGen and cannot be excluded as the true regional effector.

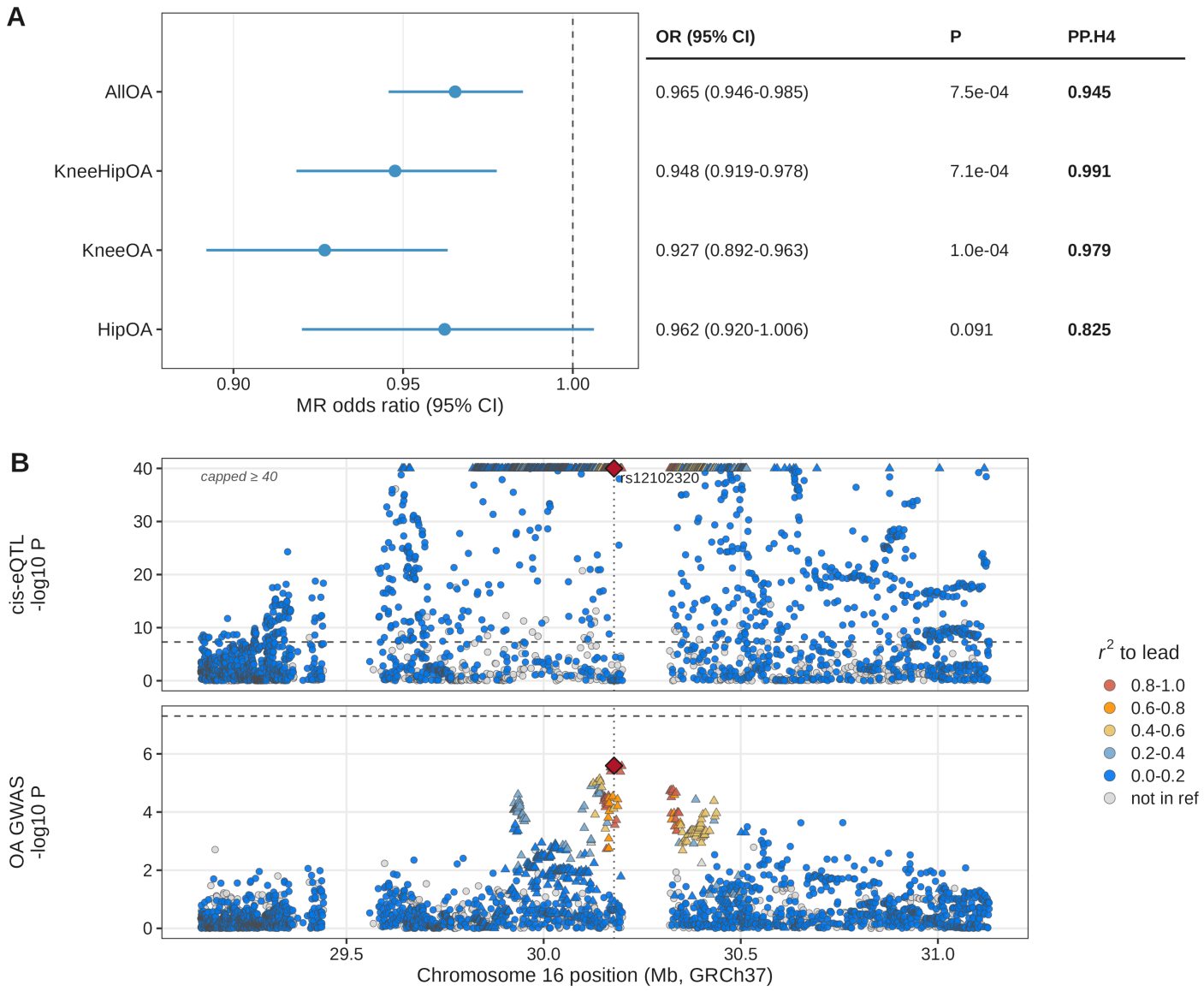


Figure 3

MAPK3 (chromosome 16) lead-locus spotlight. (A) Forest plot of MR effect estimates on OA across the four phenotypes, shown as OR (95% CI) per SD-standardized increase in MAPK3 cis-eQTL-predicted expression. Filled points indicate a nominally significant IVW estimate ($p < 0.05$); the open point marks the only non-significant MAPK3 estimate, MAPK3 × HipOA (IVW $p = 0.092$). All MAPK3 estimates are protective (OR < 1). (B) LD-colored regional colocalization plot for the representative phenotype (AllOA) across the MAPK3 locus on chromosome 16 (GRCh37); LD from the 1000 Genomes Phase 3 European reference (n = 503). Colocalization support for MAPK3 rests on coloc.abf together with SMR; across all four phenotypes coloc.susie returned no OA credible set (NO_COLOC_PAIR).

Supplementary Files

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

- FigureS1.tiff
- FigureS2.tiff
- FigureS3.tiff
- FigureS4.tiff
- FigureS5.tiff
- FigureS6.tiff
- SupplementaryMaterial.docx
- TableS1geneprovenance.csv
- TableS2all56cells.csv
- TableS3instrumentlevelciseQTL.csv
- TableS4MRestimators.csv
- TableS5colocabfposteriors.csv
- TableS6SMRHEIDI.csv
- TableS7colocsusiewindowdiagnostics.csv
- TableS8aMAPK3abfwindow.csv
- TableS8bMAPK3abfLOOsummary.csv
- TableS8cMAPK3abfLOOfull.csv
- TableS9aUQCC1COJOindependentsignals.txt
- TableS9bUQCC1conditionaloneqtllead.txt
- TableS9cUQCC1SMRmulti.csv
- TableS10tworunreproducibility.txt
- STROBEMRChecklistREVISED.docx