

Cell-resolved transcriptomics separates the components of a pathway- wide PI3K–AKT response in dilated cardiomyopathy: the AKT1 change localizes to cardiomyocytes, the PIK3CA change peaks in fibroblasts

Yingying Qian

The Sixth Medical Center of PLA General Hospital

Haitao Tian

The Sixth Medical Center of PLA General Hospital

Shuaihong Zhao

The Sixth Medical Center of PLA General Hospital

Yanjiao Bian

The Sixth Medical Center of PLA General Hospital

Hongyang Yu

The Sixth Medical Center of PLA General Hospital

Ang Di

The Sixth Medical Center of PLA General Hospital

Yingshuang Jia

The Sixth Medical Center of PLA General Hospital

Shaodan Li

1sd301@126.com

The Sixth Medical Center of PLA General Hospital

Runquan Sun

The Sixth Medical Center of PLA General Hospital

Research Article

Keywords: dilated cardiomyopathy, myocarditis, PI3K–AKT signalling, single-nucleus RNA sequencing, donor pseudobulk, network pharmacology

Posted Date: August 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10584049/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Functional & Integrative Genomics on October 3rd, 2026. See the published version at <https://doi.org/10.1007/s10142-026-02059-9>.

Abstract

Network pharmacology nominates targets but cannot show that they are active in the relevant cell or linked to disease. Using Yangxin Decoction in dilated cardiomyopathy (DCM) and viral myocarditis, we tested which nominated PI3K–AKT candidates survive independent functional-genomic evidence. Most did not. A direct-evidence intersection returned 61 shared genes, but across four bulk cardiac datasets 20 of the 61 genes reached a within-dataset false-discovery rate below 0.05, no more than the transcriptome-wide background in any human dataset, and *AKT1* reached significance in none. Donor-pseudobulk single-nucleus RNA sequencing resolved this. In DCM cardiomyocytes, *AKT1* (log2 fold change 0.351, adjusted $P = 3.32 \times 10^{-4}$), *PIK3CA* and *GSK3B* were higher and *STAT3* lower (52 cases, 18 controls), and the effects held with donor sequencing depth as a covariate, though a residual depth term remained for two genes. Resolving every population separated a pathway-wide response from a lineage-restricted node: among annotated cell types the *AKT1* change was confined to cardiomyocytes, changing elsewhere only in an unannotated group, whereas *PIK3CA* rose in eight of the nine populations and peaked in fibroblasts, not cardiomyocytes. Both are invisible to bulk tissue and to pathway enrichment, which average across cells. A prespecified screen found no qualifying instrument for *AKT1*; the single whole-blood *MTOR* estimate (adjusted $P = 0.0405$) is not positive evidence, and its colocalization was uninformative (posterior probabilities $PP.H1 = 0.559$, $PP.H4 = 0.230$). This is a pathway-wide PI3K–AKT transcriptional response carrying a cardiomyocyte-restricted *AKT1* component, not a validated therapeutic or causal mechanism.

Introduction

Dilated cardiomyopathy (DCM) is defined by ventricular dilatation and systolic dysfunction not explained solely by abnormal loading or coronary disease, and it is a major cause of heart failure and transplantation (Japp et al. 2016; Reichart et al. 2019). Myocarditis is an inflammatory myocardial disorder that can progress to chronic ventricular remodelling and DCM, although its initiating causes and clinical course are heterogeneous (Imanaka-Yoshida 2020; Kawai 1999). Their overlap lies in downstream injury, immune activation, fibrosis and contractile failure rather than in an identical disease mechanism.

The traditional principle of treating different diseases with the same therapy proposes that distinct diagnoses may share a therapeutically relevant biological state (Zhai et al. 2020). Yangxin Decoction has been used under this principle for cardiomyopathy and myocarditis. Establishing a molecular explanation is difficult because a multi-herb formula contains many compounds, predicted protein targets and indirect disease associations. Network pharmacology is useful for organizing these relationships and prioritizing candidates, but its edges are largely database-derived or computationally predicted (Hopkins 2008; Li and Zhang 2013; Nogales et al. 2022). Enrichment and centrality can identify an attractive hypothesis; they do not show that the nominated target is dysregulated in the relevant human cell, causes disease, or is engaged by the intervention.

Functional-genomic data offer independent tests of different parts of this chain. Bulk cardiac transcriptomes can assess whether a candidate is reproducibly altered across cohorts, while donor-aware single-nucleus RNA sequencing (snRNA-seq) can expose cell-type-specific changes obscured by tissue averaging (Litviňuková et al. 2020; Reichart et al. 2022). Cis-expression quantitative trait loci (cis-eQTL) can be used as instruments to test associations between genetically predicted expression and disease, but a credible interpretation requires strong instruments, multiplicity control and colocalization with complete regional summary statistics (Davey Smith and Hemani 2014; Giambartolomei et al. 2014). Static molecular docking contributes only a structural plausibility check unless it is accompanied by appropriate controls and experimental validation, and is retained here as a supplementary check rather than part of the main evidence chain (Trott and Olson 2010).

We applied these principles to the PI3K–AKT candidates nominated for Yangxin Decoction, treating the formula and its network-pharmacology map as the source of a candidate hypothesis rather than as an established mechanism. The analysis was designed to separate three questions that are often conflated: whether a target is nominated by curated disease and compound evidence, whether its expression changes in relevant cardiac cells, and whether genetically predicted expression shares a disease-associated causal variant. Because tissue averaging can hide or blur cell-restricted changes, cell-resolved expression across every annotated cardiac population was central to distinguishing a pathway-wide response from a lineage-restricted node. Negative and inconclusive results were retained as constraints on the therapeutic-hub hypothesis, keeping the strength of each conclusion proportional to its evidence.

Results

Network pharmacology nominates a flavonoid-driven PI3K–AKT candidate set

The archived compound–target dataset contained 1,148 unique genes. Disease genes were stratified before intersection: Tier 1 admitted only direct entries from OMIM, KEGG Disease or CTD DirectEvidence, whereas Tier 2 additionally admitted broad association and inference resources. Tier 1 contained 671 DCM genes and 988 viral-myocarditis genes; 61 genes were shared by both disease lists and the formula-associated target set. The broader Tier-2 intersection contained 1,060 genes (Fig. 1a, b; Online Resource 1).

Within the direct-evidence set, *AKT1* and *TP53* were supported for both diseases. *STAT3*, *SRC* and *JUN* entered only through the broader evidence tier and were not treated as equivalent to direct-evidence candidates (Table 1). KEGG over-representation analysis of the 61-gene set returned 243 pathways, 163 of them at a Benjamini–Hochberg q below 0.05 (Online Resource 2). Lipid and atherosclerosis ranked second (20/59 testable genes, $-\log_{10} P = 14.35$, $q = 5.38 \times 10^{-13}$) and PI3K–AKT signalling sixth (22/59 genes, $-\log_{10} P = 12.54$, $q = 1.08 \times 10^{-11}$), placing both among the leading terms displayed in Fig. 1c. AGE–RAGE signalling in diabetic complications (12/59 genes, $-\log_{10} P = 9.81$, $q = 1.52 \times 10^{-9}$) and TNF signalling (9/59 genes, $-\log_{10} P = 5.89$, $q = 3.86 \times 10^{-6}$) were also significant, but they ranked 25th and 82nd of the 243 terms and fall outside the twelve shown in the figure. Tier-1 membership is itself conferred in part by KEGG Disease records, so testing the same genes against KEGG pathway annotation does not produce independent hypergeometric statistics; these P values order the terms rather than test them. On the compound side, four flavonoids—quercetin, kaempferol, isorhamnetin and naringenin—converged on predicted *AKT1* relationships in the compound–target network (Fig. 2). These nominations rest on a narrow and poorly identified compound base: among the

61 Tier-1 shared genes, quercetin alone accounts for 24 of the predicted compound–target hits, kaempferol for 11, isorhamnetin for 8 and naringenin for 4, so a single promiscuous flavonoid underlies most of them, and only 3 of the 23 core compounds carry a resolvable PubChem identifier. This is the generic, hub-directed target prediction identified as a weakness of network pharmacology (Nogales et al. 2022), and the transcriptomic and genetic layers below were designed to test whether such nominations survive independent evidence rather than to confirm them. These observations nominate a pathway program for testing; none independently establishes target engagement or disease causality.

The candidate set is not reproducibly altered in bulk cardiac tissue

All 61 Tier-1 genes were assessed in four cardiac microarray comparisons: idiopathic DCM in GSE5406, type-2-diabetes-associated heart failure in GSE26887, human inflammatory cardiomyopathy/PVB19 (parvovirus B19) in GSE4172, and mouse CVB3 (coxsackievirus B3) myocarditis in GSE35182. Twenty genes reached a dataset-level Benjamini–Hochberg adjusted P value below 0.05 in at least one comparison. That count is interpretable only against the transcriptome-wide significant fraction of each dataset. In GSE5406, 2,209 of 13,039 tested genes (16.9%) reached $FDR < 0.05$, against 8 of the 60 Tier-1 genes measured on that platform (13.3%; 10.2 expected, hypergeometric $P = 0.820$). The corresponding comparisons were 253 of 19,974 (1.3%) against 1 of 60 (1.7%; 0.8 expected, $P = 0.535$) in GSE26887, 642 of 21,355 (3.0%) against 2 of 61 (3.3%; 1.8 expected, $P = 0.551$) in GSE4172, and 1,612 of 15,073 (10.7%) against 11 of 59 (18.6%; 6.3 expected, $P = 0.046$) in GSE35182. The candidate set is therefore differentially expressed no more often than the transcriptome background in any of the three human datasets, and in the largest human DCM cohort it falls below that background. The one nominal excess, in the mouse CVB3 model, is uncorrected across these four tests. Only *IL6* and *SLC2A1* were significant in two datasets; their repeated evidence arose mainly on the DCM axis, and *SLC2A1* changed in opposite directions across its two significant datasets. Most Tier-1 genes therefore lacked reproducible bulk differential expression (Fig. 3a; Online Resource 3).

The pathway hubs illustrate the same limitation. *AKT1* was not significant in any bulk dataset, with $\log_2$ fold changes of -0.103 in GSE5406, -0.080 in GSE26887, 0.506 in GSE4172 and -0.029 in GSE35182. Among the eight hubs examined at single-nucleus resolution, only *STAT3* reached $FDR < 0.05$ in the main idiopathic-DCM bulk cohort (Fig. 3b). The human and mouse myocarditis-related datasets also differed in species, initiating context and sample size, so they were retained as separate proxies rather than pooled into a single myocarditis effect.

Four PI3K–AKT hub genes differ in DCM cardiomyocytes, and the differences withstand depth adjustment

Healthy-heart snRNA-seq confirmed that the five direct-evidence PI3K–AKT genes were expressed across multiple cardiac cell populations rather than confined to one lineage, and the cardiomyopathy atlas showed the same broad distribution (Fig. 4a, b; Online Resources 4 and 5). Donor-pseudobulk analysis of the Reichart cardiomyopathy atlas then compared 52 DCM donors with 18 controls. Cardiac muscle cell is the atlas annotation for cardiomyocytes, and the two terms are used interchangeably in the text, figures and tables below. In cardiac muscle cells, *AKT1* was upregulated ($\log_2$ fold change 0.351 , adjusted $P = 3.32 \times 10^{-4}$), as were *PIK3CA* (0.407 , 4.32×10^{-8}) and *GSK3B* (0.133 , 0.0126); *STAT3* was downregulated (-0.513 , 0.0230). *TP53*, *MTOR*, *JUN* and *SRC* were not significant in these cells (Fig. 5a; Table 2). Donor-level values shifted between the groups rather than separating them: for *AKT1*, 8 of 52 DCM donors sat below the control median, and for *GSK3B*, 4 of 52 DCM donors sat below the control median while 4 of 18 controls sat above the DCM median (Online Resource 6). The donor-level area under the curve for higher expression in DCM was 0.96 for *PIK3CA*, 0.85 for *AKT1*, 0.82 for *GSK3B* and 0.26 for *STAT3*, the last value below 0.5 because *STAT3* is lower in DCM; only *PIK3CA* approaches separation, and no gene depends on a small number of donors (Fig. 5b). Thus, the negative bulk result for *AKT1* does not generalize to DCM cardiomyocytes.

Because DCM and control donors contribute unequal numbers of cardiomyocyte nuclei, and this imbalance is correlated with group, aggregation depth is a group-correlated technical covariate that the group-only differential-expression model does not adjust for. Control donors contributed a median of 6,344 cardiomyocyte nuclei ($n = 18$; minimum 1,781; interquartile range 5,454–9,872) and DCM donors a median of 2,478 ($n = 52$; minimum 35; interquartile range 884–3,891), a 2.56-fold difference in medians; three DCM donors contributed fewer than 100 nuclei, four fewer than 200 and eight fewer than 500, whereas no control donor fell below any of these thresholds. The four cardiomyocyte signals were nonetheless stable to depth. Raising the per-donor nucleus floor from the full cohort to ≥ 100 , ≥ 200 and ≥ 500 nuclei (retaining 44 DCM and 18 control donors at the strictest floor) left the group contrast essentially unchanged for *AKT1* (two-sided Mann–Whitney $P = 1.38 \times 10^{-5}$, 4.74×10^{-6} , 5.95×10^{-6} and 1.36×10^{-5}), *PIK3CA* (8.51×10^{-9} , 1.35×10^{-8} , 1.59×10^{-8} and 3.23×10^{-8}), *GSK3B* (7.16×10^{-5} , 5.72×10^{-5} , 4.47×10^{-5} and 4.38×10^{-5}) and *STAT3* (2.92×10^{-3} , 2.09×10^{-3} , 2.33×10^{-3} and 1.36×10^{-3}). A donor-level linear model with $\log_{10}(\text{nuclei})$ as a covariate alongside group retained a significant group effect for all four genes (*AKT1* $+0.510$, $P = 1.13 \times 10^{-5}$; *PIK3CA* $+0.477$, $P = 6.18 \times 10^{-7}$; *GSK3B* $+0.296$, $P = 9.18 \times 10^{-5}$; *STAT3* -0.675 , $P = 1.24 \times 10^{-3}$). The depth term itself was not significant for *AKT1* ($P = 0.252$) or *GSK3B* ($P = 0.81$) but was nominally significant for *PIK3CA* ($P = 0.0346$) and *STAT3* ($P = 0.0198$), so a residual depth contribution to these two effects cannot be excluded. Within groups, expression correlated with nucleus count only for *GSK3B* in DCM (Spearman $\rho = +0.312$, $P = 0.0245$); the *AKT1* correlation was not significant in either group (control $\rho = +0.154$, $P = 0.542$; DCM $\rho = +0.069$, $P = 0.624$).

The PI3K–AKT response is pathway-wide but its AKT1 component is lineage-restricted

Repeating the donor-pseudobulk comparison in every annotated cell population showed that the four cardiomyocyte signals do not share one cell-type distribution (Fig. 4c; Online Resource 7). The contrast is not identical across populations, because each population retains only the donors meeting its per-donor cell-count criterion, and the per-population donor counts are listed in the legend to Fig. 4. *AKT1* changed only in cardiomyocytes and in the unannotated population, whereas *PIK3CA* rose in eight of the nine populations and reached its largest effect in cardiac fibroblasts (0.922 , adjusted $P = 1.46 \times$

10^{-15}) rather than in cardiomyocytes, where its effect ranked sixth. Those eight populations include adipocytes, which contribute the fewest donors of any population (10 versus 10), so the breadth of the *PIK3CA* result rests in part on the smallest comparison in the panel. *GSK3B* changed in five populations and *STAT3* in four; *MTOR* and *JUN* changed in none. *SRC* and *TP53*, both null in cardiomyocytes, were lower in DCM in cardiac fibroblasts and mural cells (*SRC* -0.427 , adjusted $P = 3.10 \times 10^{-4}$ and -0.521 , adjusted $P = 1.99 \times 10^{-7}$; *TP53* -0.267 , adjusted $P = 3.38 \times 10^{-3}$ and -0.224 , adjusted $P = 4.62 \times 10^{-2}$), the *SRC* result in mural cells being among the most significant in the whole panel. The description of *TP53*, *MTOR*, *JUN* and *SRC* as not significant therefore holds for cardiomyocytes and not across cell types.

The depth check applied to cardiomyocytes was extended to every annotated population by recomputing donor-level values for all cell types directly from the atlas. Applying the per-donor floor of 30 nuclei to that recomputation reproduced the donor counts of all nine populations exactly, confirming that the recomputed values enter the same comparisons. Nuclei per donor were imbalanced in both directions across populations, most strongly in cardiomyocytes (control-to-DCM median ratio 2.56) and adipocytes (2.06), while endothelial cells (0.67), myeloid cells (0.58) and lymphocytes (0.72) carried more DCM than control nuclei. With $\log_{10}(\text{nuclei})$ carried alongside group, the *AKT1* change remained confined to cardiac muscle cells ($P = 1.13 \times 10^{-5}$) and the unannotated population ($P = 1.26 \times 10^{-4}$), and the nominal lymphocyte signal did not survive (unadjusted $P = 0.0188$, adjusted $P = 0.22$). *PIK3CA* remained higher in DCM in seven of the nine populations, most strongly in cardiac fibroblasts (coefficient $+0.994$, $P = 2.22 \times 10^{-13}$); the myeloid-cell signal did not survive adjustment (unadjusted $P = 0.00625$, adjusted $P = 0.0839$). The contrast between a single-lineage *AKT1* change and a broadly distributed *PIK3CA* change therefore holds when depth is held constant, with the breadth of the *PIK3CA* result reduced by one population.

The cardiomyocyte-restricted description applies to *AKT1* specifically, not to the pathway as a whole: the wider PI3K–AKT response is engaged across cardiac cell types, and only its *AKT1* component is confined to the myocyte lineage among the annotated populations. Cell-restricted expression change of this kind is averaged with other populations in whole tissue, which is consistent with the negative bulk result, but the data do not identify whether any of these responses is adaptive, pathogenic or therapeutically reversible.

Genetic appraisal does not validate AKT1 as a causal target

Seven pathway candidates—*AKT1*, *PIK3CA*, *PIK3R1*, *GSK3B*, *MTOR*, *PTEN* and *STAT3*—were screened for cis-eQTL instruments using complete GTEx v8 regional summary statistics from heart left ventricle ($n = 382$) and whole blood ($n = 670$). Instruments were required to meet $P < 5 \times 10^{-8}$, minor-allele frequency ≥ 0.01 and $F > 10$. No qualifying instrument was found for *AKT1* in either tissue. Six other candidates also lacked an instrument in heart, and only whole-blood *MTOR* yielded an eligible lead variant, rs2295079 ($F = 45.19$; Fig. 6a, b; Online Resource 8). The negative-control genes were screened separately: heart *ITGB3* had an eligible instrument ($F = 69.77$), whereas *PSEN2* did not. Both control genes are themselves members of the 61-gene Tier-1 set, each carrying OMIM direct evidence for DCM and for viral myocarditis together with an archived compound relationship, and both are among the genes plotted in Fig. 3a. They are consequently not independent of the candidate set and do not constitute a genuine negative control; because *PSEN2* yielded no instrument, the control family reduces to the single heart *ITGB3* estimate.

For the single available instrument, whole-blood *MTOR*, a Wald ratio was computable but does not constitute positive evidence. Its nominal myocarditis estimate (OR 0.736, 95% CI 0.577–0.939) survived multiple-testing correction (adjusted $P = 0.0405$ after both Benjamini–Hochberg and Bonferroni correction across the three primary estimates), yet its uncorrected P (0.0135) is numerically identical to the outcome-side single-variant association P , so the Wald ratio adds no information beyond that one SNP; a single instrument permits no assessment of pleiotropy or heterogeneity; and, as shown below, the accompanying colocalization is uninformative. The estimate is therefore not treated as support for a causal effect and is not evidence for cardiac-tissue *MTOR* regulation. The cardiomyopathy and heart-failure estimates were null (OR 1.046, 95% CI 0.921–1.188, $P = 0.485$ and OR 1.020, 95% CI 0.958–1.086, $P = 0.533$; Fig. 6c). The heart *ITGB3* comparison, which is drawn from the same 61-gene set and is therefore not an independent negative control, was null for all three outcomes after correction (cardiomyopathy OR 1.089, 95% CI 0.985–1.203; heart failure OR 1.030, 95% CI 0.980–1.082; myocarditis OR 0.947, 95% CI 0.782–1.148).

Colocalization used complete regional eQTL and outcome summary statistics with priors $p_1 = p_2 = 1 \times 10^{-4}$ and $p_{12} = 1 \times 10^{-5}$. None of the 42 primary gene–tissue–outcome comparisons reached the prespecified threshold for moderate (PP.H4 ≥ 0.5) or strong (PP.H4 ≥ 0.8) support. The largest primary posterior was the whole-blood *MTOR*–myocarditis comparison (PP.H4 = 0.230; 6,378 shared variants), while most values were below 0.05 (Fig. 7a; Online Resource 9). Its complete posterior was PP.H0 ≈ 0.000 , PP.H1 = 0.559, PP.H2 ≈ 0.000 , PP.H3 = 0.210 and PP.H4 = 0.230, so most of the posterior mass falls on the hypothesis that the region carries an eQTL signal and no detectable outcome association. Conditional on a signal being present in both traits, the split is close to even, PP.H4/(PP.H3 + PP.H4) = 0.52. The regional pattern shows why. The whole-blood *MTOR* cis-eQTL forms a sharp peak around rs2295079, whereas no variant in the same interval approaches genome-wide significance for myocarditis, so the two association profiles do not track each other across the locus (Fig. 7b, c). Raising p_{12} to 1×10^{-4} increased the *MTOR*–myocarditis PP.H4 to 0.750, showing material prior sensitivity rather than robust colocalization (Fig. 7d). This comparison is therefore uninformative rather than negative: with no detectable myocarditis association anywhere in the region, the data cannot distinguish a shared causal variant from two distinct ones, and they neither support nor exclude a shared variant behind the expression–outcome association. An *AKT1* inference cannot be drawn from a locus dataset restricted to significant eQTL records because the null regional variants required by colocalization are absent.

Discussion

PI3K–AKT remains a defensible candidate program because direct disease annotations and cell-resolved expression converge at the pathway level. The data do not, however, validate *AKT1* as a causal or therapeutic hub. *AKT1* has no qualifying GTEx cis-eQTL in heart or blood, and complete regional colocalization provides no positive genetic evidence for it. The strongest reproducible observation is instead transcriptional: DCM cardiomyocytes show higher *AKT1*,

PIK3CA and *GSK3B* and lower *STAT3*, despite weak or inconsistent signals in whole-tissue cohorts. Resolving the same comparison across all annotated populations then distinguishes two claims that the cardiomyocyte result alone would merge. *AKT1* is the one hub whose change is restricted to that lineage among the annotated populations; *PIK3CA*, and to a lesser degree *GSK3B* and *STAT3*, shift in several cardiac cell types at once. Describing the whole pathway as cardiomyocyte-specific would therefore overstate what the data show, and the distinction matters for the follow-up experiments, since a broadly engaged pathway and a lineage-restricted node call for different perturbation designs.

The distinction between pathway activity and target causality matters biologically. PI3K–AKT signalling regulates cardiomyocyte growth, survival and metabolic adaptation, and its effects can be protective or maladaptive depending on cell type, duration and pathway branch (Aoyagi and Matsui 2011; McMullen and Jay 2007). Higher transcript abundance in diseased cardiomyocytes therefore does not imply that inhibiting or activating the corresponding protein would be beneficial. It may represent compensation, altered cell state, or a downstream response. The current data locate the response and prioritize experiments; they do not determine the required therapeutic direction.

The contrast between bulk and donor-pseudobulk results shows why cellular resolution is valuable in cardiomyopathy. Bulk cardiac tissue combines cardiomyocytes, fibroblasts, vascular cells and immune populations whose relative abundance and gene regulation both change with disease. A cardiomyocyte-specific signal can consequently be diluted or cancelled even when it is statistically clear within that lineage. The donor-level design of the snRNA-seq comparison avoids treating nuclei as independent biological replicates, a common source of inflated significance in single-cell studies. Because DCM donors also contributed fewer cardiomyocyte nuclei than controls, we confirmed that the four cardiomyocyte signals persist under stricter per-donor nucleus floors and after adjusting for aggregation depth, although the depth covariate itself reached nominal significance for *PIK3CA* and *STAT3*. Repeating the check across all annotated populations left the lineage contrast intact, so it does not rest on the depth imbalance. Four qualifications remain. The breadth of the *PIK3CA* result is one population narrower after adjustment than the primary analysis reports; the adipocyte comparison rests on the fewest donors and among the fewest nuclei of any population, making it the least secure of the seven populations that retain the *PIK3CA* signal after adjustment; only eight prespecified hubs had disease pseudobulk results in the available archive; and the DCM dataset provides no direct evidence for human viral myocarditis.

The genetic analysis supplies an equally important negative constraint. Restricting an eQTL dataset to significant records removes null regional variants and is inappropriate for standard colocalization because the posterior requires the relative association pattern across the locus. Reanalysis with complete GTEx regions found no instrument for *AKT1* and no supported shared signal for any primary comparison. Whole-blood *MTOR* produced a corrected association with myocarditis, but the accompanying colocalization is uninformative rather than negative. Most of its posterior mass sits on PP.H1 = 0.559, the hypothesis that only the expression trait is associated across the region, and the remainder divides almost evenly between PP.H3 and PP.H4. The comparison therefore constrains a shared causal variant in neither direction, and together with the dependence on a single instrument this prevents a causal claim. The shift to PP.H4 = 0.750 only under the permissive p12 sensitivity setting shows how weakly the data determine the posterior. A larger tissue-matched eQTL study or molecular QTL measured in disease-relevant cardiac cells would be needed to revisit the locus.

The bulk datasets also impose clear scope limits. GSE5406 is a sizeable human idiopathic-DCM cohort, whereas GSE26887 is a small type-2-diabetes heart-failure cohort and should not be treated as an independent idiopathic-DCM replication. GSE4172 represents human inflammatory cardiomyopathy/PVB19, while GSE35182 is a mouse CVB3 model; neither is a complete representation of clinical viral myocarditis. We preserved dataset-level estimates and did not combine these heterogeneous contrasts into a pooled disease effect. Their value lies in testing whether a candidate survives distinct contexts, not in estimating one common fold change.

Several additional limitations define what the study can support. Compound–target associations and disease annotations are database-derived and inherit database coverage, update and curation biases. The archived source did not retain exact access dates for every network-pharmacology database. The receptor and ligand PDBQT files permit a computationally reproducible Vina rerun, but their preparation history is incomplete; docking lacks decoy benchmarking and dynamic or experimental validation. The snRNA-seq evidence is transcript-level and limited to DCM. The genetic outcomes are broad FinnGen cardiomyopathy, myocarditis and heart-failure phenotypes rather than DCM- or virus-specific endpoints. All analyses are computational and use public data; no experiment demonstrates formula exposure, protein modulation or functional rescue.

These boundaries point to a focused experimental program. Protein abundance and phosphorylation of AKT1, PIK3CA, GSK3B and STAT3 should be measured in independently collected DCM cardiomyocytes or cardiac tissue. Formula or compound perturbation should then test pathway direction and phenotype rescue with dose–response designs, orthogonal target-engagement assays and target-specific genetic perturbation. Human myocarditis single-cell data would determine whether the DCM cell-state pattern generalizes to an inflammatory myocardial context. Until those tests are available, Yangxin Decoction should be described as the source of a PI3K–AKT candidate hypothesis rather than as a validated modulator of that mechanism.

Conclusions

Cell-resolved transcriptomics shows that the PI3K–AKT response in dilated cardiomyopathy is engaged across cardiac cell types while only its *AKT1* component is confined to cardiomyocytes among the annotated populations, a distinction that neither bulk transcriptomes nor pathway enrichment can make because both average across cells. Curated target evidence, complete-region eQTL and GWAS analysis, and a supplementary docking check add no causal support: there is no qualifying *AKT1* instrument and no positive colocalization, so the analysis supports neither an *AKT1* causal claim nor a demonstrated therapeutic mechanism for Yangxin Decoction. This evidence-calibrated conclusion is narrower than network pharmacology alone but provides a clearer experimental target—the direction and functional consequence of PI3K–AKT dysregulation, tested separately in cardiomyocytes and in the non-myocyte populations that carry the broader signal, with direct validation in myocarditis still required.

Methods

Study design and evidence hierarchy

The study used archived network-pharmacology outputs to nominate candidates and tested them against independent transcriptomic and genetic datasets. Evidence was interpreted in ordered categories: direct disease curation, expression in relevant tissue or cell types, and genetic association with colocalization, with structural plausibility from docking retained only as a supplementary check. A positive result in one category did not substitute for a negative or unavailable result in another. All analyses used public summary statistics or de-identified public transcriptomic data; no new human or animal material was collected.

Compound and disease target nomination

Yangxin Decoction is a classical calming-the-spirit formula recorded in volume 11 of the twenty-six-volume *Renzhai Zhizhi Fanglun* (Renzhai's Direct Guidance on Formulas), whose author's preface is dated 1264, where it is prescribed for deficiency of blood with fright and unquiet of the heart (Yang 1264). It comprises thirteen medicinals, given here with their standard Latin drug names: Astragali Radix Praeparata cum Melle (Zhi Huangqi), Poria (Baifuling), Poria cum Radix Pini (Fushen), Banxiaqu (fermented Pinelliae Rhizoma, a preparation of Pinelliae Rhizoma powder fermented with wheat flour), Angelicae Sinensis Radix (Danggui), Chuanxiong Rhizoma (Chuanxiong), Polygalae Radix (Yuanzhi), Cinnamomi Cortex (Rougui), Platycladi Semen (Baiziren), Ziziphi Spinosa Semen (Suanzaoren), Schisandrae Chinensis Fructus (Beiwuweizi), Ginseng Radix et Rhizoma (Renshen) and Glycyrrhizae Radix et Rhizoma Praeparata cum Melle (Zhi Gancao). The source text specifies Huangqi and Gancao only as roasted; the honey-roasted forms named here follow the modern reading used throughout this analysis. Figure 2 shows the curated core of that network rather than the complete compound–target map: it draws six of the thirteen medicinals and the twenty-six herb–compound edges retained in the core set, while the screened compound set as a whole spans all thirteen. Some compounds drawn in the figure therefore derive from medicinals that are not themselves drawn as nodes.

The study is computational throughout and drew on archived compound–target records rather than on a physically prepared decoction; no plant material was collected or extracted, so voucher-specimen and batch information does not apply. The source text also records the proportions, which are reproduced here in the units in which they are given rather than converted: Zhi Huangqi, Baifuling, Fushen, Banxiaqu, Danggui and Chuanxiong at half a liang each; Yuanzhi, Rougui, Baiziren, Suanzaoren, Beiwuweizi and Renshen at one fen each; and Zhi Gancao at four qian. The formula is prepared as a coarse powder and taken at three qian per dose, decocted with five slices of fresh ginger and two jujubes before meals (Yang 1264). These proportions describe the classical preparation and are given for completeness; the present analysis used archived compound–target records for the constituent medicinals and did not prepare or dose a decoction, so the proportions were not applied experimentally.

Candidate compounds and protein targets for Yangxin Decoction were obtained from TCMSP and BATMAN-TCM. Compounds had been filtered using oral bioavailability $\geq 30\%$, drug-likeness (DL) ≥ 0.18 and Lipinski criteria. The fixed archived compound–target dataset contained 1,148 unique genes. The exact database access dates and the software implementation used to compute the absorption, distribution, metabolism and excretion values were not retained; the network was therefore treated as an archived nomination set rather than a current database snapshot.

Coverage of the thirteen medicinals in TCMSP was uneven, and the archived record documents how each gap was handled. The honey-roasted preparations Zhi Huangqi and Zhi Gancao are not indexed separately from the unprocessed medicinals, so the records for the unprocessed forms were used. Fushen and Yuanzhi returned no entry at all. For Baifuling, Danggui and Baiziren no compound met the $\text{AlogP} \leq 5$ criterion, a calculated-lipophilicity threshold that the archived record lists as a filter in its own right rather than as one of the Lipinski criteria, and for Rougui none met the drug-likeness threshold $\text{DL} \geq 0.18$. Compounds for Yuanzhi were therefore assembled from TCM Database@Taiwan, TCM-ID, HIT 2.0, ETCM, TCMIP, HERB and SymMap and screened with SwissADME instead, retaining compounds with high predicted gastrointestinal absorption, no more than ten rotatable bonds and at least two positive drug-likeness verdicts including Lipinski. Compounds that could not be retrieved from the primary sources were supplemented from NPASS, ZINC, NPACT and ChEMBL. Two screening standards were thus applied: the TCMSP oral-bioavailability and drug-likeness thresholds for most medicinals, and the SwissADME criteria for the Yuanzhi set. This heterogeneity is a property of the archived nomination set and is reported rather than corrected, because the underlying database queries were not re-run for this study.

DCM and viral-myocarditis genes were divided into two evidence tiers. Tier 1 admitted OMIM, KEGG Disease and CTD records with explicit DirectEvidence only. Tier 2 was the union of Tier 1 with GeneCards, NCBI, GeneMap, Expasy, Harmonizome and full CTD inference. Gene symbols were standardized to uppercase and deduplicated. The three-way intersections between the formula-associated set and the two disease sets were calculated separately for each tier.

Enrichment and network visualization

KEGG over-representation analysis was computed on the 61 Tier-1 genes. Gene symbols were mapped to Entrez identifiers with org.Hs.eg.db and tested with clusterProfiler (Wu et al. 2021). The background was set explicitly to the 5,868 human genes carrying at least one KEGG pathway annotation rather than left at its default, because a background that includes unannotated genes inflates significance for every term. P values were adjusted across all returned terms by the Benjamini–Hochberg procedure and are reported as q values; the term ranks quoted in the text are by unadjusted P. The complete result table for all 243 returned terms, giving the gene counts, the member genes, the unadjusted P and the Benjamini–Hochberg q of each term, accompanies the submission as Online Resource 2, so that terms outside the twelve drawn in Fig. 1c can be looked up directly. Tier 1 admits genes on the strength of OMIM, KEGG Disease and CTD DirectEvidence records, and those genes are then tested against KEGG pathway annotation. The resulting P values are consequently not independent hypergeometric statistics: they are used to order the pathway terms, not to test them, and no over-representation claim rests on them.

Archived Metascape output supplied the pathway-node display names used in the compound–target network (Zhou et al. 2019). The displayed core network retained higher-degree compounds, the four flavonoids relevant to *AKT1*, Tier-1 targets connected to at least two displayed compounds, and leading KEGG pathways containing at least three displayed targets. The network is a visualization of archived associations and was not used as an inferential statistical model.

Bulk cardiac transcriptomics

Four GEO microarray datasets were analyzed separately. GSE5406 contributed idiopathic DCM/heart-failure samples (86 cases, 16 non-failing controls); ischemic samples were excluded. GSE26887 contributed type-2-diabetes-associated heart failure (7 cases, 5 controls) and was treated as exploratory. GSE4172 contributed human inflammatory cardiomyopathy/PVB19 (8 cases, 4 controls). GSE35182 contributed mouse CVB3-infected and PBS-treated hearts (12 per group). Expression was log2-transformed when required, quantile-normalized, mapped to gene symbols with platform annotation, and collapsed to the probe with the highest mean expression. Differential expression used limma with empirical-Bayes moderation, trend and robust options (Ritchie et al. 2015). GSE35182 models included sex and infection time point; mouse symbols were mapped to human orthologues. Benjamini–Hochberg adjusted $P < 0.05$ defined dataset-level statistical evidence in the 61-gene summary; effect sizes were retained without an additional threshold. The correction was applied across the full transcriptome of each dataset (13,039 genes in GSE5406, 19,974 in GSE26887, 21,355 in GSE4172 and 15,073 in GSE35182) and not within the 61-gene subset, so the adjusted P values quoted for Tier-1 genes are transcriptome-wide values. The statement that a gene was significant in at least one dataset carries no multiplicity control across the four datasets. To put the 61-gene counts on an interpretable scale, the proportion of Tier-1 genes reaching adjusted $P < 0.05$ was compared with the transcriptome-wide proportion in the same dataset using a one-sided hypergeometric test, with the Tier-1 genes measured on that platform as the denominator.

Cardiac single-nucleus RNA sequencing

Healthy cell-type localization used the Litviňuková adult-human-heart atlas (486,134 nuclei from 14 donors) (Litviňuková et al. 2020). Disease differential expression used the Reichart cardiomyopathy atlas and was restricted to 52 DCM and 18 normal donors (Reichart et al. 2022). Raw integer counts were aggregated by donor within each cell type. Genes expressed in more than 1% of cells were retained, and cell-type datasets required at least three donors per group and 30 cells per donor. Differential expression was modeled with PyDESeq2 0.5.4 using group as the design factor (Muzellec et al. 2023). Adjusted P values were calculated across all genes tested within each cell type. Disease pseudobulk results were available for eight prespecified hub genes; no unreported gene-level result was imputed. Two gene–cell-type combinations, *SRG* in endothelial cells and *SRG* in lymphocytes, did not meet the retention criteria and are shown as untested rather than as a null result. A third combination, *SRG* in adipocytes, was tested but removed by the DESeq2 independent-filtering step, so it carries a fold-change estimate and no adjusted P value; it is likewise not a null result. Donor numbers differ between cell types because each type retains only the donors meeting its per-donor cell-count criterion, so the DCM-versus-control contrast ranges from 52 against 18 donors down to 10 against 10 in adipocytes. Healthy and disease normalized expression values were not compared directly across studies; the two atlases also differ in annotation granularity, so their cell-type panels are read separately.

Donor-level expression values used for display were recomputed from the Reichart atlas by summing raw integer counts per donor within the cardiac-muscle-cell population and dividing by that donor's cardiomyocyte library size, giving counts per million, which were placed on a log2 scale after addition of a pseudocount of one, $\log_2(\text{CPM} + 1)$. This step regenerates the per-donor distribution behind the published comparison; the differential statistics remain the PyDESeq2 estimates and were not recalculated.

Depth sensitivity of the cardiomyocyte result

Because DCM and control donors contribute unequal numbers of cardiomyocyte nuclei and the imbalance is correlated with group, aggregation depth is a group-correlated technical covariate not captured by the group-only differential-expression model. Three checks on the released donor-level values, which are the same values shown in the donor-level figure panel, tested whether the group contrast depends on depth. The nuclei-per-donor distribution was summarized by group; the contrast was recomputed with a two-sided Mann–Whitney test after excluding donors below successively stricter nucleus floors (≥ 100 , ≥ 200 and ≥ 500 nuclei), so that a result driven by the sparsest donors would surface as an unstable P value; and a donor-level ordinary-least-squares model, $\log_2(\text{CPM} + 1) \sim \text{group} + \log_{10}(\text{nuclei})$, estimated the group effect with depth held constant, complemented by the within-group Spearman correlation between expression and nucleus count. These rank and linear-model statistics on donor-level values do not refit the count model and therefore do not re-estimate dispersion. The check was then repeated for every annotated population by recomputing donor-level counts for all cell types from the atlas itself, whose SHA-256 digest was verified against the value recorded in the provenance file before use. Mast cells were not carried into this comparison because too few donors clear the nuclei floor for a model to be fitted, which is also why they are absent from the nine populations of the primary analysis. The analysis is implemented in the accompanying scripts and its outputs are included in the analysis archive.

Cis-eQTL instrument screen and Wald-ratio analysis

Complete GTEx v8 summary statistics processed by the eQTL Catalogue were retrieved for heart left ventricle (QTL Catalogue dataset QTD000256; $n = 382$) and whole blood (QTD000356; $n = 670$) (GTEx Consortium 2020; Kerimov et al. 2021). Seven pathway genes and two prespecified controls were examined within gene boundaries extended by 1 Mb on each side, using GRCh38 coordinates. The two control genes, *ITGB3* and *PSEN2*, are themselves members of the 61-gene Tier-1 direct-evidence set, each carrying OMIM records for both diseases and an archived compound relationship. They are therefore not independent of the candidate set and do not provide a genuine negative control, and because no instrument was available for *PSEN2* the control family consists of a single heart *ITGB3* estimate. The analysis is reported as performed and this limitation is stated rather than corrected. The strongest cis-eQTL was selected before looking at outcome results if it met $P < 5 \times 10^{-8}$, minor-allele frequency ≥ 0.01 and $F = (\text{beta}/\text{standard error})^2 > 10$. A single strongest variant was used so that no LD-dependent multi-instrument estimate was implied.

FinnGen R12 regional summary statistics were retrieved for cardiomyopathy (I9_CARDMYO; 7,649 cases and 375,343 controls), myocarditis (I9_MYOCARD; 2,062 cases and 252,810 controls) and heart failure (I9_HEARTFAIL; 37,653 cases and 462,695 controls) (Kurki et al. 2023). Exposure and outcome variants were joined by chromosome, position, reference and alternate allele on GRCh38. The alternate allele was the effect allele in both resources. Wald-ratio log odds were calculated as $\beta_{\text{outcome}}/\beta_{\text{exposure}}$, with standard error equal to outcome standard error divided by the absolute exposure β . Odds ratios and 95% confidence intervals were exponentiated from the log-odds scale. Benjamini–Hochberg and Bonferroni corrections were applied separately to the primary and control families across estimates that could be performed.

Colocalization

All valid regional variants were used, not only significant eQTL. Duplicate records at the same chromosome, position and allele combination were collapsed after ordering by P value. Quantitative eQTL datasets used β , variance of β , minor-allele frequency and tissue sample size. FinnGen datasets used log-odds β , variance, alternate-allele frequency, total sample size and case fraction. Exact GRCh38 allele matches were retained. Colocalization was performed with coloc 5.2.3 and coloc.abf (Giambartolomei et al. 2014). Prespecified priors were $p_1 = p_2 = 1 \times 10^{-4}$ and $p_{12} = 1 \times 10^{-5}$; p_{12} values of 1×10^{-6} and 1×10^{-4} were sensitivity analyses. $PP_4 \geq 0.8$ was considered strong and 0.5–0.8 moderate support for a shared variant. Loci with fewer than 100 shared variants were not analyzed.

Repeated molecular docking (supplementary analysis)

Repeated docking was run as a reproducibility check only and was excluded from the main evidence chain: three of its five receptors (STAT3, SRC and JUN) are Tier-2 candidates and none of the five (AKT1, JUN, SRC, STAT3, TP53) is one of the signal-carrying hubs *PIK3CA*, *GSK3B* or *MTOR*; the mean scores of the five ligands against AKT1 span only 0.51 kcal/mol, within the scoring function's error; and no matched decoy set and no biochemical target-engagement assay were available, so no threshold was used to classify binding. The archived receptor and ligand PDBQT files were rerun with AutoDock Vina 1.2.7 under fixed cubic search boxes, exhaustiveness 32 and three prespecified random seeds, with input files recorded by SHA-256 checksum; the full procedure and the resulting scores are given in Online Resources 10 and 11 (Trott and Olson 2010).

Reproducibility and quality control

Regional data files were checked by SHA-256 digest. Analysis scripts stopped on missing columns, incomplete pair–seed combinations or absent outputs. The genetic analysis used R 4.5.2, data.table 1.18.4 and coloc 5.2.3. All figures were produced as single-page vector graphics and are supplied as EPS. Source scripts, result tables, run logs and input checksums are included in the accompanying analysis archive.

Declarations

Author contributions

All authors contributed to the conception and design of the study, the analysis and interpretation of the data, and the drafting and critical revision of the manuscript, and approved the final version for submission. Yingying Qian and Haitao Tian contributed equally to this work.

Acknowledgements

The authors thank the FinnGen study, the eQTL Catalogue and the GTEx Project, and the data contributors to the Gene Expression Omnibus, the CELLxGENE Census and the Protein Data Bank, for making the data used in this study publicly available.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare that they have no conflict of interest.

Ethics approval

This study is a computational secondary analysis of publicly available, de-identified datasets and summary statistics. No participants were recruited and no new human or animal material was collected. The work therefore qualified for exemption from ethics committee review under the institutional policy covering research that uses only already-published, publicly accessible de-identified data, and no separate ethics approval was sought or required for the present analysis. Each primary study from which the analyzed data originate was approved by its own institutional ethics committee and was conducted in accordance with the 1964 Declaration of Helsinki and its later amendments; the corresponding approval details are reported in those publications.

Consent to participate

Not applicable. The study did not recruit participants and used only publicly available, de-identified data and summary statistics, so no consent to participate was sought by the authors.

Consent to publish

Not applicable. The manuscript contains no individually identifiable data, images or other personal details for which consent to publish would be required.

Clinical trial number

Clinical trial number: not applicable. This study is a computational secondary analysis of previously published, publicly available datasets and summary statistics; it is not a clinical trial, no participants were prospectively enrolled or assigned to an intervention, and no trial registration was therefore required or obtained.

Compliance with national legislation

The analyses reported in this study were performed in China and comply with the current laws of that country. The study is entirely computational and used only publicly available data; no laboratory experiments involving human participants, animals or regulated biological material were carried out.

Data availability

Bulk expression data are available from GEO under GSE5406, GSE26887, GSE4172 and GSE35182. Cardiac snRNA-seq data are available through CELLxGENE for the Litviňuková adult-heart and Reichart cardiomyopathy studies. GTEx v8 regional eQTL statistics are available through the eQTL Catalogue (study QTS000015; datasets QTD000256 and QTD000356). Outcome summary statistics are available from FinnGen release 12 under I9_CARDMYO, I9_MYOCARD and I9_HEARTFAIL. Protein structures are available from the Protein Data Bank under 3CQW, 1A02, 1Y57, 6NJS and 8WD2. The processed result tables and the supplementary docking figure are provided as Online Resources 1 to 11, and the analysis scripts as Online Resource 12.

Code availability

The scripts used for regional data retrieval, genetic analysis, expression integration, repeated docking and figure generation are provided as Online Resource 12. They include fixed parameters, software-version logs and checksums for the locally archived inputs.

References

1. Aoyagi T, Matsui T (2011) Phosphoinositide-3 kinase signaling in cardiac hypertrophy and heart failure. *Curr Pharm Des* 17:1818–1824. <https://doi.org/10.2174/138161211796390976>
2. Davey Smith G, Hemani G (2014) Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 23:R89–R98. <https://doi.org/10.1093/hmg/ddu328>
3. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C et al (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 10:e1004383. <https://doi.org/10.1371/journal.pgen.1004383>
4. GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369:1318–1330. <https://doi.org/10.1126/science.aaz1776>
5. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 4:682–690. <https://doi.org/10.1038/nchembio.118>
6. Imanaka-Yoshida K (2020) Inflammation in myocardial disease: from myocarditis to dilated cardiomyopathy. *Pathol Int* 70:1–11. <https://doi.org/10.1111/pin.12868>
7. Japp AG, Gulati A, Cook SA, Cowie MR, Prasad SK (2016) The diagnosis and evaluation of dilated cardiomyopathy. *J Am Coll Cardiol* 67:2996–3010. <https://doi.org/10.1016/j.jacc.2016.03.590>
8. Kawai C (1999) From myocarditis to cardiomyopathy: mechanisms of inflammation and cell death. *Circulation* 99:1091–1100. <https://doi.org/10.1161/01.CIR.99.8.1091>
9. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L et al (2021) A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet* 53:1290–1299. <https://doi.org/10.1038/s41588-021-00924-w>
10. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM et al (2023) FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 613:508–518. <https://doi.org/10.1038/s41586-022-05473-8>
11. Li S, Zhang B (2013) Traditional Chinese medicine network pharmacology: theory, methodology and application. *Chin J Nat Med* 11:110–120. [https://doi.org/10.1016/S1875-5364\(13\)60037-0](https://doi.org/10.1016/S1875-5364(13)60037-0)
12. Litviňuková M, Talavera-López C, Maatz H, Reichart D, Worth CL, Lindberg EL et al (2020) Cells of the adult human heart. *Nature* 588:466–472. <https://doi.org/10.1038/s41586-020-2797-4>

13. McMullen JR, Jay PY (2007) PI3K(p110alpha) inhibitors as anti-cancer agents: minding the heart. *Cell Cycle* 6:910–913. <https://doi.org/10.4161/cc.6.8.4124>
14. Muzellec B, Teleńczuk M, Cabeli V, Andreux M (2023) PyDESeq2: a Python package for bulk RNA-seq differential expression analysis. *Bioinformatics* 39:btad547. <https://doi.org/10.1093/bioinformatics/btad547>
15. Nogales C, Mamdouh ZM, List M, Kiel C, Casas AI, Schmidt HHHW (2022) Network pharmacology: curing causal mechanisms instead of treating symptoms. *Trends Pharmacol Sci* 43:136–150. <https://doi.org/10.1016/j.tips.2021.11.004>
16. Reichart D, Lindberg EL, Maatz H, Miranda AMA, Viveiros A, Shvetsov N et al (2022) Pathogenic variants damage cell composition and single-cell transcription in cardiomyopathies. *Science* 377:eabo1984. <https://doi.org/10.1126/science.abo1984>
17. Reichart D, Magnussen C, Zeller T, Blankenberg S (2019) Dilated cardiomyopathy: from epidemiologic to genetic phenotypes. *J Intern Med* 286:362–372. <https://doi.org/10.1111/joim.12944>
18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47. <https://doi.org/10.1093/nar/gkv007>
19. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem* 31:455–461. <https://doi.org/10.1002/jcc.21334>
20. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z et al (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2:100141. <https://doi.org/10.1016/j.xinn.2021.100141>
21. Yang SY (1264) Renzhai zhizhi fanlung [Renzhai's direct guidance on formulas], vol 11. Siku Quanshu recension, Chinese Wikisource digital edition. [https://zh.wikisource.org/wiki/%E4%BB%81%E9%BD%8B%E7%9B%B4%E6%8C%87_\(%E5%9B%9B%E5%BA%AB%E5%85%A8%E6%9B%B8%E6%9C%AC\)/%E5%9B%9B%E5%BA%AB%E5%85%A8%E6%9B%B8%E6%9C%AC](https://zh.wikisource.org/wiki/%E4%BB%81%E9%BD%8B%E7%9B%B4%E6%8C%87_(%E5%9B%9B%E5%BA%AB%E5%85%A8%E6%9B%B8%E6%9C%AC)/%E5%9B%9B%E5%BA%AB%E5%85%A8%E6%9B%B8%E6%9C%AC) Accessed 2 August 2026
22. Zhai X, Wang X, Wang L, Xiu L, Wang W, Pang X (2020) Treating different diseases with the same method—a traditional Chinese medicine concept analyzed for its biological basis. *Front Pharmacol* 11:946. <https://doi.org/10.3389/fphar.2020.00946>
23. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O et al (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10:1523. <https://doi.org/10.1038/s41467-019-09234-6>

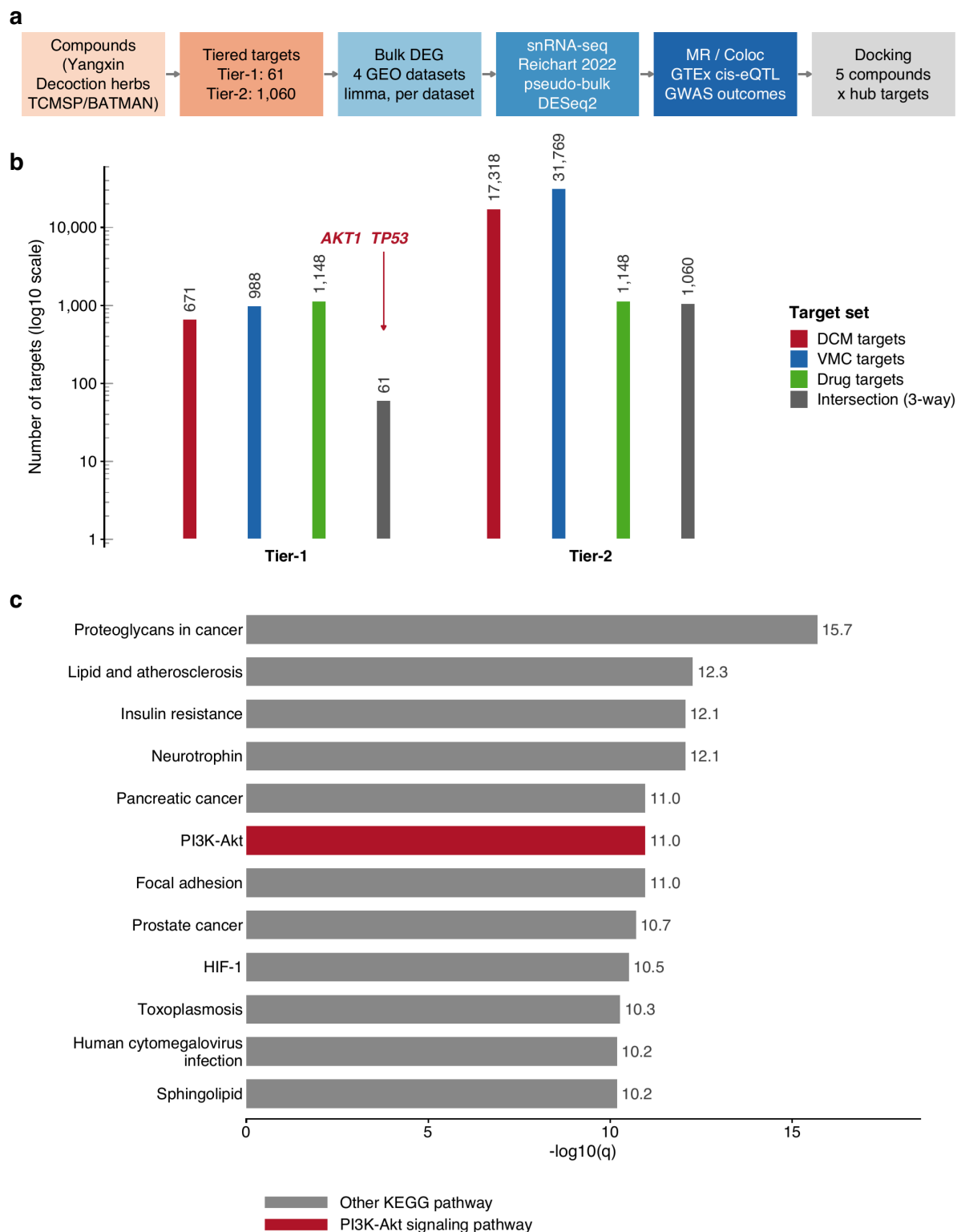


Figure 1

Candidate nomination from archived network-pharmacology evidence

(a) Analytical workflow separating candidate nomination, transcriptomic resolution and genetic appraisal. (b) Sizes of the DCM, viral-myocarditis, formula-associated and three-way intersection sets under the direct-evidence Tier 1 and broad Tier 2 definitions. (c) The twelve most significant KEGG terms for the 61 Tier-1 shared targets, ordered by Benjamini-Hochberg q ; bar length and the value printed beside each bar give $-\log_{10} q$. The red bar marks PI3K-AKT signalling, the one term followed up mechanistically in the later analyses; every other term is grey. Pathway names carry the same shortening rule as Fig. 2: a trailing “signaling pathway” is dropped for space, so Neurotrophin, PI3K-Akt, HIF-1 and Sphingolipid are drawn without it, and no other name is altered.

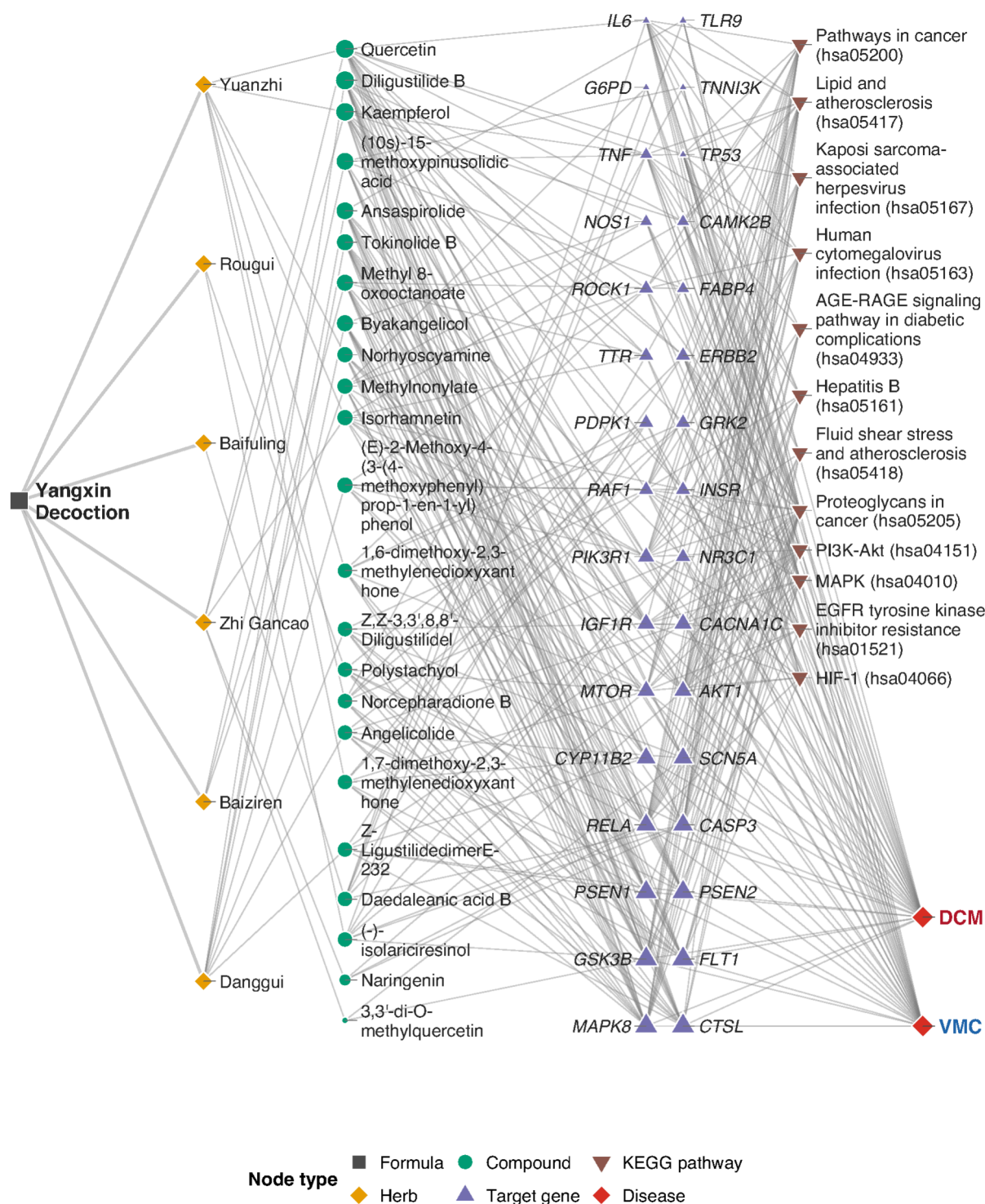


Figure 2

Curated core compound–target–pathway–disease network

Archived relationships linking the Yangxin Decoction formula, its constituent herbs, the retained compounds, Tier-1 target genes, leading KEGG pathways and the two diseases. Node and edge density represents archived database relationships, not measured biochemical interaction. Pathway nodes give the KEGG term name with its hsa accession; for three terms whose names end in “signaling pathway” that suffix is dropped for space, so PI3K-Akt (hsa04151), MAPK (hsa04010) and HIF-1 (hsa04066) are shown without it. Compound names are drawn from the curated compound table and shown in sentence case, with systematic locant and stereochemical prefixes left exactly as recorded. One entry carried only the abbreviated designation “quercetin der.”; it is identified in the source table by PubChem CID 5316900 and is shown here under the corresponding chemical name, 3,3'-di-O-methylquercetin.

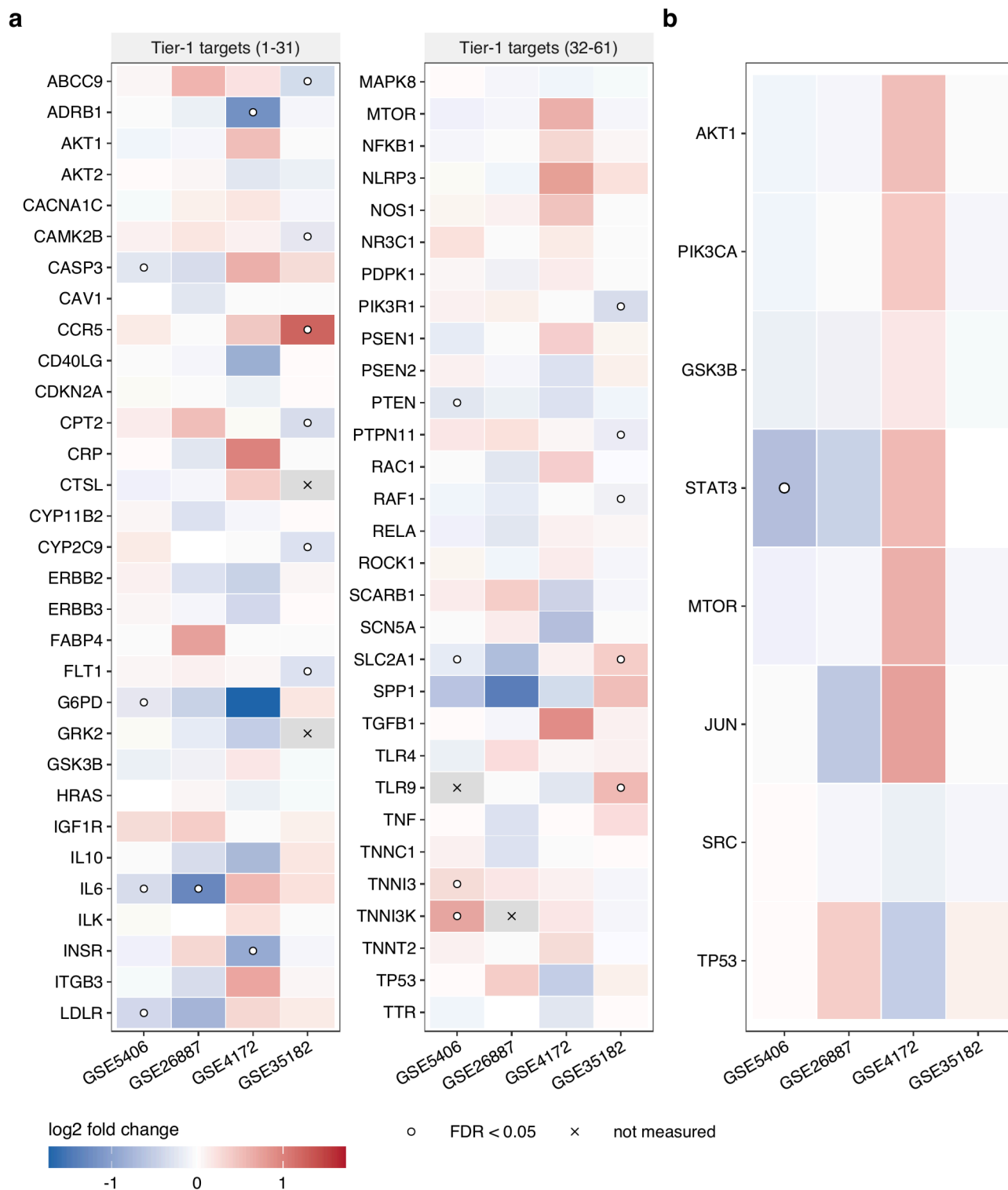


Figure 3

Bulk cardiac transcriptomes rarely show reproducible change in the candidate set

(a) Log2 fold changes for all 61 direct-evidence shared targets in four separately analyzed cardiac microarray datasets, split into two column blocks. GSE5406 is idiopathic DCM; GSE26887 is type-2-diabetes-associated heart failure; GSE4172 is human inflammatory cardiomyopathy/PVB19; GSE35182 is mouse CVB3 myocarditis. (b) The same four datasets for the eight pathway hub genes. Both panels use one shared colour scale, so the muted appearance of panel (b) reflects genuinely smaller hub effects rather than a difference in scaling. Open circles mark dataset-level Benjamini–Hochberg adjusted $P < 0.05$; crosses mark four gene–dataset combinations for which the gene was not measured on that platform. The four datasets are not pooled because their disease definitions and species differ. Panel (b) is drawn from the full per-dataset differential-expression tables because *PIK3CA*, *STAT3*, *JUN* and *SRC* are Tier-2 candidates and fall outside the 61-gene Tier-1 set shown in panel (a).

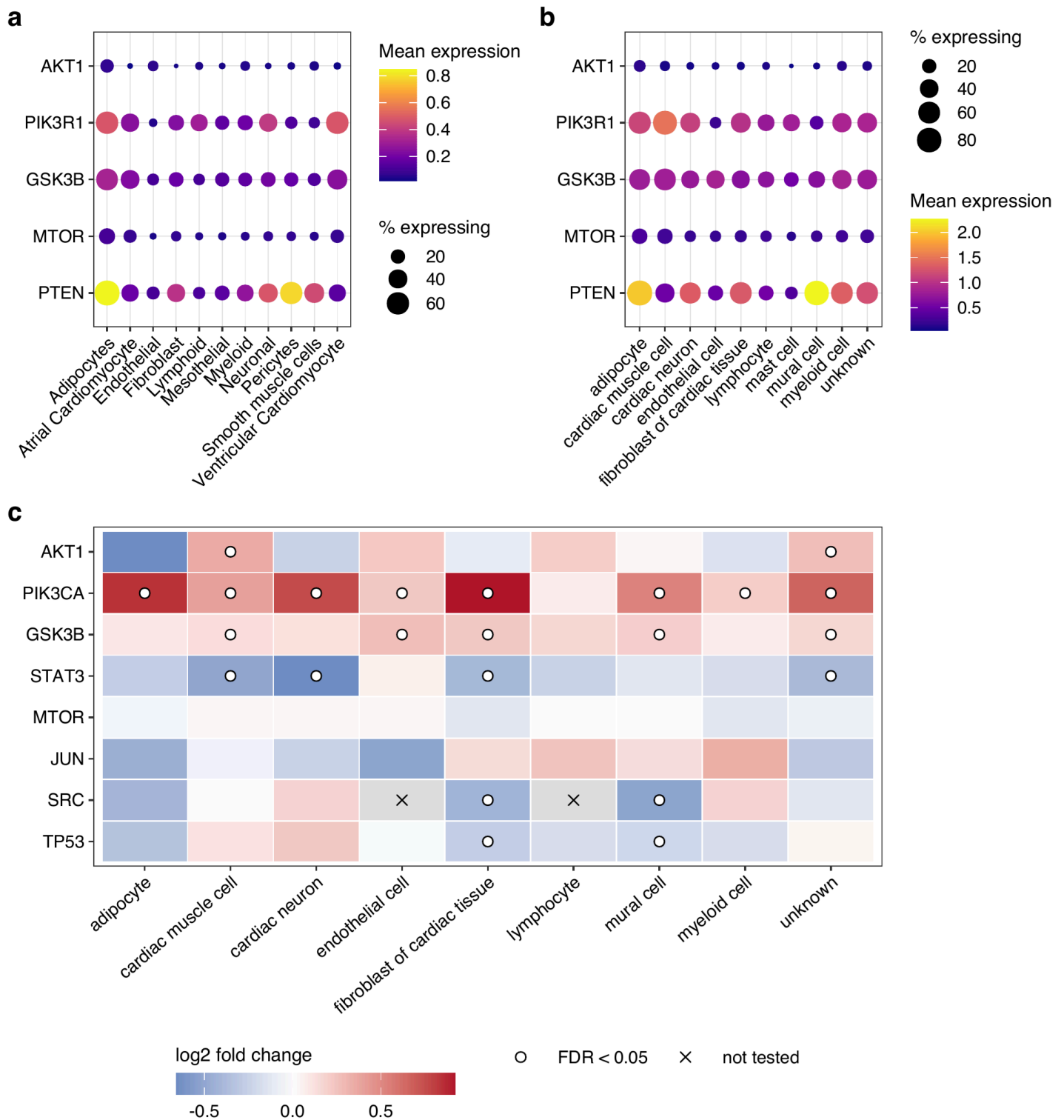


Figure 4

Cell-resolved data localize the AKT1 change, but not the whole pathway, to cardiomyocytes among the annotated populations

(a) Healthy adult-heart expression of five Tier-1 PI3K–AKT genes in the Litviňuková atlas. (b) The same genes in the Reichart cardiomyopathy atlas. In both panels dot size indicates the percentage of expressing nuclei and colour indicates mean normalized expression within each cell type. The two atlases use different annotation granularity and different normalization, so they are read as separate descriptions of where the genes are expressed and are not compared quantitatively; their colour scales differ accordingly. (c) Donor-pseudobulk log2 fold changes for the eight hub genes across the nine donor-level populations, eight annotated cell types and one unannotated group of the Reichart atlas. Panel (b) displays ten cell types and panel (c) nine because mast cells were not carried into the donor-level comparison: too few donors clear the per-donor nucleus floor for a model to be fitted in that population. The cell-type label unknown is the group described in the text as the unannotated population. Donor numbers differ between populations because each population

retains only the donors meeting its per-donor cell-count criterion: 52 DCM against 18 control donors in cardiac muscle cells, endothelial cells, cardiac fibroblasts, mural cells and the unannotated population, 51 against 18 in myeloid cells, 46 against 16 in lymphocytes, 40 against 18 in cardiac neurons, and 10 against 10 in adipocytes. Open circles mark adjusted $P < 0.05$; crosses mark the two gene–cell-type combinations, *SRC* in endothelial cells and *SRC* in lymphocytes, that did not meet the retention criteria and were therefore not tested. One further combination, *SRC* in adipocytes, was tested but removed by the DESeq2 independent-filtering step, so it has a fold-change estimate and no adjusted P value; it is drawn as an ordinary cell and should not be read as a null result.

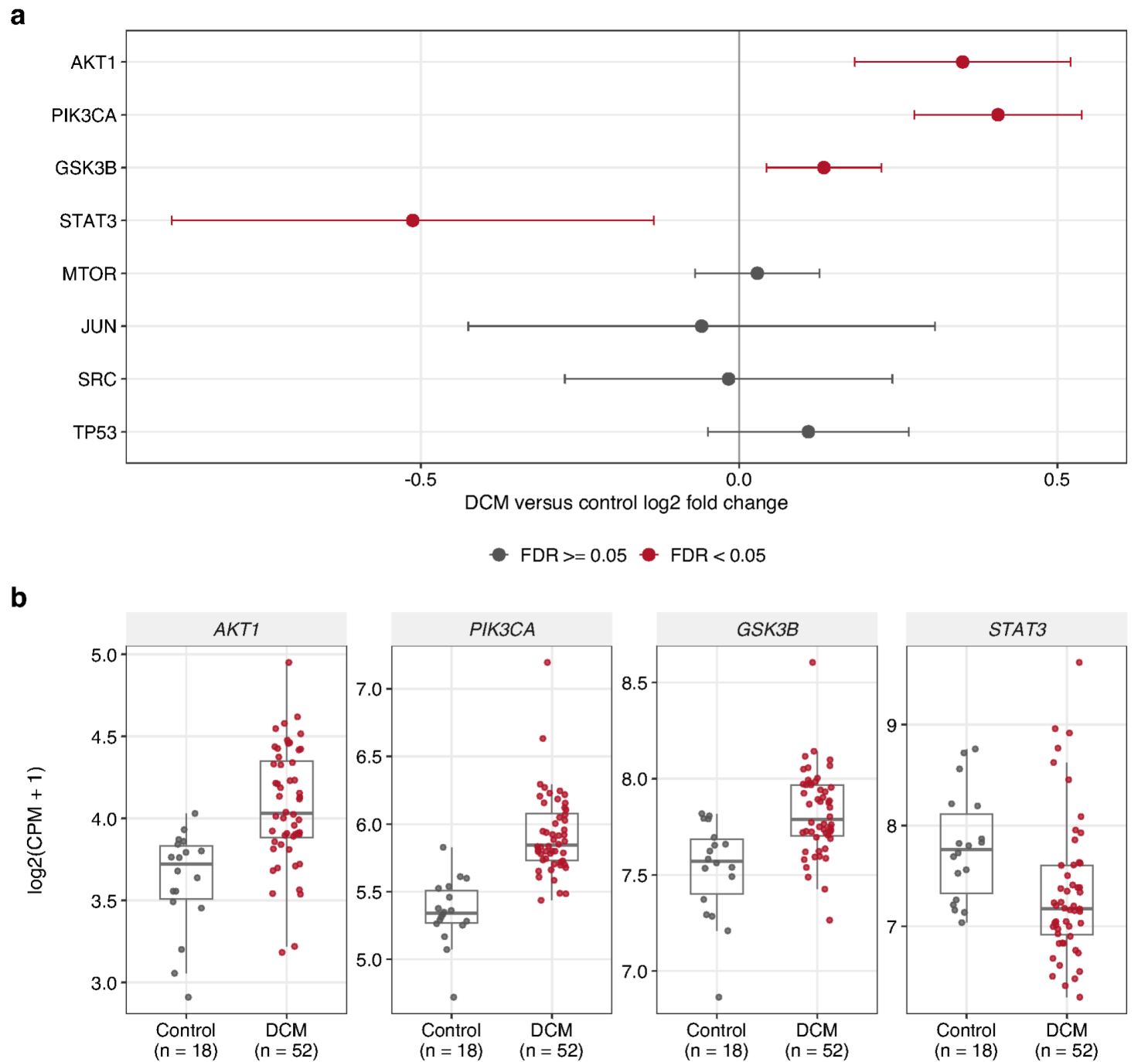


Figure 5

Donor-pseudobulk hub-gene effects in DCM cardiomyocytes

(a) The cardiac-muscle-cell comparison of 52 DCM with 18 control donors, with points showing log2 fold change and bars showing 95% Wald confidence intervals; red indicates adjusted $P < 0.05$. (b) Donor-level pseudobulk expression for the four genes significant in cardiomyocytes, given as counts per million of each donor's cardiomyocyte library on a log2 scale after addition of a pseudocount of one, log2(CPM + 1), as labelled on the vertical axis. Boxes show the median and interquartile range; each point is one donor, and the per-donor values are given in Online Resource 6.

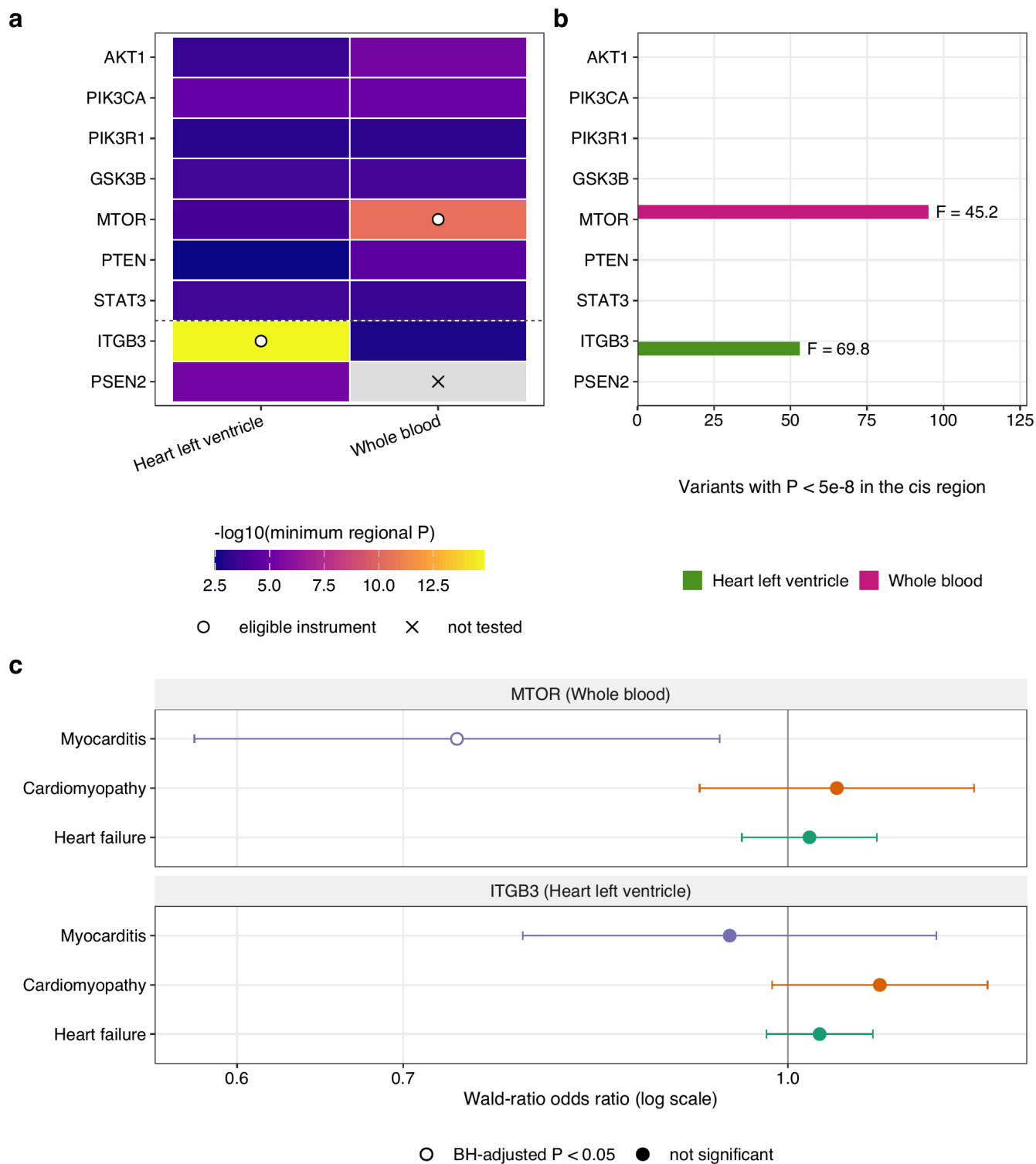


Figure 6

A prespecified instrument screen yields one eligible cis-eQTL

(a) Minimum regional GTEx cis-eQTL P value for the seven primary candidates and the two genes designated as controls in heart left ventricle and whole blood. *ITGB3* and *PSEN2* were designated as controls before the screen, but both are themselves members of the 61-gene Tier-1 candidate set and therefore do not constitute a genuine negative control. The dashed rule separates the primary candidates above from the two designated controls below. Open circles identify gene–tissue combinations meeting $P < 5 \times 10^{-8}$, minor-allele frequency ≥ 0.01 and $F > 10$; a cross marks whole-blood *PSEN2*, for which no regional variant was available. (b) Number of variants reaching $P < 5 \times 10^{-8}$ in each cis region, with the F statistic annotated for the two eligible instruments. (c) Single-variant Wald-ratio estimates across the three FinnGen R12 outcomes for the whole-blood *MTOR* instrument and for the heart *ITGB3* designated

control. Points show odds ratios and bars show 95% confidence intervals on a logarithmic scale; the open point marks the one estimate below 0.05 after Benjamini–Hochberg correction.

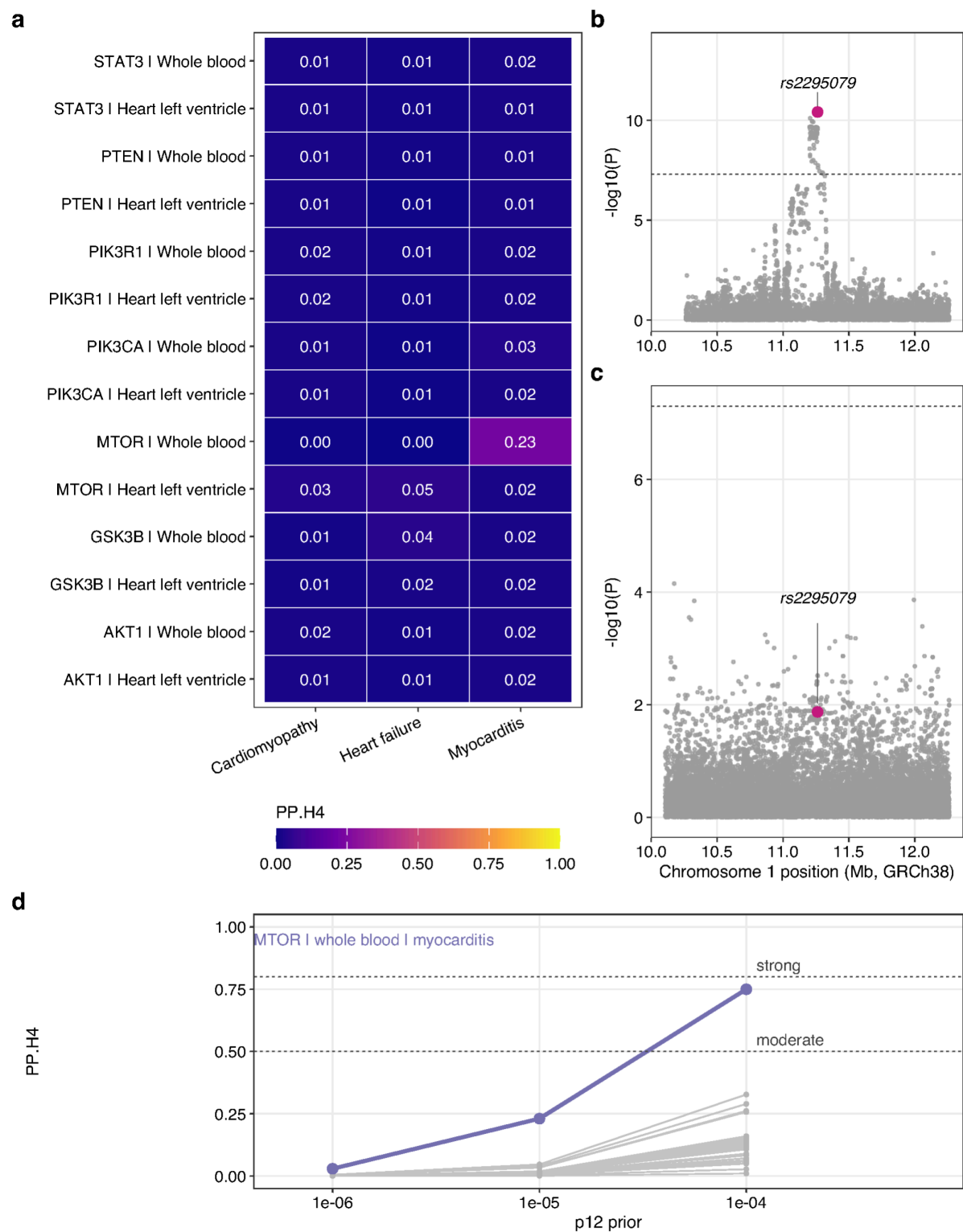


Figure 7

Complete-region colocization gives no positive support for a shared causal variant

(a) PP.H4 from coloc.abf using complete regional summary statistics and priors $p_1 = p_2 = 1 \times 10^{-4}$ and $p_{12} = 1 \times 10^{-5}$. No comparison reached $PP.H4 \geq 0.5$. (b) Regional whole-blood cis-eQTL association for *MTOR*, with the lead instrument *rs2295079* highlighted. (c) Myocarditis association across the same interval, on the same genomic axis. The dashed line in both regional panels is $P = 5 \times 10^{-8}$. The eQTL peak has no counterpart in the outcome data, which is

the pattern behind the low posterior in panel (a). (d) PP.H4 for every primary comparison across the three p12 settings, with the whole-blood *MTOR*–myocarditis comparison highlighted. Its rise from 0.230 to 0.750 at the permissive setting indicates prior sensitivity rather than robust colocalization.

Supplementary Files

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

- [ESM1.xlsx](#)
- [ESM3.xlsx](#)
- [ESM7.xlsx](#)
- [ESM6.xlsx](#)
- [ESM5.xlsx](#)
- [ESM2.xlsx](#)
- [ESM11.xlsx](#)
- [ESM4.xlsx](#)
- [ESM8.xlsx](#)
- [ESM9.xlsx](#)
- [ESM10.pdf](#)
- [ESM12.zip](#)