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.

A reproducible computational pipeline integrating network toxicology, nested machine learning, single-cell transcriptomics and molecular docking nominates cytoskeletal and BMP genes (ACTA2, BMP4, FLNA, FLNB) as bisphenol A–responsive markers of meningioma tissue

Kaihua Cao
Shanghai Jiao Tong University
Zhihua Cheng
Shanghai Jiao Tong University
Yu Zeng
Shanghai Jiao Tong University
Pengran Yu
Shanghai Jiao Tong University
WenChuan Zhang
zhangwench88@sjtu.edu.cn
Shanghai Jiao Tong University

Research Article

Keywords: meningioma, bisphenol A, network toxicology, machine learning, single-cell transcriptomics, molecular docking, WGCNA, computational reproducibility, endocrine disruptor

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11493293/v1

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

Additional Declarations: No competing interests reported.

A reproducible computational pipeline integrating network toxicology, nested machine learning, single-cell transcriptomics and molecular docking nominates cytoskeletal and *BMP* genes (*ACTA2*, *BMP4*, *FLNA*, *FLNB*) as bisphenol A–responsive markers of meningioma tissue

Kaihua Cao, MD, PhD[1]; Zhihua Cheng, MD, PhD[1]; Yu Zeng, MD, PhD[1]; Pengran Yu, MD[1]; WenChuan Zhang, MD, PhD[1],*

[1] Department of Neurosurgery, Shanghai Ninth People's Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200011, China

* Correspondence: WenChuan Zhang, MD, PhD, Department of Neurosurgery, Shanghai Ninth People's Hospital, School of Medicine, Shanghai Jiao Tong University, No. 639 Zhizaoju Road, Shanghai 200011, China. Email: zhangwench88@sjtu.edu.cn

Author emails: Kaihua Cao — caokaihua@shsmu.edu.cn; Zhihua Cheng — eric_cheng027@outlook.com; Yu Zeng — gzzengyu@163.com; Pengran Yu — yupengran@126.com; WenChuan Zhang — zhangwench88@sjtu.edu.cn

Target journal: BMC Bioinformatics (Research article)

Article type: Research article — computational bioinformatics pipeline (no new wet-lab experiments)

Abstract

Background: Bisphenol A (BPA) is a high-production-volume endocrine-disrupting chemical, and meningioma is a common hormone-sensitive intracranial tumour. A systematic, reproducible molecular map linking BPA-responsive genes to meningioma transcriptomes has been lacking. We therefore developed a fully computational pipeline that integrates multi-cohort transcriptomics, chemical–target databases, nested machine learning, single-cell localisation and molecular docking to prioritise candidate

markers while treating all docking and classification outputs as computational evidence rather than clinical causality.

Results: Three GEO microarray cohorts (GSE43290, GSE54934, GSE84263; 165 tumours, 23 normals) were merged on 10,938 genes after ComBat correction. Differential expression (682 DEGs) and WGCNA tumour modules (153 genes) were intersected with BPA targets from SwissTargetPrediction (96), SEA (19), PharmMapper (174), CTD and ChEMBL (five-source union 2,250) and meningioma genes from Open Targets/CTD. Twenty-five BPA-responsive core candidates emerged; consensus PPI×SHAP hubs were *ACTA2*, *BMP4*, *FLNA* and *FLNB* (Table 2 ranks 7.0 / 12.0 / 13.0 / 12.0). Nested tumour-versus-normal classifiers approached a performance ceiling (best external AUC 1.000; 83/109 models = 1.0), interpreted cautiously under class imbalance, whereas WHO-grade prediction was weak (external AUC ≤ 0.677). In GSE183655, *BMP4*/*FLNB* mapped to tumour mesenchymal cells and *ACTA2* to pericyte/smooth-muscle niches. AutoDock Vina affinities were weaker at hubs (−5.17 to −5.70 kcal/mol) than at *ESR1*/*ESR2*/*PTGS2* (−8.21 to −8.49), with symmetry-aware redock RMSDs supporting control-site validity.

Conclusions: This reproducible pipeline provides computational evidence that cytoskeletal and *BMP* genes mark meningioma tissue within a BPA-responsive intersection. Direct high-affinity BPA–hub binding is not supported relative to classical oestrogen receptors and *PTGS2*. Wet-lab and exposure-linked studies are required before translational inference.

Keywords: meningioma; bisphenol A; network toxicology; machine learning; single-cell transcriptomics; molecular docking; WGCNA; computational reproducibility; endocrine disruptor

Background

Meningiomas are among the most frequent primary intracranial tumours in adults and account for a substantial fraction of central nervous system neoplasms in contemporary CBTRUS reports [1]. Beyond epidemiology, integrative molecular classifications and DNA-

methylation groups have refined biological subtypes and clinical risk strata [2,17,18]. A durable clinical observation is female predominance together with abundant sex-steroid receptor expression, particularly progesterone signalling [2,7]. Pharmacoepidemiologic evidence has strengthened this hormonal link: prolonged use of selected progestogens—including cyproterone, nomegestrol and related agents—is associated with intracranial meningioma [3,6,8], depot medroxyprogesterone acetate has been linked to meningioma risk in recent US data [4], and contraceptive exposure continues to be examined in surgically treated cohorts [5]. Together, these observations place meningioma at the interface of neurosurgery and endocrinology and motivate computational investigation of environmental xenoestrogens as candidate modifiers of meningeal biology.

Bisphenol A (BPA) is a high-production-volume plasticiser and prototypical endocrine-disrupting chemical. It engages classical oestrogen receptors (*ESR1*/*ESR2*) and related signalling cascades, and reviews have linked BPA exposure to hormone-related phenotypes and neurological effects [9–12]. Beyond nuclear receptors, BPA can engage G protein-coupled oestrogen receptor pathways and perturb inflammatory and metabolic programmes, providing a plausible—but unproven—route toward hormone-sensitive intracranial tumours [10,12,13]. To our knowledge, a systematic molecular map connecting BPA-responsive genes to meningioma transcriptomes has not been published.

Network toxicology integrates chemical–target prediction and curated interaction databases with disease transcriptomics to generate mechanistic hypotheses for environmental chemicals. Recent applications include BPA neurotoxicity and BPA-associated cholestatic injury [14,15]. Zhan et al. combined network toxicology with multi-cohort transcriptomics, weighted gene co-expression network analysis (WGCNA), large-scale machine learning (ML), single-cell

analysis and molecular docking—including wet-lab validation—for 6PPD/6PPD-Q and atrial fibrillation [16]. Here we transplant a comparable in silico framework to BPA and meningioma, with emphasis on computational reproducibility, explicit software versions and structural docking as orthogonal evidence layers suitable for a bioinformatics readership. We prioritise cytoskeletal and *BMP*-pathway genes (*ACTA2*, *BMP4*, *FLNA*, *FLNB*), interpret classification performance in light of class imbalance and tissue composition, and report limitations of each data source openly (sparse normal meninges, database coverage limits, AUC saturation, absent exposure metadata, and no wet-lab assays). All docking, ML and network outputs are framed as computational evidence rather than clinical causality.

Methods

Study design and computational pipeline

This study is entirely computational: no new human participants, animals or wet-lab assays were enrolled. The aim was to build a reproducible priority list of BPA-responsive genes marking meningioma tissue identity. The workflow comprises (i) multi-cohort GEO transcriptomic acquisition and ComBat merging; (ii) differential expression and WGCNA; (iii) BPA target assembly from SwissTargetPrediction, SEA, PharmMapper, CTD and ChEMBL plus meningioma genes from Open Targets/CTD; (iv) intersection, STRING PPI and GO/KEGG enrichment; (v) nested feature-selection × classifier machine learning with SHAP and external validation; (vi) single-cell localisation (GSE183655); (vii) median-split single-gene GSEA; and (viii) AutoDock Vina docking with redocking controls. All numerical claims were taken from the project results/ directory; analysis scripts reside in scripts/. Random seeds for docking (three

independent Vina runs) and stratified ML cross-validation repeats are fixed in the accompanying code.

Software environment and versions

Analyses were run in a pinned conda/mamba environment (Python 3.11.17; R 4.4.3). Key packages and tools (versions verified in the analysis environment): limma 3.62.1; sva 3.54.0; WGCNA 1.74; clusterProfiler 4.14.0; scikit-learn 1.9.1; SHAP 0.51.0; scanpy 1.11.5; harmonypy 2.1.0; pandas 2.3.3; numpy 2.4.6; AutoDock Vina 1.2 (build f458505-mod); STRING API v12; PyMOL for pose rendering. Large language models do not meet authorship criteria and were not used to generate scientific content, analyses, figures or conclusions; pipeline logic and numerical outputs derive solely from the scripted analyses described below.

Transcriptomic data and cohorts

GEO series matrices were downloaded for GSE43290 (Affymetrix GPL96; 47 tumours, 4 commercial normal meninges), GSE54934 (GPL6244; 22 tumours, 3 normals) and GSE84263 (GPL10558; 112 samples: 96 tumours, 16 normals). Expression values were log2-transformed where required and quantile-normalised. Probes were collapsed to gene symbols by highest mean intensity. After merging on 10,938 common genes (188 samples: 165 tumours, 23 normals), batch effects were removed with ComBat (sva) while preserving tumour/normal group labels and inspected by principal component analysis (PCA) [25,26]. GSE285889 (FPKM; 33 tumour and 7 normal multi-region samples from few donors) served as the external tumour-versus-normal cohort. For WHO-grade prediction, GSE16581 (GPL570; 43 WHO grade 1 versus 25 grade 2/3) was used for training and GSE43290 tumour samples with grade annotation (33 versus 14) for external validation. Cohort composition is summarised in Table 1.

Differential expression and WGCNA

Differential expression used limma with thresholds |log2 fold-change (FC)| > 1 and Benjamini–Hochberg (BH)–adjusted $P < 0.05$ [24]. WGCNA used the 5,000 genes with highest median absolute deviation, soft-thresholding power 10 (scale-free topology fit index $R^2 \geq 0.85$), minimum module size 30 and merge cut height 0.25 [23]. Modules with absolute Pearson correlation $|r| \geq 0.5$ and $P < 0.05$ for tumour status were retained as tumour-associated modules.

Chemical and disease target databases

BPA putative targets were assembled from five sources: SwissTargetPrediction (Homo sapiens; probability > 0; 96 unique gene symbols from 100 ranked targets) [38], SEA human UniProt hits (19 genes) [39], PharmMapper human pharmacophore hits mapped to HGNC symbols (174 genes) [40], CTD chemical–gene interactions for C006780 filtered to ≥ 3 supporting PubMed identifiers (2,081 genes) [29], and ChEMBL CHEMBL418971 with pChEMBL ≥ 5 (6 genes) [30]. The five-source union contained 2,250 genes; the three web-tool exports contributed 168 genes absent from CTD/ChEMBL. SwissTargetPrediction, SEA and PharmMapper were queried via their public web interfaces (BPA SMILES CC(C)(C1=CC=C(C=C1)O)C2=CC=C(C=C2)O); UniProt accessions/entry names were mapped to gene symbols before merging. Meningioma-associated genes comprised Open Targets MONDO_0016642 with association score ≥ 0.1 (355 genes) [31] and CTD curated disease genes (16), union 360; a GeneCards manual export was not available. CTD chemical–gene links predominantly reflect expression associations rather than confirmed ligand–protein binding and were interpreted accordingly.

Intersection, PPI network and enrichment

Core candidates were defined as BPA targets ∩ (DEG ∩ WGCNA tumour modules). An extended set was BPA ∩ (DEG ∩ WGCNA ∪ disease genes). Protein–protein interactions (PPI) were retrieved from STRING v12 (combined score ≥ 0.4) [27]. Genes were ranked by the sum of ranks of degree, betweenness, closeness and maximal-clique centrality. Gene Ontology (GO) and KEGG enrichment used clusterProfiler with the array gene universe as background and BH correction [28].

Nested machine learning and hub prioritisation

Eight feature selectors × 12 classifiers yielded 109 unique combinations after removal of redundant selector–classifier pairs. Feature selection was nested within repeated stratified 5-fold cross-validation (CV; 5 repeats) to limit information leakage. Models were ranked by mean CV area under the receiver-operating characteristic curve (AUC) and then by external AUC; Brier scores were recorded as calibration summaries. SHapley Additive exPlanations (SHAP; permutation explainer) interpreted the best tumour-versus-normal model. Final hubs combined PPI rank with SHAP rank (results/05b_hub_ranking.csv), yielding *ACTA2*, *BMP4*, *FLNA* and *FLNB* (Table 2). Oriented single-gene ROC curves are reported for training and external cohorts. An earlier 98-model pipeline that used a broader candidate set (including *PTGS2*, *FLT1*, *UNG* and *JUN*) was retained only as sensitivity analysis and was not used for hub calling. Classification performance is reported as computational evidence of tissue separation, not as a clinical diagnostic claim.

Single-cell RNA-seq

Public 10x Genomics data from GSE183655 (10 samples: 6 tumours, 2 brain–tumour interfaces [BTI], 2 dura) were processed with quality filters of 200–7,000 detected genes,

mitochondrial read fraction < 20% and > 500 unique molecular identifiers, yielding 80,901 cells of which 30,000 were randomly sampled for downstream analysis. After normalisation and Harmony batch integration in a Seurat-compatible / scanpy workflow, Leiden clustering (19 clusters) was annotated into eight cell types by marker-gene scoring; copy-number variation inference was not performed [35,36]. Recent single-cell atlases of dura and meningioma informed interpretation of mesenchymal and immune niches but were not re-analysed here [19,20].

Single-gene GSEA

Within the 165 training tumours, samples were split at the median expression of each hub gene. Genes were ranked by the limma t-statistic contrasting High versus Low groups (adjusting for batch), and KEGG gene-set enrichment analysis (gseKEGG) was run with BH-adjusted $P < 0.05$, following the GSEA conceptual framework [37].

Molecular docking

Docking used AutoDock Vina 1.2 with three independent random seeds [32,33]. Hub receptors were crystal structures of *FLNA* (PDB 3HOC) and *FLNB* (3FER), or AlphaFold models trimmed to per-residue confidence pLDDT ≥ 70 for *ACTA2* (UniProt P62736) and *BMP4* (P12644), with fpocket pocket 1 as the search site [34]. Positive controls were *ESR1* (PDB 3UU7, co-crystallised BPA), *ESR2* (3OLS) and *PTGS2* (5KIR). Protocol validity was assessed by symmetry-aware redocking root-mean-square deviation (RMSD; obrms). Protein–ligand interactions were inspected with PLIP; poses were rendered in PyMOL (ray-traced PNG, 300 dpi). Docking affinities are reported as computational structural evidence of relative pocket preference and are not interpreted as measured binding constants.

Statistical reporting

Unless stated otherwise, adjusted P values are BH-corrected. AUC values are reported as mean ± standard deviation across CV repeats where applicable. Reporting follows computational reproducibility practice: scripted pipelines, deposited GEO accessions, fixed thresholds and versioned software.

Results

Cohort overview, DEGs and WGCNA (Figure 1; Table 1)

ComBat removed batch-dominant PCA clustering while preserving tumour/normal separation. We identified 682 DEGs (221 up-regulated, 461 down-regulated) at $|\log2FC| > 1$ and adjusted $P < 0.05$; a more permissive $|\log2FC| > 0.585$ threshold yielded 1,649 genes. WGCNA produced seven modules; brown and black modules correlated with tumour status and together contained 153 genes. Soft-threshold power 10 satisfied $R^2 \geq 0.85$.

Table 1. Transcriptomic cohorts used in this study.

Cohort (GEO)	Platform	Role	Tumours (n)	Normals (n)	Notes
GSE43290	GPL96	Training (TN); grade external	47 (TN); 33 vs 14 (grade)	4	Commercial normal meninges
GSE54934	GPL6244	Training (TN)	22	3	—
GSE84263	GPL10558	Training (TN)	96	16	—
Merged training	—	DEG / WGCNA / ML train	165	23	10,938 common genes; ComBat
GSE285889	RNA-seq FPKM	External TN validation	33	7	Multi-region; few donors
GSE16581	GPL570	Grade training	43 grade 1 / 25 grade 2–3	—	WHO grade labels
GSE183655	10x scRNA-seq	Cell localisation	6 tumour / 2 BTI / 2 dura samples	—	30,000 cells analysed

TN, tumour versus normal; BTI, brain–tumour interface.

BPA–meningioma candidates and PPI (Figure 2)

Intersection of DEGs with tumour modules yielded 139 genes, of which 25 were BPA targets: *ACTA2*, *AGT*, *ALDOC*, *ALPL*, *BMP4*, *DHRS3*, *DIO2*, *EDN1*, *FGFR3*, *FLNA*, *FLNB*, *GDPD5*, *GLG1*, *GPRC5B*, *KDR*, *MAPT*, *MYO1B*, *NDRG2*, *NOS3*, *PDE4B*, *PDGFRA*, *RAMP1*, *SLC2A1*, *SORBS2* and *TINAGL1* (*MAPT* from SwissTargetPrediction; *PDE4B* from PharmMapper). Three (*FGFR3*, *KDR*, *PDGFRA*) were also present in meningioma disease databases. The extended BPA-intersected set contained 125 genes. The STRING network of the 25 core candidates contained 29 edges among 19 connected nodes; *MAPT* (degree 7) and *KDR* (degree 6) ranked highest by degree, with *FLNA*, *BMP4* and *ACTA2* remaining among the top topological hubs. Consensus PPI×SHAP ranking continued to prioritise *ACTA2*, *BMP4*, *FLNA* and *FLNB*.

Functional enrichment (Figure 3)

Enrichment recovered 267 biological process, 18 cellular component and 11 molecular function GO terms (BH-adjusted $P < 0.05$). Top biological processes centred on vasculature development; receptor complex and growth factor binding ranked among top cellular component / molecular function terms. Fourteen KEGG pathways were significant (adjusted P 0.024–0.048), including HIF-1 signalling, vascular smooth muscle contraction, fluid shear stress and atherosclerosis, integrin signalling (*FLNA*/*FLNB*/*KDR*/*PDGFRA*), focal adhesion and pathways in cancer.

Machine learning and hub genes (Figure 4; Table 2)

Tumour-versus-normal classification approached a performance ceiling: external AUC = 1.000 in 83 of 109 models. The top model (elastic-net feature selection + support vector machine; Enet+SVM) used 11 genes (*ACTA2*, *AGT*, *BMP4*, *FLNA*, *FLNB*, *GDPD5*, *GLG1*,

GPRC5B, *PDGFRA*, *SORBS2*, *TINAGL1*) and achieved CV AUC 0.999 ± 0.004 (Brier 0.009) and external AUC 1.000 (Brier 0.001). Top SHAP features were *ACTA2*, *GPRC5B*, *FLNB* and *TINAGL1*. Consensus hubs (lower sum of PPI and SHAP ranks = higher priority) were *ACTA2* (7.0), *BMP4* (12.0), *FLNB* (12.0) and *FLNA* (13.0). In bulk tumour-versus-normal limma contrasts, *ACTA2* and *FLNA* were down-regulated (log2FC −2.17 and −1.46) whereas *BMP4* and *FLNB* were up-regulated (log2FC 2.20 and 1.71); all four had adjusted $P < 1\times10^{-19}$. Oriented single-gene external AUCs were 0.883 (*ACTA2*), 0.857 (*BMP4*), 0.827 (*FLNA*) and 0.939 (*FLNB*). WHO-grade discrimination was weak: the best model (glmBoost+XGBoost) reached CV AUC 0.763 ± 0.116 and external AUC 0.658 (no model exceeded external AUC 0.677). High AUCs are presented as computational evidence that the hubs mark meningioma tissue identity, not as proof of BPA causality or grade prognostication.

Table 2. Consensus hub genes prioritised for BPA–meningioma biology.

Gene	PPI rank	SHAP rank	Consensus	log2FC (T vs N)	Adj. P	Ext. AUC (oriented)	Predominant scRNA niche
ACTA2	6.0	1.0	7.0	−2.17	3.3×10^{-22}	0.883	Pericyte/SMC (65.7% cells)
BMP4	5.0	7.0	12.0	+2.20	2.0×10^{-20}	0.857	Tumour mesenchymal (62.5%)
FLNA	4.0	9.0	13.0	−1.46	3.1×10^{-23}	0.827	Broad (29–72% non-dura)
FLNB	9.0	3.0	12.0	+1.71	5.6×10^{-35}	0.939	Tumour mesenchymal (85.2%)

Consensus = PPI rank + SHAP rank (lower is better). Ext. AUC from GSE285889 sample-level oriented ROC. scRNA percentages from GSE183655 tumour compartment where applicable. Hub calling order follows results/hub_top4.txt (*ACTA2*, *BMP4*, *FLNA*, *FLNB*).

Single-cell localisation (Figure 5)

Eight annotated cell types were recovered. Tumour-resident fibroblast/mesenchymal cells were the largest compartment in tumour (45.2%) and BTI (43.1%) samples. *BMP4* was detected in 62.5% and *FLNB* in 85.2% of tumour mesenchymal cells; *ACTA2* was detected in 65.7% of pericyte/smooth-muscle cells; *FLNA* was broadly expressed (approximately 29–72% across non-dura types).

Single-gene GSEA (Figure 6)

Significant KEGG gene sets numbered 38 (*ACTA2*), 26 (*BMP4*), 34 (*FLNA*) and 45 (*FLNB*). High-*BMP4* tumours were enriched for cytoskeleton in muscle cells and integrin/neurotrophin signalling; high-*FLNA* tumours were enriched for cytoskeleton in muscle cells, focal adhesion and integrin signalling. Oxidative phosphorylation and ribosome gene sets were depleted in high-*BMP4* and high-*FLNB* groups.

Molecular docking (Figure 7)

Symmetry-aware redock RMSDs were 1.27 Å (*ESR1*), 0.42 Å (*ESR2*) and 0.89 Å (*PTGS2*), supporting protocol validity at the control sites. BPA affinities were −8.21 (*ESR1*), −8.35 (*ESR2*) and −8.49 kcal/mol (*PTGS2*). Affinities at hub pockets were substantially weaker: *ACTA2* −5.70, *BMP4* −5.68, *FLNA* −5.70 and *FLNB* −5.17 kcal/mol. Relative Vina scores therefore provide structural-computational evidence against preferential high-affinity BPA engagement of the four hubs compared with classical oestrogen receptors and *PTGS2*.

Discussion

Principal findings

Using an integrated network-toxicology design, we prioritised four BPA-responsive genes as computational markers of meningioma tissue: contractile/cytoskeletal genes that are down-regulated in bulk meningioma (*ACTA2*, *FLNA*) and up-regulated *BMP*/filamin partners (*BMP4*, *FLNB*). These hubs sit at the intersection of BPA target lists, multi-cohort differential expression and tumour-associated co-expression modules, and they recur in both PPI topology and SHAP explanations of the best tumour-versus-normal classifier. Docking, however, does not support high-affinity direct binding of BPA to the hub proteins relative to classical oestrogen receptors and *PTGS2*. The overall picture is therefore one of indirect, receptor- and pathway-mediated association rather than of BPA as a high-affinity ligand of *ACTA2*/*BMP4*/*FLNA*/*FLNB*. We deliberately avoid clinical causal claims: no GEO cohort carries BPA exposure metadata, and no functional assays were performed.

Mechanistic synthesis framed as computational evidence

Meningioma biology is tightly coupled to hormonal context [2–8]. BPA is a xenoestrogen that binds *ESR1*/*ESR2* and can engage G protein-coupled oestrogen receptor signalling [9,11,13]. In parallel, BPA has been linked to inflammatory and metabolic reprogramming in multiple tissues [10,12,14,15]. Our enrichment signature—vascular development, HIF-1 signalling, integrin/focal-adhesion programmes and pathways in cancer—is consistent with a stromal and angiogenic phenotype rather than a single oncogene driver. Filamins (*FLNA*, *FLNB*) mechanically couple integrins to the actin cytoskeleton [22]; their joint appearance with *ACTA2* (smooth-muscle actin) and *BMP4* (a *TGF*-β/*BMP* family morphogen) suggests a mesenchymal/contractile axis that could shape meningeal niche architecture. *TGF*-β family signalling has long been implicated in meningioma pathogenesis and microenvironment crosstalk [21]. Single-cell atlases of dura and the brain–tumour interface likewise emphasise stromal,

immune and interface compartments as dominant axes of meningioma heterogeneity [19,20]. Up-regulation of *BMP4* in tumour-resident mesenchymal cells in GSE183655 therefore aligns with a developmental morphogen being redeployed in the meningioma stroma, whereas bulk down-regulation of *ACTA2* may partly reflect altered perivascular/smooth-muscle composition rather than a cell-intrinsic tumour-cell loss of contractile identity alone.

A coherent working model—presented strictly as a hypothesis for experimental testing—is: BPA → oestrogen-receptor and inflammatory intermediates (including *PTGS2*-related eicosanoid tone as a docking-supported control axis) → transcriptional shifts in cytoskeletal and *BMP* programmes within meningioma-associated mesenchymal and perivascular niches → bulk expression contrasts that separate tumour from normal meninges. Because the chemical link is database-mediated rather than exposure-linked, this model remains computational evidence pending validation.

Comparison with related literature and methodological templates

Pharmacoepidemiology has established that selected exogenous progestogens increase meningioma risk [3–8], whereas environmental xenoestrogens such as BPA remain comparatively underexplored in neuro-oncology. Network-toxicology studies of BPA in other diseases [14,15] and the multi-layered 6PPD/6PPD-Q atrial fibrillation template of Zhan et al. [16] demonstrate that intersecting chemical targets with multi-cohort transcriptomics, WGCNA, ML, single-cell maps and docking can surface biologically coherent candidates. Molecular classifications of meningioma [17,18] provide a complementary disease context against which stromal hub genes should eventually be stratified. Our study transplants that design into meningioma and deliberately retains honesty regarding database coverage and performance

ceilings. Unlike Zhan et al., we report no wet-lab exposure or functional assays; translational claims are accordingly withheld.

Interpreting AUC = 1.000 and class imbalance honestly

External tumour-versus-normal AUC reached 1.000 for the top model and for 83/109 combinations. Perfect separation on a small external set (33 tumours, 7 normals from few donors) should not be read as a clinically deployable diagnostic signature. Several non-exclusive explanations apply: (i) strong tissue-composition differences between meningioma and normal meninges; (ii) residual handling or stress programmes (immediate-early genes are prominent among down-regulated DEGs); (iii) severe class imbalance and limited independent normal donors, which inflate apparent discrimination and destabilise patient-level estimates; and (iv) feature sets that capture stroma-rich tumour identity rather than BPA-specific causation. Consistent with this reading, the same gene space failed to stratify WHO grade (best external AUC 0.658). We therefore present high AUCs as computational evidence that the hubs (and related candidates) mark meningioma tissue identity, not as proof of BPA causality or of grade prognostication.

Translational implications (hypothesis-generating only)

For a bioinformatics audience, the immediate value of these results is a transparent, multi-layer, versioned priority list rather than bedside application. If future exposure-linked cohorts confirm BPA-responsive shifts in *BMP4*/*FLNB*-high mesenchymal niches or *ACTA2*-low perivascular programmes, research priorities would include: (i) immunohistochemistry or spatial transcriptomics of the four hubs in hormone-related meningioma subsets; (ii) in vitro BPA titration in meningioma and meningeal primary cultures with ESR blockade; and (iii) careful

pharmacoepidemiologic co-modelling of xenoestrogen biomarkers alongside progestogen exposure. Until such data exist, the hubs should be treated as candidate biomarkers of tissue state, not as therapeutic targets, exposure assays or clinical decision tools.

Limitations

Several constraints bound interpretation and are listed explicitly: (1) normal meninges are few and heterogeneous (23 training samples; 7 external samples from few donors), so tumour-versus-normal contrasts are fragile; (2) ML performance saturates (83/109 external AUC = 1.0), and separation may reflect composition or handling rather than BPA-specific biology; (3) GeneCards and STITCH exports were unavailable, so meningioma disease genes relied on Open Targets/CTD, whereas BPA targets include web SwissTargetPrediction/SEA/PharmMapper alongside CTD/ChEMBL; CTD links remain largely expression associations [29–31,38–40]; (4) 109 rather than 113 model combinations remained after removing identical pairs; (5) single-cell annotation used marker scoring without CNV calling and analysed a 30,000-cell subset; (6) no cohort provides BPA exposure metadata and no wet-lab validation was performed; *ACTA2*/*BMP4* docking used AlphaFold models [34]; (7) docking scores are relative computational estimates, not experimental Kd/Ki values. These limitations reinforce the hypothesis-generating framing of all conclusions.

Conclusions

An integrated bioinformatics and structural docking pipeline prioritises *ACTA2*, *BMP4*, *FLNA* and *FLNB* as BPA-responsive computational markers of meningioma tissue, with single-cell support for mesenchymal and perivascular niches. Direct high-affinity binding of BPA to

these hub proteins is not supported relative to *ESR1*/*ESR2*/*PTGS2*. Experimental validation and exposure-linked clinical cohorts are required before any translational or clinical inference.

List of abbreviations

AUC: area under the receiver-operating characteristic curve

BH: Benjamini–Hochberg

BPA: bisphenol A

BTI: brain–tumour interface

CV: cross-validation

DEG: differentially expressed gene

FC: fold-change

GEO: Gene Expression Omnibus

GO: Gene Ontology

GSEA: gene set enrichment analysis

KEGG: Kyoto Encyclopedia of Genes and Genomes

ML: machine learning

PCA: principal component analysis

PDB: Protein Data Bank

PPI: protein–protein interaction

RMSD: root-mean-square deviation

SHAP: SHapley Additive exPlanations

SMC: smooth-muscle cell

STRING: Search Tool for the Retrieval of Interacting Genes/Proteins

TN: tumour versus normal

WGCNA: weighted gene co-expression network analysis

WHO: World Health Organization

Declarations

Ethics approval and consent to participate

Not applicable. This study analysed only publicly available, de-identified transcriptomic datasets from the NCBI Gene Expression Omnibus; no new human participants, human tissue or animals were enrolled, and no institutional ethics approval was required for secondary analysis of public data.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analysed during the current study are available in the NCBI Gene Expression Omnibus repository under accessions GSE43290 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE43290), GSE54934 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54934), GSE84263 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84263), GSE285889

(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE285889), GSE16581 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16581) and GSE183655 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183655). Numerical outputs supporting the findings are provided in the project results/ directory; analysis scripts are in scripts/. Further materials are available from the corresponding author on reasonable request.

Software and pipeline summary (BMC software-statement style):

- Project name: BPA–meningioma network toxicology computational pipeline
- Project home page / code: available from the corresponding author on reasonable request (scripts/ in the project archive)
- Operating system(s): Platform independent (developed on Linux)
- Programming language: R 4.4.3; Python 3.11
- Other requirements: packages listed in Methods (limma, sva, WGCNA, clusterProfiler, scikit-learn, SHAP, scanpy, AutoDock Vina 1.2)
- License: analysis scripts intended for academic reuse upon request
- Any restrictions to use by non-academics: none beyond third-party tool licences

Competing interests

The authors declare that they have no competing interests.

Funding

None. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. The funder (none) had no role in study design, data collection, analysis, interpretation or manuscript preparation.

Authors' contributions

KC conceived the study, developed the computational pipeline, performed analyses and drafted the manuscript. ZC contributed to methodology validation and formal analysis. YZ contributed to investigation and manuscript revision. PY contributed to data curation, visualisation and manuscript revision. WZ supervised the project, administered the study and revised the manuscript critically. All authors read and approved the final manuscript. Large language models do not satisfy authorship criteria and are not authors of this work.

Acknowledgements

Not applicable.

Authors' information

Not applicable.

References

1. Price M, Ballard C, Benedetti J, Neff C, Cioffi G, Waite KA, et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2017-2021. Neuro Oncol. 2024;26(Supplement_6):vi1-85. https://doi.org/10.1093/neuonc/noae145

2. Wang JZ, Landry AP, Raleigh DR, Sahm F, Walsh KM, Goldbrunner R, et al. Meningioma: International Consortium on Meningiomas consensus review on scientific advances and treatment paradigms for clinicians, researchers, and patients. Neuro Oncol. 2024;26(10):1742-80. https://doi.org/10.1093/neuonc/noae082

3. Roland N, Neumann A, Hoisnard L, Duranteau L, Froelich S, Zureik M, et al. Use of progestogens and the risk of intracranial meningioma: national case-control study. BMJ. 2024;384:e078078. https://doi.org/10.1136/bmj-2023-078078

4. Xiao T, Kumar P, Lobbous M, Yogi-Morren D, Soni P, Recinos PF, et al. Depot Medroxyprogesterone Acetate and Risk of Meningioma in the US. JAMA Neurol. 2025;82(11):1094-102. https://doi.org/10.1001/jamaneurol.2025.3011

5. Chen S, Jugl S, Jackson L, Rahman M, Antonelli PJ, Bruggeman B, et al. Risk of intracranial meningioma requiring surgical intervention among recent hormonal contraceptives users. J Neurooncol. 2026;176(2):156. https://doi.org/10.1007/s11060-025-05402-5

6. Nguyen P, Roland N, Neumann A, Hoisnard L, Passeri T, Duranteau L, et al. Prolonged use of nomegestrol acetate and risk of intracranial meningioma: a population-based cohort study. Lancet Reg Health Eur. 2024;42:100928. https://doi.org/10.1016/j.lanepe.2024.100928

7. Miyagishima DF, Moliterno J, Claus E, Günel M. Hormone therapies in meningioma-where are we?. J Neurooncol. 2023;161(2):297-308. https://doi.org/10.1007/s11060-022-04187-1

8. Hudelist B, Roux A, Huet-Mignaton E, Dufaure-Gare I, Moiraghi A, Elia A, et al. Progestogen use and the risk of intracranial meningioma: a systematic review and meta-analysis. EClinicalMedicine. 2026;92:103791. https://doi.org/10.1016/j.eclinm.2026.103791

9. Murata M, Kang JH. Bisphenol A (BPA) and cell signaling pathways. Biotechnol Adv. 2018;36(1):311-27. https://doi.org/10.1016/j.biotechadv.2017.12.002

10. Costa HE, Cairrao E. Effect of bisphenol A on the neurological system: a review update. Arch Toxicol. 2024;98(1):1-73. https://doi.org/10.1007/s00204-023-03614-0

11. Patisaul HB. Achieving CLARITY on bisphenol A, brain and behaviour. J Neuroendocrinol. 2020;32(1):e12730. https://doi.org/10.1111/jne.12730

12. Ma Y, Liu H, Wu J, Yuan L, Wang Y, Du X, et al. The adverse health effects of bisphenol A and related toxicity mechanisms. Environ Res. 2019;176:108575. https://doi.org/10.1016/j.envres.2019.108575

13. Prossnitz ER, Barton M. The G protein-coupled oestrogen receptor GPER in health and disease: an update. Nat Rev Endocrinol. 2023;19(7):407-24. https://doi.org/10.1038/s41574-023-00822-7

14. Xu Z, Wang A, Hu J, Li Z, Zheng Y, Liu Y, et al. Elucidating the mechanisms of environmental pollutant BPA inducing neurotoxicity through metabolomics and network toxicology strategy. Environ Pollut. 2025;381:126591. https://doi.org/10.1016/j.envpol.2025.126591

15. Tang Y, She L, Sun Y, Zhu Z, Zhu P. Network toxicology and molecular docking reveal the mechanistic basis of bisphenol A-mediated cholestatic liver injury. Toxicol Rep. 2026;17:102167. https://doi.org/10.1016/j.toxrep.2025.102167

16. Zhan W, Zhao J, Li Y, Ran S, Lai L, Wu J, et al. Network toxicology, machine learning, single-cell analysis, and molecular docking reveal 6PPD/6PPD-Q-associated atrial fibrillation signatures. J Adv Res. 2026. https://doi.org/10.1016/j.jare.2026.10.032

17. Nassiri F, Liu J, Patil V, Mamatjan Y, Wang JZ, Hugh-White R, et al. A clinically applicable integrative molecular classification of meningiomas. Nature. 2021;597(7874):119-25. https://doi.org/10.1038/s41586-021-03850-3

18. Choudhury A, Magill ST, Eaton CD, Prager BC, Chen WC, Cady MA, et al. Meningioma DNA methylation groups identify biological drivers and therapeutic vulnerabilities. Nat Genet. 2022;54(5):649-59. https://doi.org/10.1038/s41588-022-01061-8

19. Wang AZ, Bowman-Kirigin JA, Desai R, Kang LI, Patel PR, Patel B, et al. Single-cell profiling of human dura and meningioma reveals cellular meningeal landscape and insights into meningioma immune response. Genome Med. 2022;14(1):49. https://doi.org/10.1186/s13073-022-01051-9

20. Huang B, Liang J, Tang X, Zhu Y, Gao L, Fu D, et al. Single-cell transcriptomic analysis reveals metastatic and immunosuppressive characteristics in meningioma brain-tumor interface. J Transl Med. 2025;23(1):930. https://doi.org/10.1186/s12967-025-06935-z

21. Johnson MD. Transforming Growth Factor Beta Family in the Pathogenesis of Meningiomas. World Neurosurg. 2017;104:113-9. https://doi.org/10.1016/j.wneu.2017.03.058

22. Stossel TP, Condeelis J, Cooley L, Hartwig JH, Noegel A, Schleicher M, et al. Filamins as integrators of cell mechanics and signalling. Nat Rev Mol Cell Biol. 2001;2(2):138-45. https://doi.org/10.1038/35052082

23. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. https://doi.org/10.1186/1471-2105-9-559

24. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007

25. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118-27. https://doi.org/10.1093/biostatistics/kxj037

26. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882-3. https://doi.org/10.1093/bioinformatics/bts034

27. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638-46. https://doi.org/10.1093/nar/gkac1000

28. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2(3):100141. https://doi.org/10.1016/j.xinn.2021.100141

29. Davis AP, Wiegers TC, Sciaky D, Barkalow F, Strong M, Wyatt B, et al. Comparative Toxicogenomics Database's 20th anniversary: update 2025. Nucleic Acids Res. 2025;53(D1):D1328-34. https://doi.org/10.1093/nar/gkae883

30. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1):D945-54. https://doi.org/10.1093/nar/gkw1074

31. Buniello A, Suveges D, Cruz-Castillo C, Llinares MB, Cornu H, Lopez I, et al. Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. Nucleic Acids Res. 2025;53(D1):D1467-75. https://doi.org/10.1093/nar/gkae1128

32. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-61. https://doi.org/10.1002/jcc.21334

33. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model. 2021;61(8):3891-8. https://doi.org/10.1021/acs.jcim.1c00203

34. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. https://doi.org/10.1038/s41586-021-03819-2

35. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289-96. https://doi.org/10.1038/s41592-019-0619-0

36. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

37. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50. https://doi.org/10.1073/pnas.0506580102

38. Daina A, Michielin O, Zoete V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. Nucleic Acids Res. 2019;47(W1):W357-64. https://doi.org/10.1093/nar/gkz382

39. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. Nat Biotechnol. 2007;25(2):197-206. https://doi.org/10.1038/nbt1284

40. Wang X, Shen Y, Wang S, Li S, Zhang W, Liu X, et al. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. Nucleic Acids Res. 2017;45(W1):W356-60. https://doi.org/10.1093/nar/gkx374

Figure legends

Figure 1. Training-cohort overview, differential expression and WGCNA.

(A) Sample composition by GEO batch and tumour/normal group (n = 188; 165 tumours, 23 normals).

(B, C) PCA before and after ComBat batch correction.

(D) Volcano plot of DEGs ($|\log2FC| > 1$, BH-adjusted $P < 0.05$; 682 DEGs: 221 up, 461 down).

(E) Heatmap of the top 50 DEGs (enlarged for readability).

(F) WGCNA soft-threshold selection with labelled powers, scale-free R^2 curve, mean connectivity, $R^2 = 0.85$ threshold and selected power = 10.

(G) Gene dendrogram and module colours.

(H) Module–trait correlations with tumour status (brown and black retained: $|r| \geq 0.5$, $P < 0.05$; 153 genes).

(I) Module membership versus gene significance in the selected modules.

Figure 2. BPA–meningioma candidate identification and PPI.

(A) BPA target source counts for CTD (≥3 PMIDs; 2,081), PharmMapper (174), SwissTargetPrediction (96), SEA (19) and ChEMBL (pChEMBL ≥ 5; 6), with five-source union 2,250; companion CTD ∩ ChEMBL Venn.

(B) Custom UpSet-style plot of BPA targets, DEGs, WGCNA module genes and meningioma database genes (bars coloured by intersection degree; set sizes at right).

(C) Venn of BPA targets ∩ DEG ∩ WGCNA (25 core candidates highlighted).

(D) Hub prioritisation bar plot: PPI rank + SHAP rank consensus (lower = better); final hubs *ACTA2*, *BMP4*, *FLNA*, *FLNB*; legend placed upper-right to avoid overlap with bottom consensus values.

(E) Compound–target–disease network (BPA–25 cores–meningioma).

(F) Circular STRING PPI layout of the 25 candidates (edge width ∝ combined score ≥ 0.4; red = final hubs).

Figure 3. Functional enrichment and PPI topology (dense 2×3 journal grid).

(A) GO Biological Process (top 10).

(B) GO Cellular Component (top 10).

(C) GO Molecular Function (all 11 significant terms; count labels kept inside bars).

(D) KEGG enrichment bar plot (14 significant pathways; adjP 0.024–0.048).

(E) KEGG enrichment dot plot (gene ratio vs −log10 adjusted P; point size = gene count).

(F) Horizontal bar chart of STRING PPI degree ranking (top 12); red bars = final hubs (*ACTA2*, *BMP4*, *FLNA*, *FLNB*).

Figure 4. Nested machine learning, SHAP interpretation and single-gene ROC.

(A) Full heatmap of all 109 tumour-versus-normal model combinations (CV, external sample-level, patient-level and mean AUC) in the left column at natural aspect ratio (no anisotropic

stretch); note ceiling (83/109 models reach external AUC = 1.0). The same heatmap is also provided as Additional file 1 (Figure 4S).

(B, C) SHAP bar and beeswarm plots for the best model (Enet+SVM, 11 genes).

(D, E) SHAP dependence and waterfall plots for the best tumour-versus-normal model.

(F) Top 25 WHO-grade models by mean AUC.

(G) SHAP bar plot for the best grade model (glmBoost+XGBoost).

(H) Single-gene ROC curves for *ACTA2*, *BMP4*, *FLNA* and *FLNB* (training and external).

Figure 4S (Additional file 1). Full heatmap of CV and external AUCs for all 109 tumour-versus-normal model combinations (readable tall layout).

Figure 5. Single-cell localisation in GSE183655 (30,000 sampled cells after QC).

(A) UMAP coloured by annotated cell type (8 types); shown alone on a full-width row with the right-side legend kept in place and the gutter widened so legend text does not overlap the plot.

(B–E) Nebulosa-style density of *ACTA2*, *BMP4*, *FLNA* and *FLNB* (panels separated for readability).

(F) Dot plot of hub-gene expression by cell type.

(G) Cell-type proportions by tissue (tumour / BTI / dura).

(H–K) UMAP feature plots of the four hubs (separated).

Figure 6. Median-split single-gene GSEA (KEGG) enrichment curves (gseaplot2) within 165 training tumours; pathway labels kept in the upper-right of each panel with collision avoidance so they do not overlap the running-enrichment curves.

(A) *ACTA2* (38 significant gene sets).

(B) *BMP4* (26).

(C) *FLNA* (34).

(D) *FLNB* (45).

Figure 7. Molecular docking of BPA (structural-computational evidence layer).

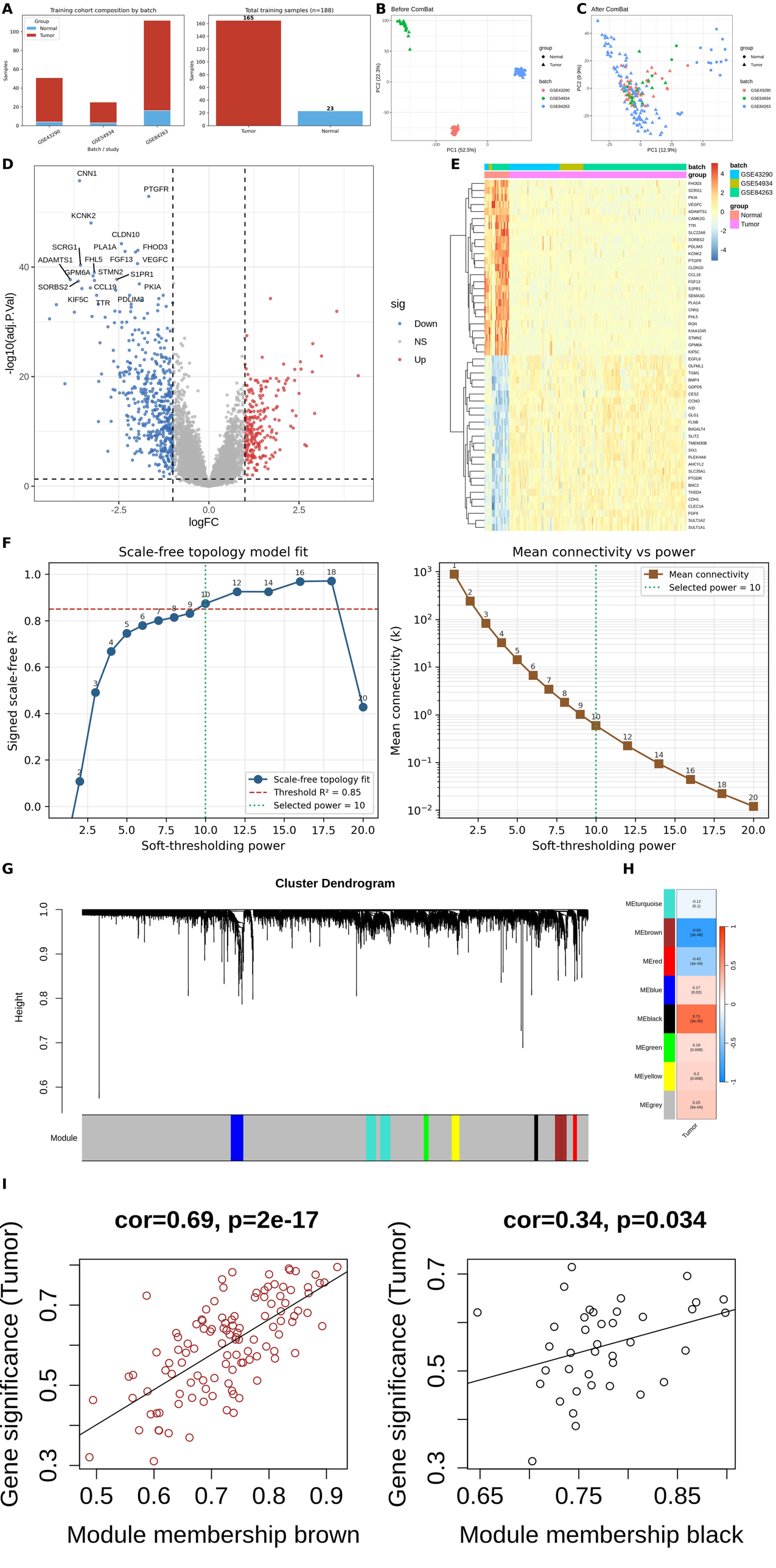
(A–C) PyMOL ray-traced 3D poses for *ESR1* (PDB 3UU7), *ESR2* (3OLS) and *PTGS2* (5KIR): protein cartoon, BPA sticks (green carbons), translucent wheat pocket surface (residues within 7 Å of ligand), key-residue sticks and yellow H-bond dashes; affinities −8.21, −8.35 and −8.49 kcal/mol; symmetry-aware redock RMSD 1.27, 0.42 and 0.89 Å.

(D–G) Weaker BPA poses at hubs *ACTA2*, *BMP4* (AlphaFold, fpocket P1) and *FLNA*/*FLNB* (PDB 3HOC/3FER): −5.70, −5.68, −5.70 and −5.17 kcal/mol.

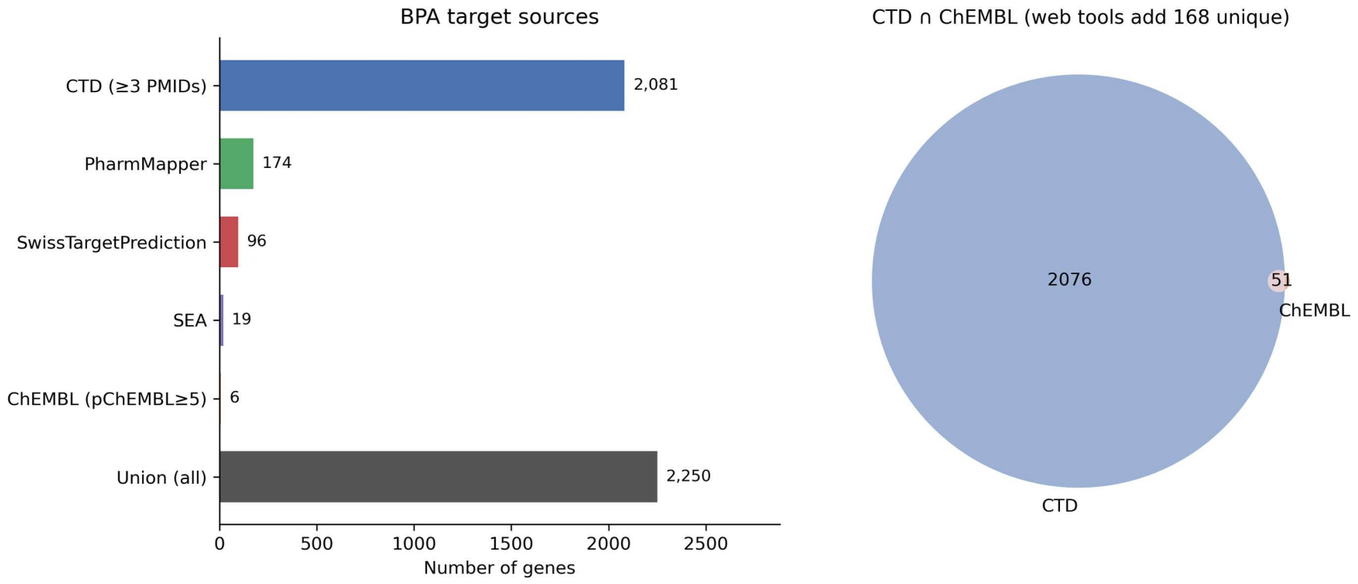
(H) Summary bar chart of BPA Vina affinities (green = positive controls; red = hubs).

Additional files

Additional file 1 (Figure 4S). Full heatmap of CV and external AUCs for all 109 tumour-versus-normal model combinations (readable tall layout; PNG/PDF).

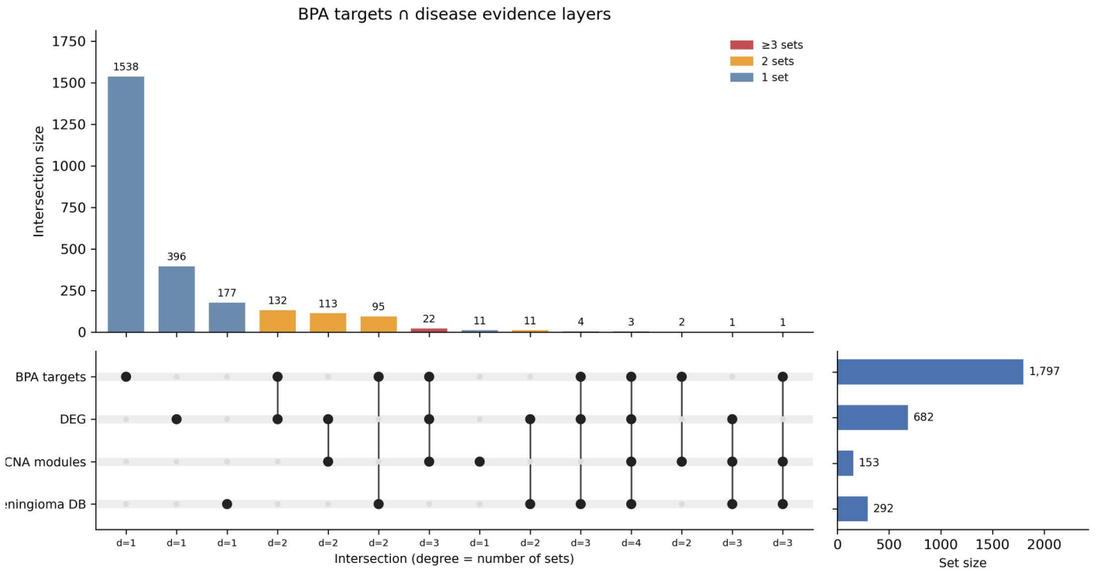

A
Training cohort composition by batch
Group
Normal
Tumor
Samples
Batch / study
GSE43290
GSE54934
GSE84263
Total training samples (n=188)
165
23
Tumor
Normal
B
Before ComBat
PC2 (22.3%)
PC1 (52.5%)
group
Normal
Tumor
batch
GSE43290
GSE54934
GSE84263
C
After ComBat
PC2 (8.9%)
PC1 (12.9%)
D
CNN1
PTGFR
KCNK2
CLDN10
SCRG1
PLA1A
FHOD3
ADAMTS1
FHL5
FGF13
VEGFC
GPM6A
STMN2
S1PR1
SORBS2
CCL19
PKIA
KIF5C
TTR
PDLIM3
-log10(adj.P.Val)
logFC
sig
Down
NS
Up
E
batch
group
FHOD3
SCRG1
PKIA
VEGFC
ADAMTS1
CAMK2G
TTR
SLC22A8
SORBS2
PDLIM3
KCNK2
PTGFR
CLDN10
CCL19
FGF13
S1PR1
SEMA3G
PLA1A
CNN1
FHL5
RGN
KIAA1045
STMN2
GPM6A
KIF5C
EGFL6
OLFML1
TGM1
BMP4
GDPD5
CES2
CCNO
IVD
GLG1
FLNB
B4GALT4
SLIT2
TMEM30B
SIX1
PLEKHA6
AHCYL2
SLC35A1
PTGDR
BNC2
THSD4
CDH1
CLEC1A
FGF9
SULT1A2
SULT1A1
F
Scale-free topology model fit
Signed scale-free R²
Soft-thresholding power
Scale-free topology fit
Threshold R² = 0.85
Selected power = 10
Mean connectivity vs power
Mean connectivity (k)
Mean connectivity
Selected power = 10
G
Cluster Dendrogram
Height
Module
H
MEturquoise
MEbrown
MEred
MEblue
MEblack
MEgreen
MEyellow
MEgrey
Tumor
I
cor=0.69, p=2e-17
Gene significance (Tumor)
Module membership brown
cor=0.34, p=0.034
Gene significance (Tumor)
Module membership black

A

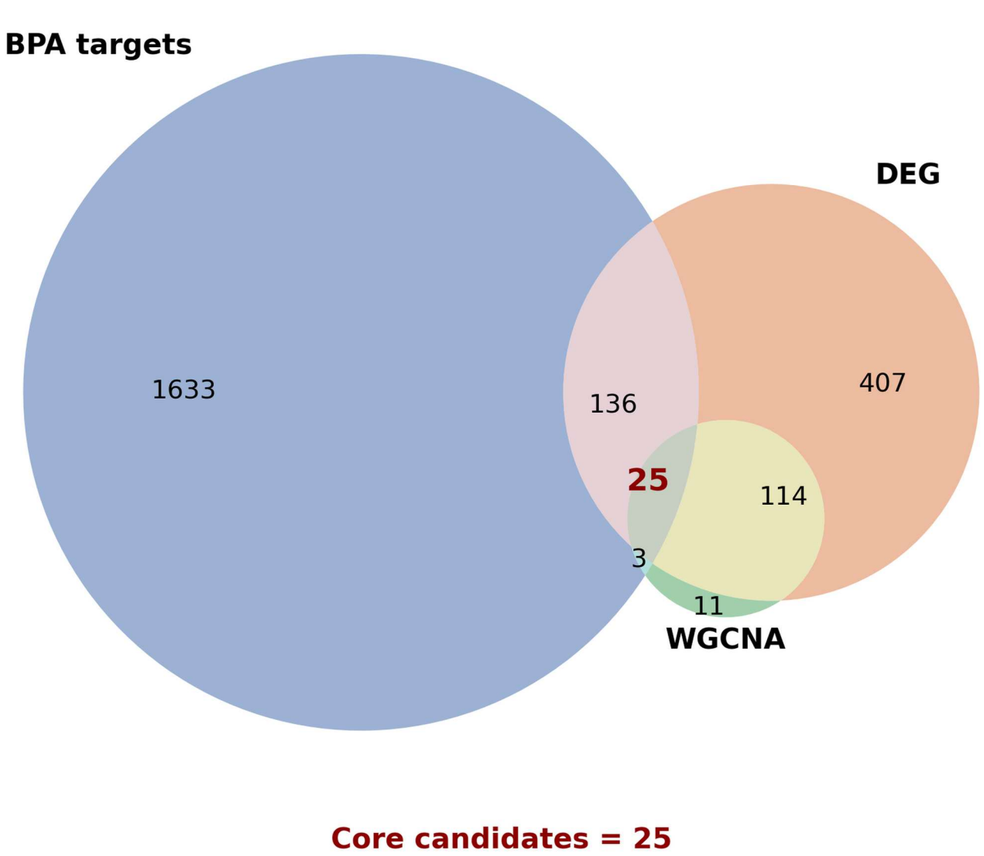

BPA target sources
CTD (≥3 PMIDs)
2,081
PharmMapper
174
SwissTargetPrediction
96
SEA
19
ChEMBL (pChEMBL≥5)
6
Union (all)
2,250
0
500
1000
1500
2000
2500
Number of genes

CTD ∩ ChEMBL (web tools add 168 unique)
2076
51
ChEMBL
CTD

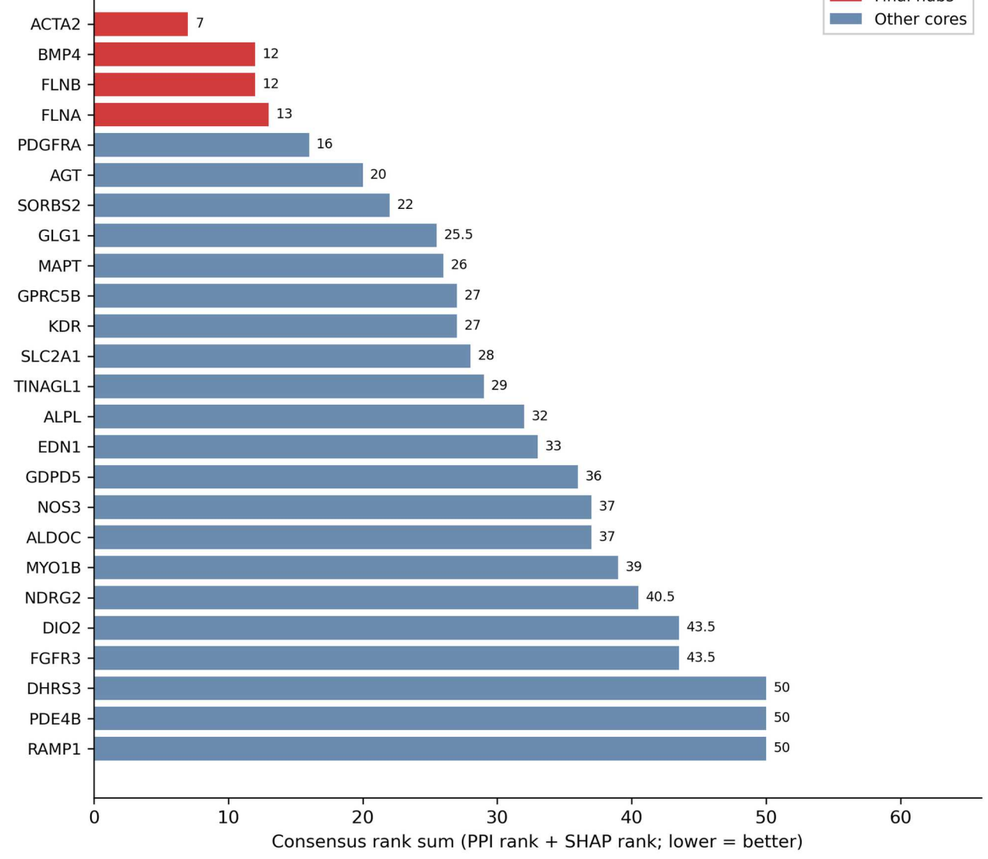
B


BPA targets ∩ disease evidence layers
≥3 sets
2 sets
1 set
Intersection size
1538
396
177
132
113
95
22
11
11
4
3
2
1
1
BPA targets
DEG
CNA modules
ningioma DB
1,797
682
153
292
Intersection (degree = number of sets)
Set size

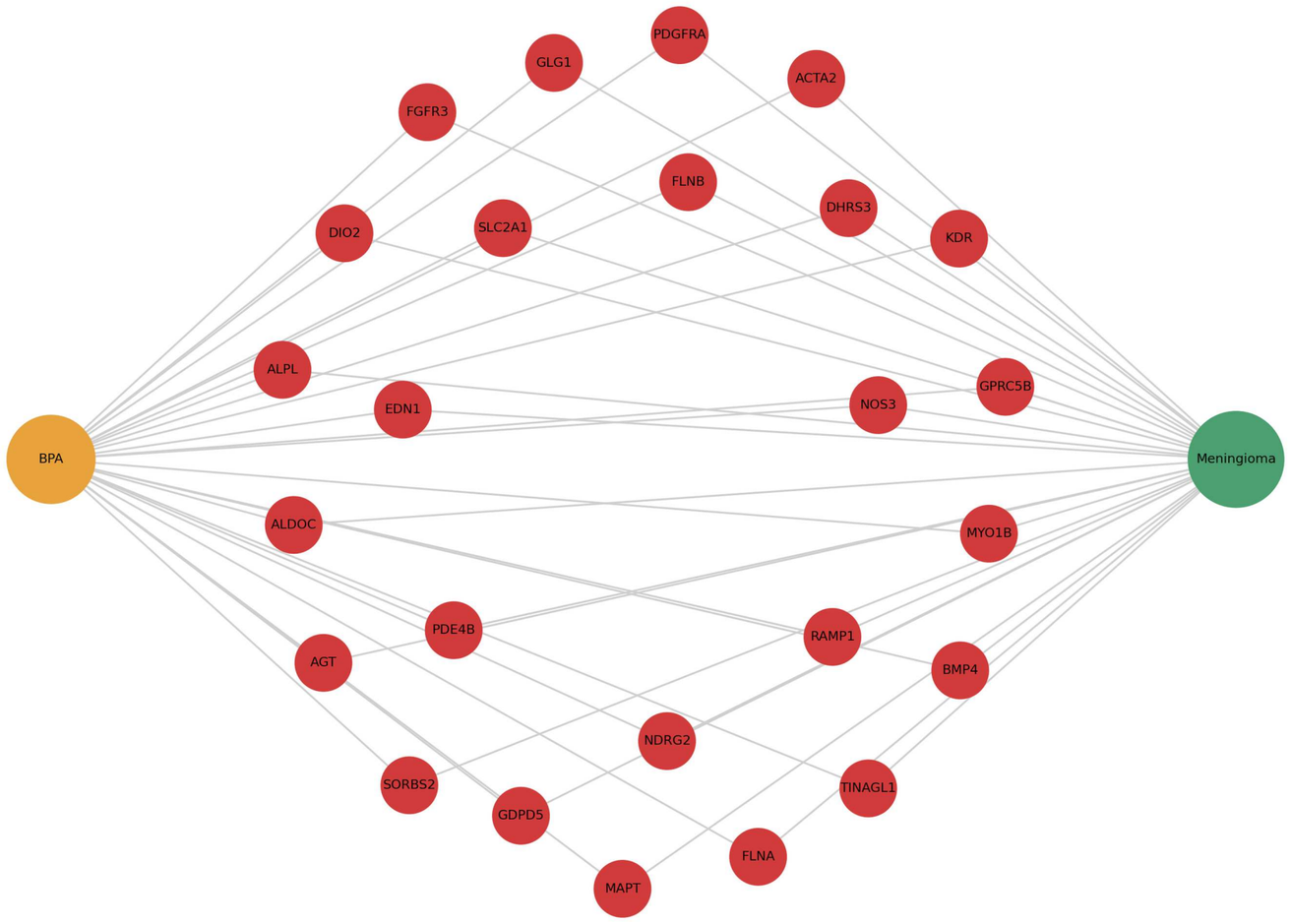
C


Core overlap (BPA ∩ DEG ∩ WGCNA)
BPA targets
DEG
WGCNA
1633
136
407
25
114
3
11
Core candidates = 25

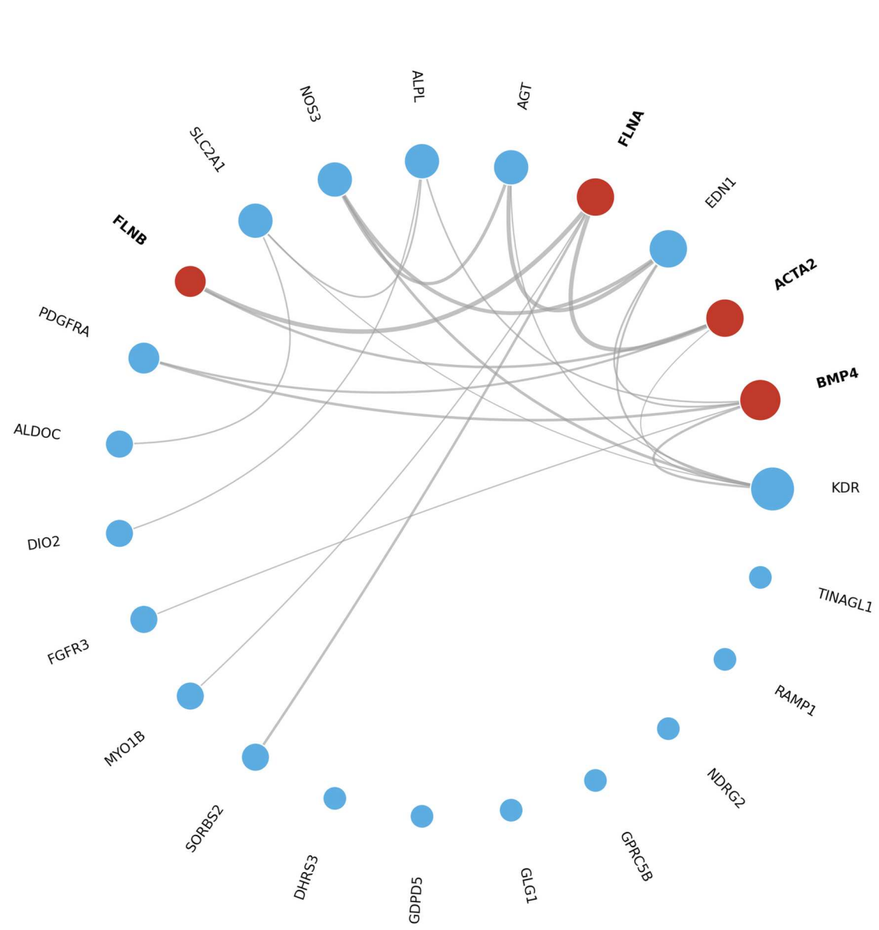
D

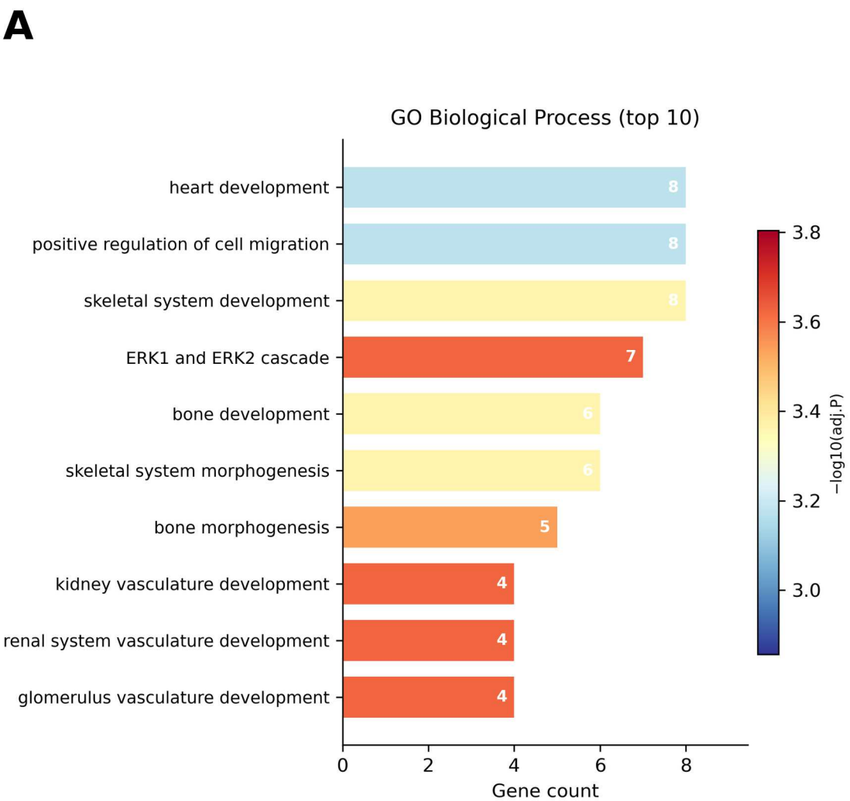

Hub prioritisation (PPI × SHAP consensus)
Final hubs
Other cores
ACTA2 7
BMP4 12
FLNB 12
FLNA 13
PDGFRA 16
AGT 20
SORBS2 22
GLG1 25.5
MAPT 26
GPRC5B 27
KDR 27
SLC2A1 28
TINAGL1 29
ALPL 32
EDN1 33
GDPD5 36
NOS3 37
ALDOC 37
MYO1B 39
NDRG2 40.5
DIO2 43.5
FGFR3 43.5
DHRS3 50
PDE4B 50
RAMP1 50
Consensus rank sum (PPI rank + SHAP rank; lower = better)

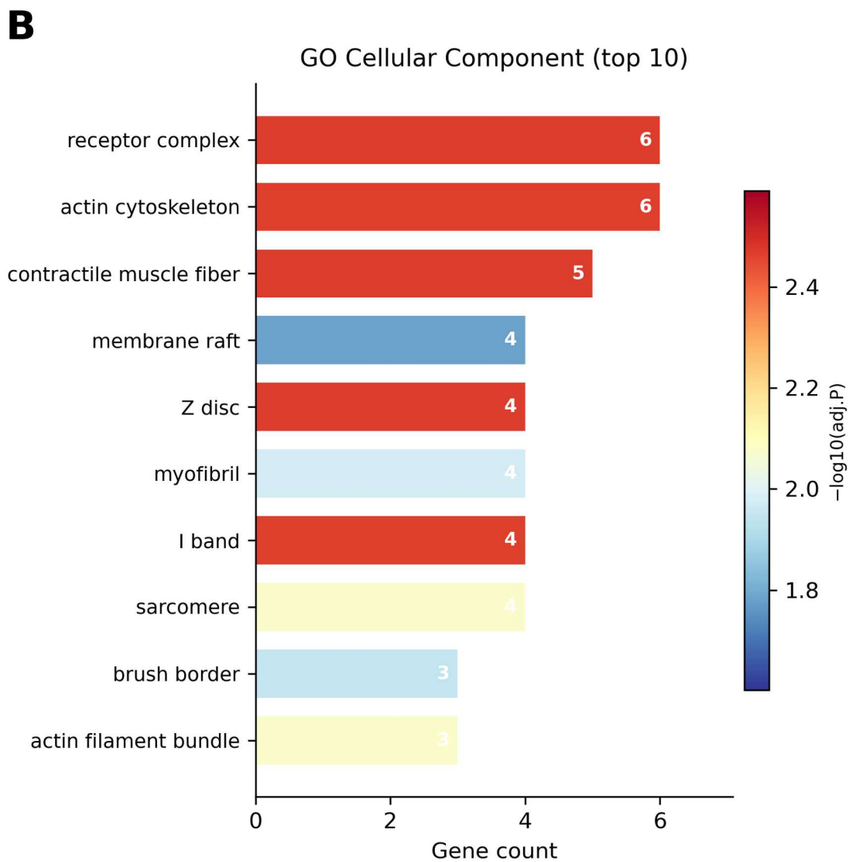
E

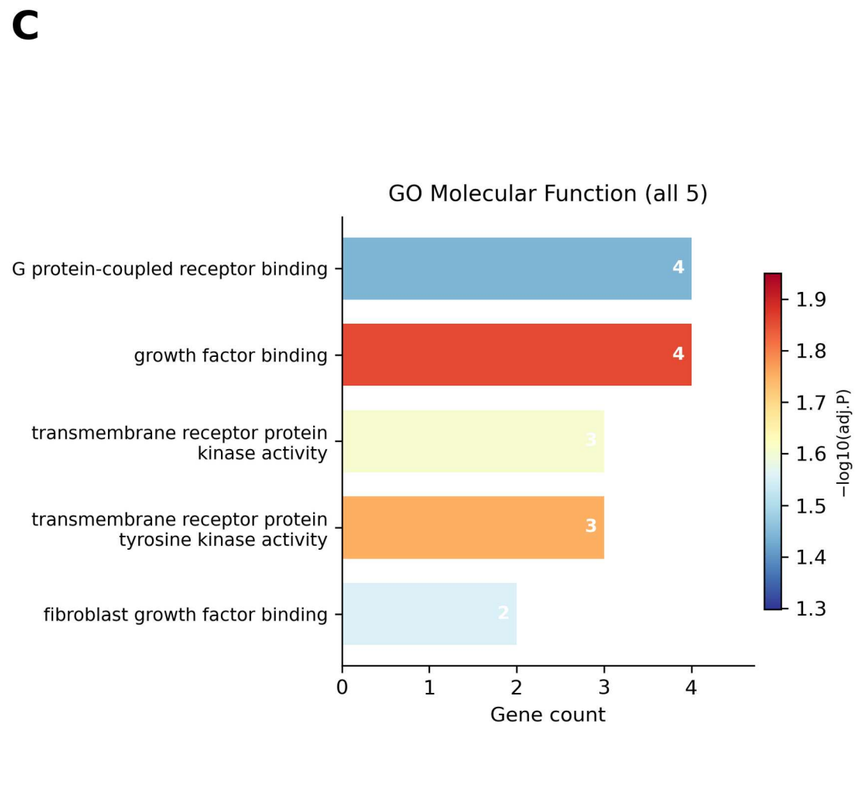

BPA
Meningioma
PDGFRA
GLG1
FGFR3
ACTA2
FLNB
DIO2
SLC2A1
DHRS3
KDR
ALPL
EDN1
NOS3
GPRC5B
ALDOC
MYO1B
PDE4B
RAMP1
AGT
BMP4
NDRG2
SORBS2
GDPD5
TINAGL1
FLNA
MAPT

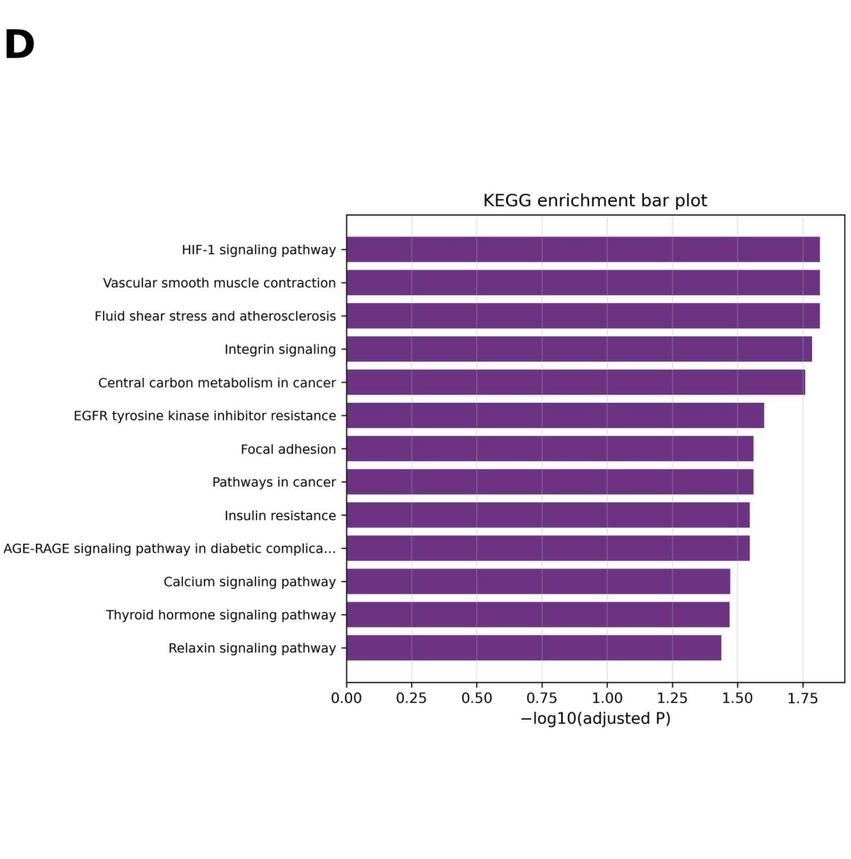
F

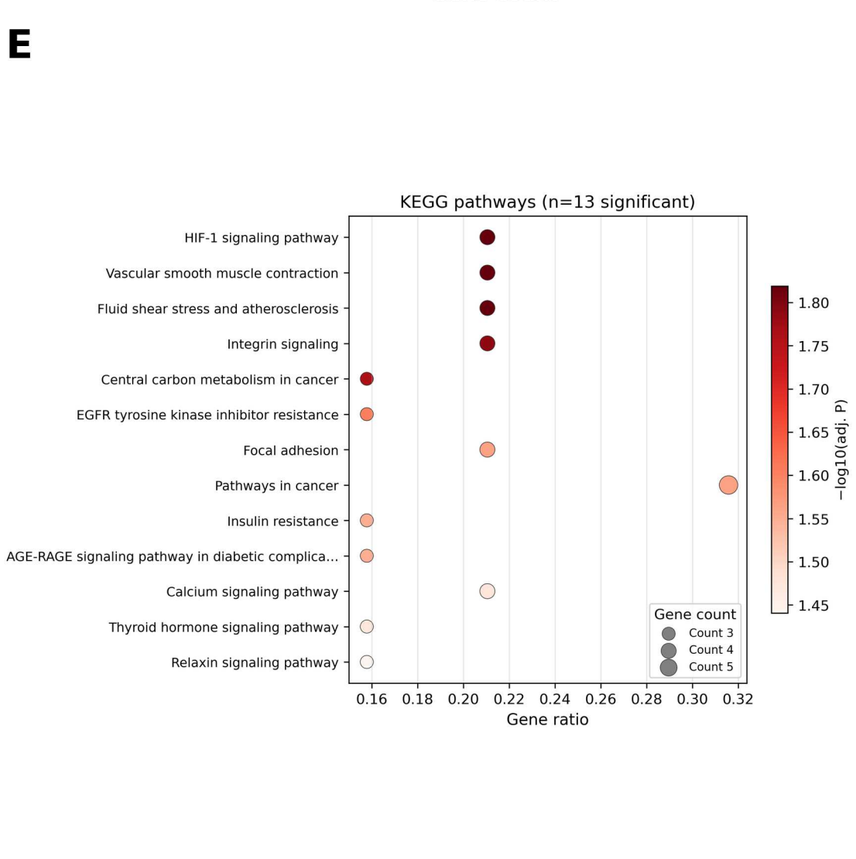

Circular PPI layout of 23 core candidates
(edge width ∝ STRING score; red = hubs)
NOS3
ALPL
AGT
FLNA
SLC2A1
EDN1
FLNB
ACTA2
PDGFRA
BMP4
ALDOC
KDR
DIO2
TINAGL1
FGFR3
RAMP1
MYO1B
NDRG2
SORBS2
DHRS3
GDPD5
GLG1
GPRC5B

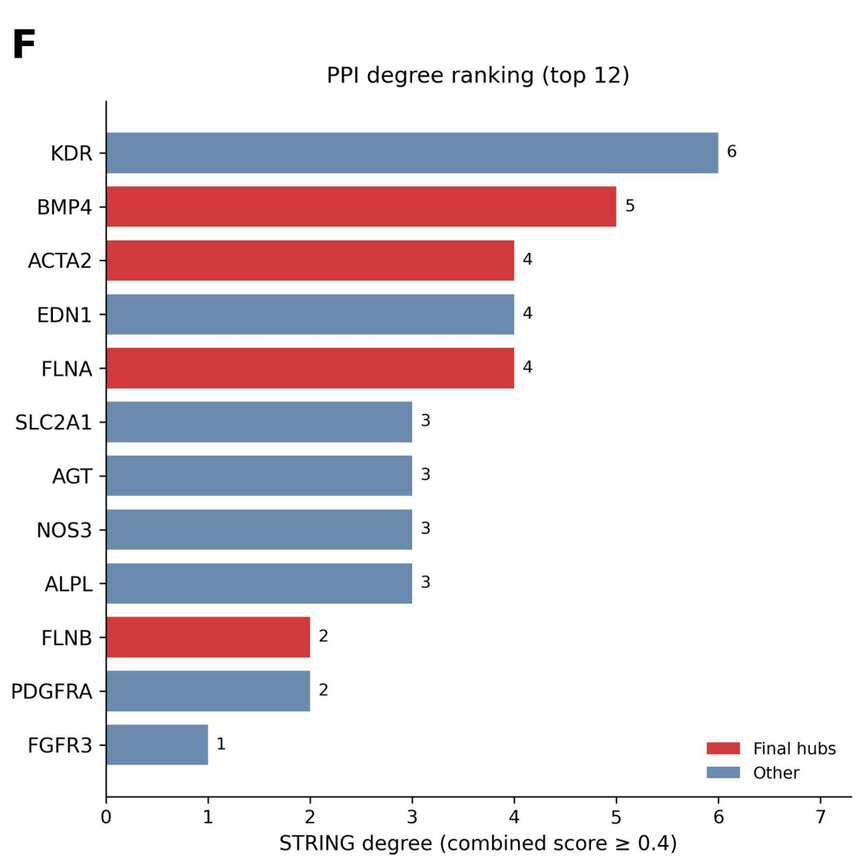

A
GO Biological Process (top 10)
heart development
positive regulation of cell migration
skeletal system development
ERK1 and ERK2 cascade
bone development
skeletal system morphogenesis
bone morphogenesis
kidney vasculature development
renal system vasculature development
glomerulus vasculature development
8
8
8
7
6
6
5
4
4
4
0
2
4
6
8
Gene count
3.8
3.6
3.4
3.2
3.0
−log10(adj.P)


B
GO Cellular Component (top 10)
receptor complex
actin cytoskeleton
contractile muscle fiber
membrane raft
Z disc
myofibril
I band
sarcomere
brush border
actin filament bundle
6
6
5
4
4
4
4
4
3
3
0
2
4
6
Gene count
2.4
2.2
2.0
1.8
−log10(adj.P)

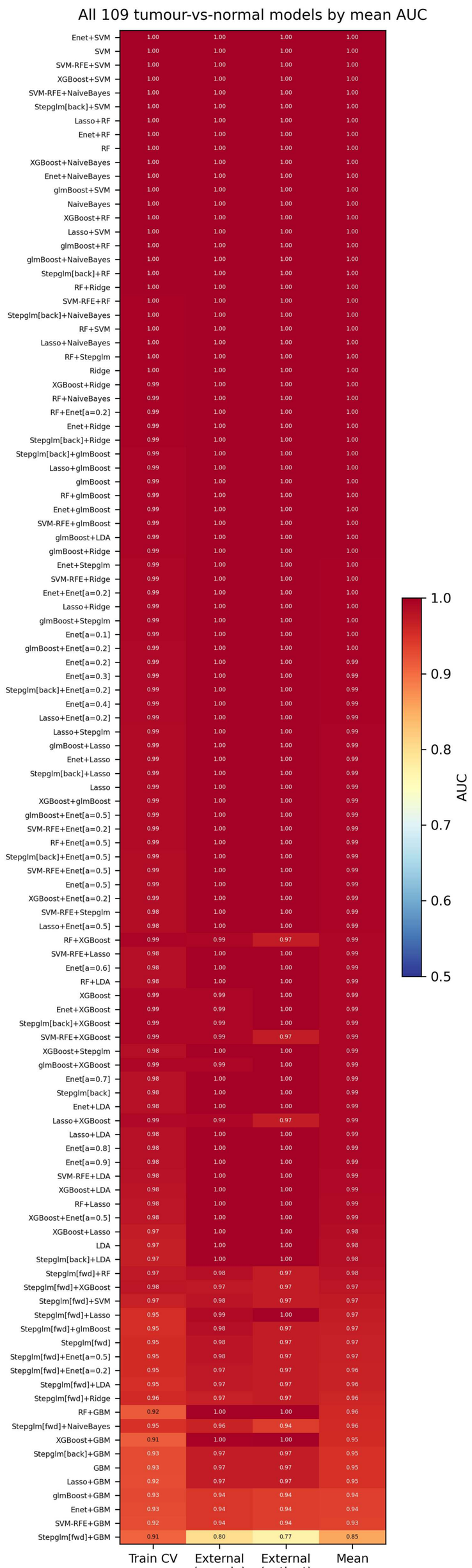

C
GO Molecular Function (all 5)
G protein-coupled receptor binding
growth factor binding
transmembrane receptor protein kinase activity
transmembrane receptor protein tyrosine kinase activity
fibroblast growth factor binding
4
4
3
3
2
0
1
2
3
4
Gene count
1.9
1.8
1.7
1.6
1.5
1.4
1.3
−log10(adj.P)


D
KEGG enrichment bar plot
HIF-1 signaling pathway
Vascular smooth muscle contraction
Fluid shear stress and atherosclerosis
Integrin signaling
Central carbon metabolism in cancer
EGFR tyrosine kinase inhibitor resistance
Focal adhesion
Pathways in cancer
Insulin resistance
AGE-RAGE signaling pathway in diabetic complica...
Calcium signaling pathway
Thyroid hormone signaling pathway
Relaxin signaling pathway
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
−log10(adjusted P)

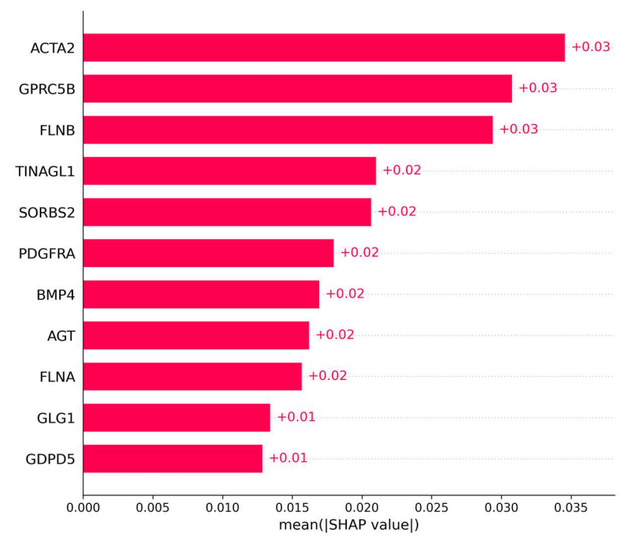

E
KEGG pathways (n=13 significant)
HIF-1 signaling pathway
Vascular smooth muscle contraction
Fluid shear stress and atherosclerosis
Integrin signaling
Central carbon metabolism in cancer
EGFR tyrosine kinase inhibitor resistance
Focal adhesion
Pathways in cancer
Insulin resistance
AGE-RAGE signaling pathway in diabetic complica...
Calcium signaling pathway
Thyroid hormone signaling pathway
Relaxin signaling pathway
Gene count
Count 3
Count 4
Count 5
0.16
0.18
0.20
0.22
0.24
0.26
0.28
0.30
0.32
Gene ratio
1.80
1.75
1.70
1.65
1.60
1.55
1.50
1.45
−log10(adj. P)


F
PPI degree ranking (top 12)
KDR
BMP4
ACTA2
EDN1
FLNA
SLC2A1
AGT
NOS3
ALPL
FLNB
PDGFRA
FGFR3
6
5
4
4
4
3
3
3
3
2
2
1
Final hubs
Other
0
1
2
3
4
5
6
7
STRING degree (combined score ≥ 0.4)

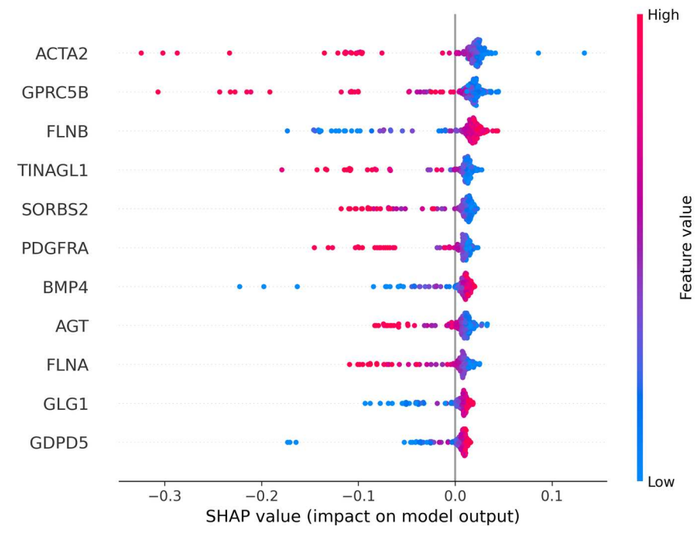
A


All 109 tumour-vs-normal models by mean AUC
Enet+SVM 1.00 1.00 1.00 1.00
SVM 1.00 1.00 1.00 1.00
SVM-RFE+SVM 1.00 1.00 1.00 1.00
XGBoost+SVM 1.00 1.00 1.00 1.00
SVM-RFE+NaiveBayes 1.00 1.00 1.00 1.00
Stepglm[back]+SVM 1.00 1.00 1.00 1.00
Lasso+RF 1.00 1.00 1.00 1.00
Enet+RF 1.00 1.00 1.00 1.00
RF 1.00 1.00 1.00 1.00
XGBoost+NaiveBayes 1.00 1.00 1.00 1.00
Enet+NaiveBayes 1.00 1.00 1.00 1.00
glmBoost+SVM 1.00 1.00 1.00 1.00
NaiveBayes 1.00 1.00 1.00 1.00
XGBoost+RF 1.00 1.00 1.00 1.00
Lasso+SVM 1.00 1.00 1.00 1.00
glmBoost+RF 1.00 1.00 1.00 1.00
glmBoost+NaiveBayes 1.00 1.00 1.00 1.00
Stepglm[back]+RF 1.00 1.00 1.00 1.00
RF+Ridge 1.00 1.00 1.00 1.00
SVM-RFE+RF 1.00 1.00 1.00 1.00
Stepglm[back]+NaiveBayes 1.00 1.00 1.00 1.00
RF+SVM 1.00 1.00 1.00 1.00
Lasso+NaiveBayes 1.00 1.00 1.00 1.00
RF+Stepglm 1.00 1.00 1.00 1.00
Ridge 1.00 1.00 1.00 1.00
XGBoost+Ridge 0.99 1.00 1.00 1.00
RF+NaiveBayes 0.99 1.00 1.00 1.00
RF+Enet[a=0.2] 0.99 1.00 1.00 1.00
Enet+Ridge 0.99 1.00 1.00 1.00
Stepglm[back]+Ridge 0.99 1.00 1.00 1.00
Stepglm[back]+glmBoost 0.99 1.00 1.00 1.00
Lasso+glmBoost 0.99 1.00 1.00 1.00
glmBoost 0.99 1.00 1.00 1.00
RF+glmBoost 0.99 1.00 1.00 1.00
Enet+glmBoost 0.99 1.00 1.00 1.00
SVM-RFE+glmBoost 0.99 1.00 1.00 1.00
glmBoost+LDA 0.99 1.00 1.00 1.00
glmBoost+Ridge 0.99 1.00 1.00 1.00
Enet+Stepglm 0.99 1.00 1.00 1.00
SVM-RFE+Ridge 0.99 1.00 1.00 1.00
Enet+Enet[a=0.2] 0.99 1.00 1.00 1.00
Lasso+Ridge 0.99 1.00 1.00 1.00
glmBoost+Stepglm 0.99 1.00 1.00 1.00
Enet[a=0.1] 0.99 1.00 1.00 1.00
glmBoost+Enet[a=0.2] 0.99 1.00 1.00 1.00
Enet[a=0.2] 0.99 1.00 1.00 0.99
Enet[a=0.3] 0.99 1.00 1.00 0.99
Stepglm[back]+Enet[a=0.2] 0.99 1.00 1.00 0.99
Enet[a=0.4] 0.99 1.00 1.00 0.99
Lasso+Enet[a=0.2] 0.99 1.00 1.00 0.99
Lasso+Stepglm 0.99 1.00 1.00 0.99
glmBoost+Lasso 0.99 1.00 1.00 0.99
Enet+Lasso 0.99 1.00 1.00 0.99
Stepglm[back]+Lasso 0.99 1.00 1.00 0.99
Lasso 0.99 1.00 1.00 0.99
XGBoost+glmBoost 0.99 1.00 1.00 0.99
glmBoost+Enet[a=0.5] 0.99 1.00 1.00 0.99
SVM-RFE+Enet[a=0.2] 0.99 1.00 1.00 0.99
RF+Enet[a=0.5] 0.99 1.00 1.00 0.99
Stepglm[back]+Enet[a=0.5] 0.99 1.00 1.00 0.99
SVM-RFE+Enet[a=0.5] 0.99 1.00 1.00 0.99
Enet[a=0.5] 0.99 1.00 1.00 0.99
XGBoost+Enet[a=0.2] 0.99 1.00 1.00 0.99
SVM-RFE+Stepglm 0.98 1.00 1.00 0.99
Lasso+Enet[a=0.5] 0.98 1.00 1.00 0.99
RF+XGBoost 0.99 0.99 0.97 0.99
SVM-RFE+Lasso 0.98 1.00 1.00 0.99
Enet[a=0.6] 0.98 1.00 1.00 0.99
RF+LDA 0.98 1.00 1.00 0.99
XGBoost 0.99 0.99 1.00 0.99
Enet+XGBoost 0.99 0.99 1.00 0.99
Stepglm[back]+XGBoost 0.99 0.99 1.00 0.99
SVM-RFE+XGBoost 0.99 0.99 0.97 0.99
XGBoost+Stepglm 0.98 1.00 1.00 0.99
glmBoost+XGBoost 0.99 0.99 1.00 0.99
Enet[a=0.7] 0.98 1.00 1.00 0.99
Stepglm[back] 0.98 1.00 1.00 0.99
Enet+LDA 0.98 1.00 1.00 0.99
Lasso+XGBoost 0.99 0.99 0.97 0.99
Lasso+LDA 0.98 1.00 1.00 0.99
Enet[a=0.8] 0.98 1.00 1.00 0.99
Enet[a=0.9] 0.98 1.00 1.00 0.99
SVM-RFE+LDA 0.98 1.00 1.00 0.99
XGBoost+LDA 0.98 1.00 1.00 0.99
RF+Lasso 0.98 1.00 1.00 0.99
XGBoost+Enet[a=0.5] 0.98 1.00 1.00 0.99
XGBoost+Lasso 0.97 1.00 1.00 0.98
LDA 0.97 1.00 1.00 0.98
Stepglm[back]+LDA 0.97 1.00 1.00 0.98
Stepglm[fwd]+RF 0.97 0.98 0.97 0.98
Stepglm[fwd]+XGBoost 0.98 0.97 0.97 0.97
Stepglm[fwd]+SVM 0.97 0.98 0.97 0.97
Stepglm[fwd]+Lasso 0.95 0.99 1.00 0.97
Stepglm[fwd]+glmBoost 0.95 0.98 0.97 0.97
Stepglm[fwd] 0.95 0.98 0.97 0.97
Stepglm[fwd]+Enet[a=0.5] 0.95 0.98 0.97 0.97
Stepglm[fwd]+Enet[a=0.2] 0.95 0.97 0.97 0.96
Stepglm[fwd]+LDA 0.95 0.97 0.97 0.96
Stepglm[fwd]+Ridge 0.96 0.97 0.97 0.96
RF+GBM 0.92 1.00 1.00 0.96
Stepglm[fwd]+NaiveBayes 0.95 0.96 0.94 0.96
XGBoost+GBM 0.91 1.00 1.00 0.95
Stepglm[back]+GBM 0.93 0.97 0.97 0.95
GBM 0.93 0.97 0.97 0.95
Lasso+GBM 0.92 0.97 0.97 0.95
glmBoost+GBM 0.93 0.94 0.94 0.94
Enet+GBM 0.93 0.94 0.94 0.94
SVM-RFE+GBM 0.92 0.94 0.94 0.93
Stepglm[fwd]+GBM 0.91 0.80 0.77 0.85
Train CV
External (sample)
External (patient)
Mean
AUC
1.0
0.9
0.8
0.7
0.6
0.5

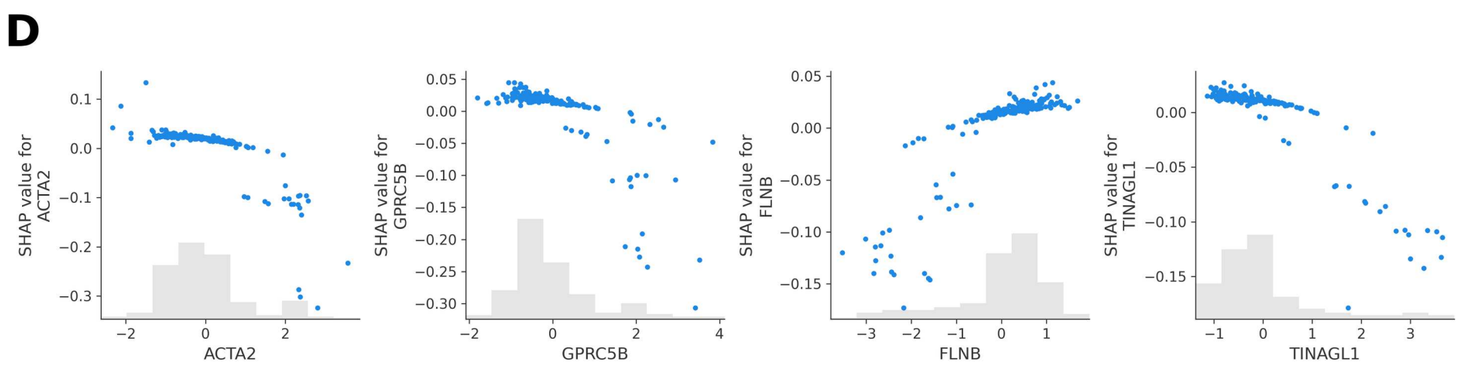
B


ACTA2 +0.03
GPRC5B +0.03
FLNB +0.03
TINAGL1 +0.02
SORBS2 +0.02
PDGFRA +0.02
BMP4 +0.02
AGT +0.02
FLNA +0.02
GLG1 +0.01
GDPD5 +0.01
0.000 0.005 0.010 0.015 0.020 0.025 0.030 0.035
mean(|SHAP value|)

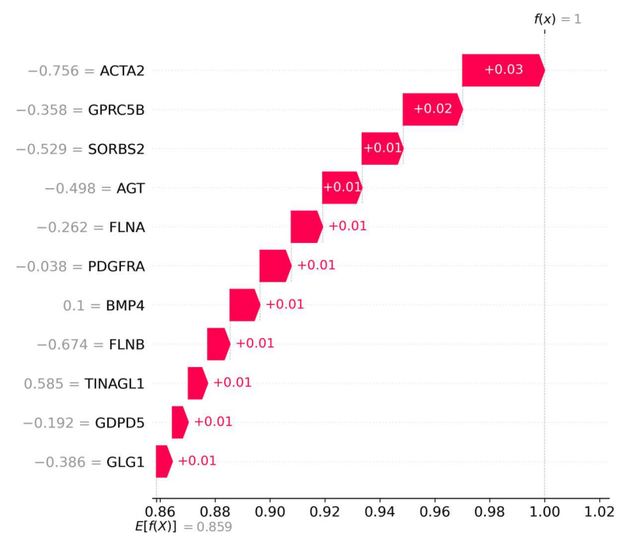
C


ACTA2
GPRC5B
FLNB
TINAGL1
SORBS2
PDGFRA
BMP4
AGT
FLNA
GLG1
GDPD5
−0.3 −0.2 −0.1 0.0 0.1
SHAP value (impact on model output)
High
Feature value
Low

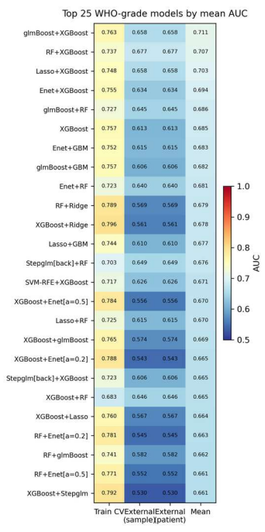
D


SHAP value for ACTA2
ACTA2
SHAP value for GPRC5B
GPRC5B
SHAP value for FLNB
FLNB
SHAP value for TINAGL1
TINAGL1

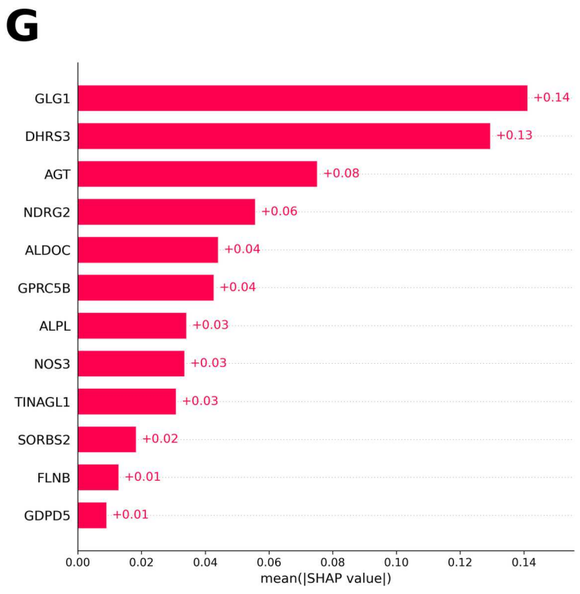
E


f(x) = 1
−0.756 = ACTA2 +0.03
−0.358 = GPRC5B +0.02
−0.529 = SORBS2 +0.01
−0.498 = AGT +0.01
−0.262 = FLNA +0.01
−0.038 = PDGFRA +0.01
0.1 = BMP4 +0.01
−0.674 = FLNB +0.01
0.585 = TINAGL1 +0.01
−0.192 = GDPD5 +0.01
−0.386 = GLG1 +0.01
0.86 0.88 0.90 0.92 0.94 0.96 0.98 1.00 1.02
E[f(X)] = 0.859

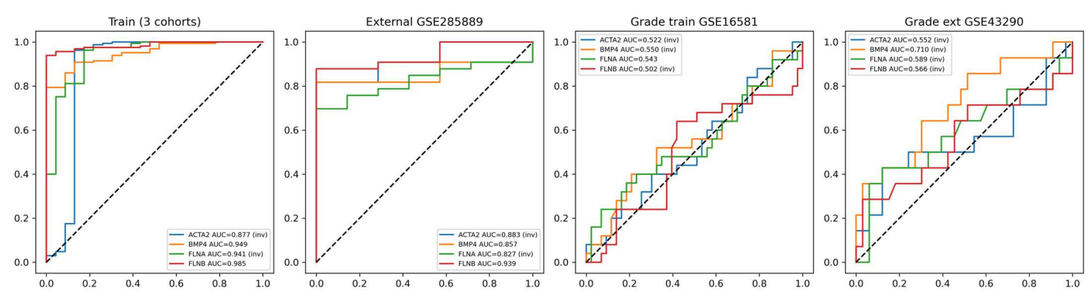
F


Top 25 WHO-grade models by mean AUC
glmBoost+XGBoost 0.763 0.658 0.658 0.711
RF+XGBoost 0.737 0.677 0.677 0.707
Lasso+XGBoost 0.748 0.658 0.658 0.703
Enet+XGBoost 0.755 0.634 0.634 0.694
glmBoost+RF 0.727 0.645 0.645 0.686
XGBoost 0.757 0.613 0.613 0.685
Enet+GBM 0.752 0.615 0.615 0.683
glmBoost+GBM 0.757 0.606 0.606 0.682
Enet+RF 0.723 0.640 0.640 0.681
RF+Ridge 0.789 0.569 0.569 0.679
XGBoost+Ridge 0.796 0.561 0.561 0.678
Lasso+GBM 0.744 0.610 0.610 0.677
Stepglm[back]+RF 0.703 0.649 0.649 0.676
SVM-RFE+XGBoost 0.717 0.626 0.626 0.671
XGBoost+Enet[a=0.5] 0.784 0.556 0.556 0.670
Lasso+RF 0.725 0.615 0.615 0.670
XGBoost+glmBoost 0.765 0.574 0.574 0.669
XGBoost+Enet[a=0.2] 0.788 0.543 0.543 0.665
Stepglm[back]+XGBoost 0.723 0.606 0.606 0.665
XGBoost+RF 0.683 0.646 0.646 0.665
XGBoost+Lasso 0.760 0.567 0.567 0.664
RF+Enet[a=0.2] 0.781 0.545 0.545 0.663
RF+glmBoost 0.741 0.582 0.582 0.662
RF+Enet[a=0.5] 0.771 0.552 0.552 0.661
XGBoost+Stepglm 0.792 0.530 0.530 0.661
Train CV
External (sample)
External (patient)
Mean
AUC
1.0
0.9
0.8
0.7
0.6
0.5

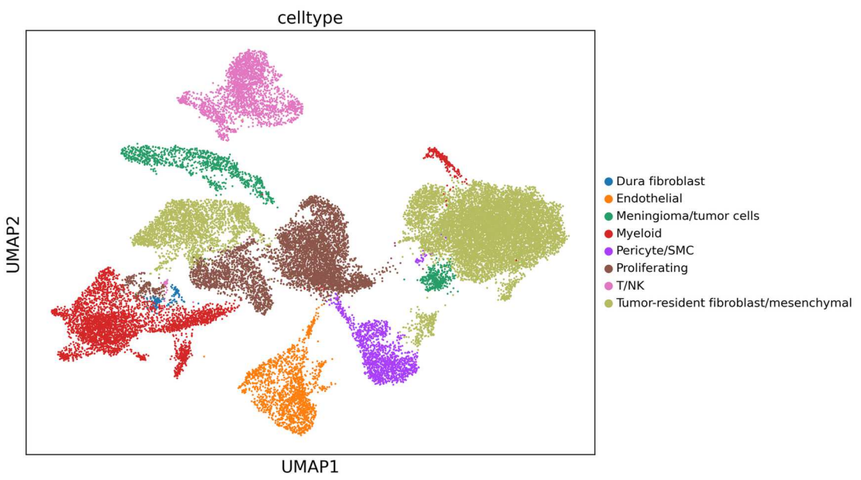
G

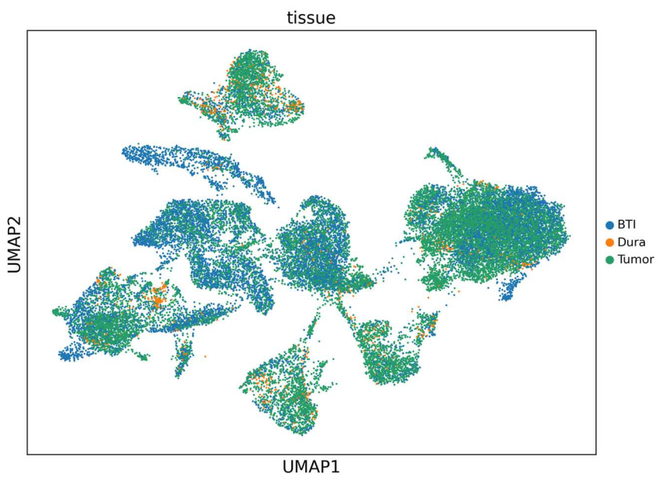

GLG1 +0.14
DHRS3 +0.13
AGT +0.08
NDRG2 +0.06
ALDOC +0.04
GPRC5B +0.04
ALPL +0.03
NOS3 +0.03
TINAGL1 +0.03
SORBS2 +0.02
FLNB +0.01
GDPD5 +0.01
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
mean(|SHAP value|)

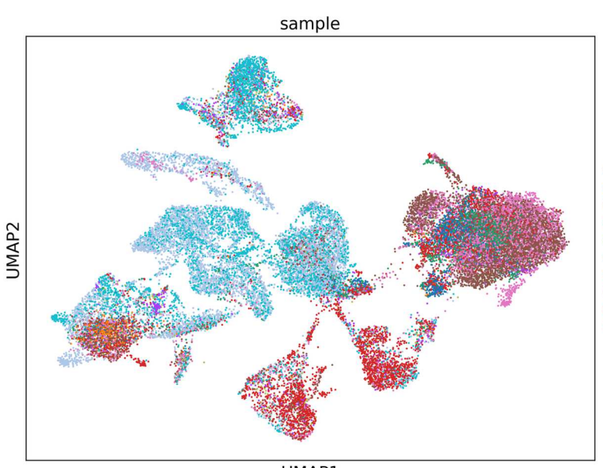
H

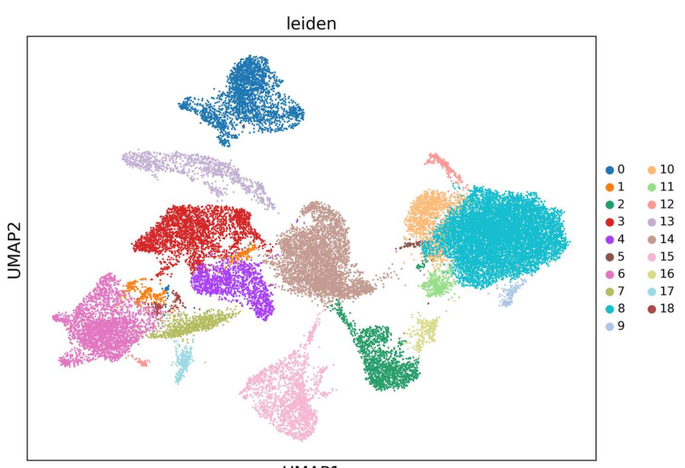

Train (3 cohorts)
ACTA2 AUC=0.877 (inv)
BMP4 AUC=0.949
FLNA AUC=0.941 (inv)
FLNB AUC=0.985
External GSE285889
ACTA2 AUC=0.883 (inv)
BMP4 AUC=0.857
FLNA AUC=0.827 (inv)
FLNB AUC=0.939
Grade train GSE16581
ACTA2 AUC=0.522 (inv)
BMP4 AUC=0.550 (inv)
FLNA AUC=0.543
FLNB AUC=0.502 (inv)
Grade ext GSE43290
ACTA2 AUC=0.552 (inv)
BMP4 AUC=0.710 (inv)
FLNA AUC=0.589 (inv)
FLNB AUC=0.566 (inv)

A





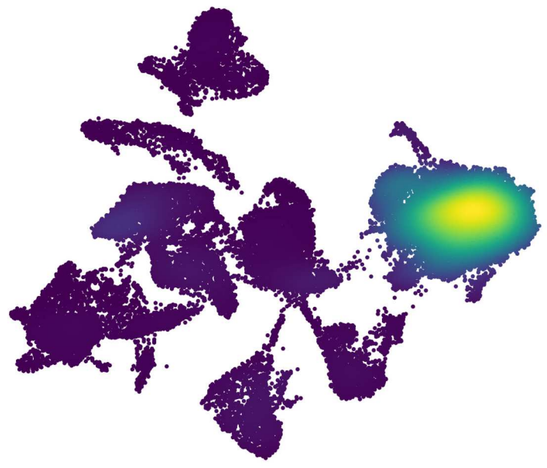




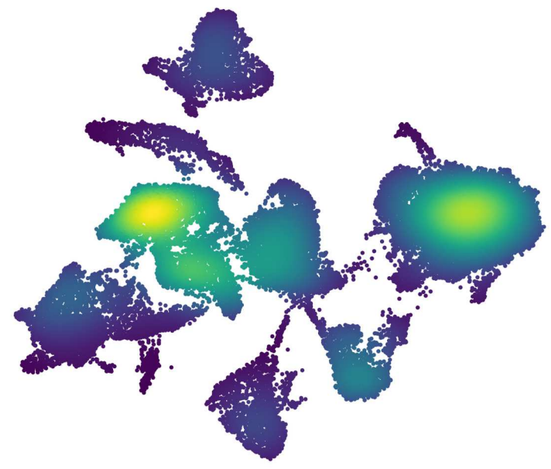
B



C



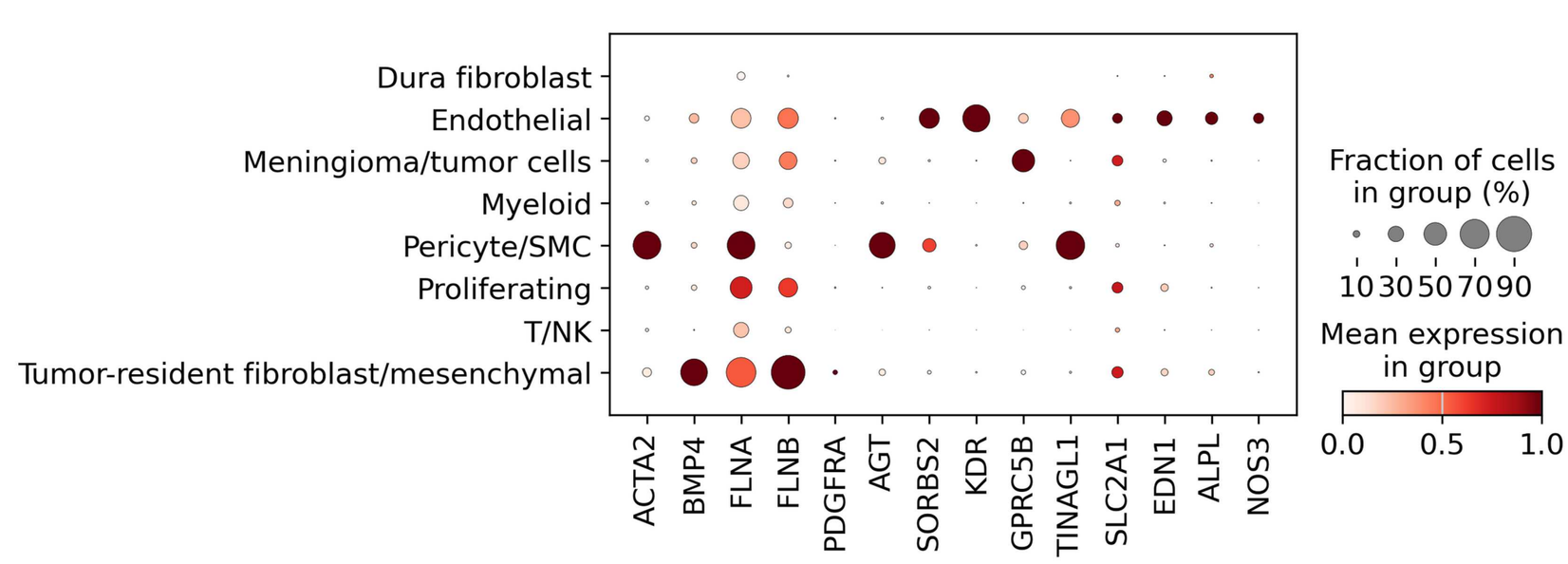
D



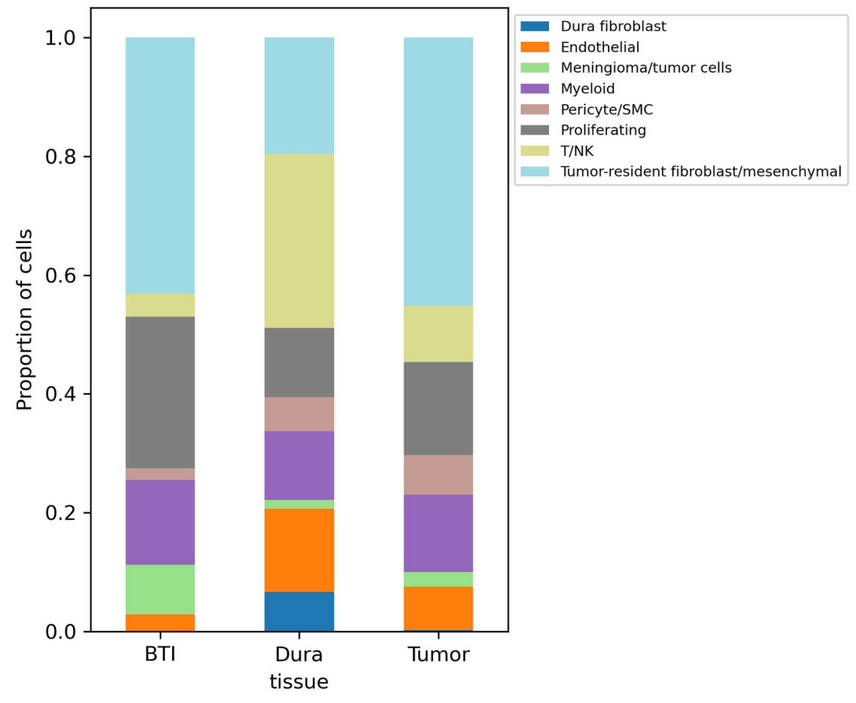
E



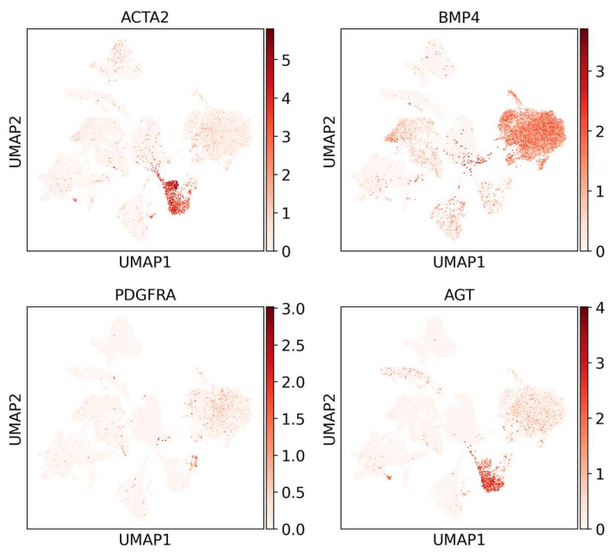
F



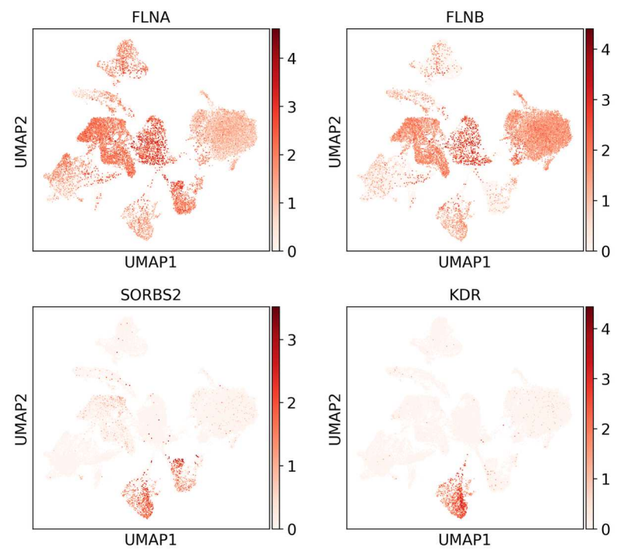
G



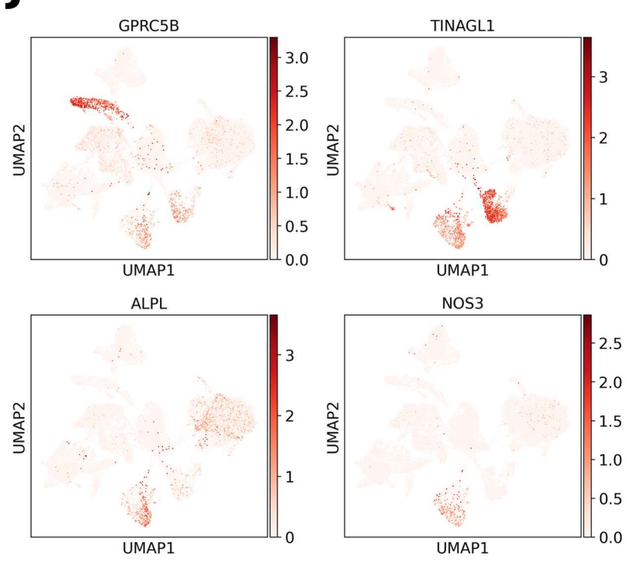
H



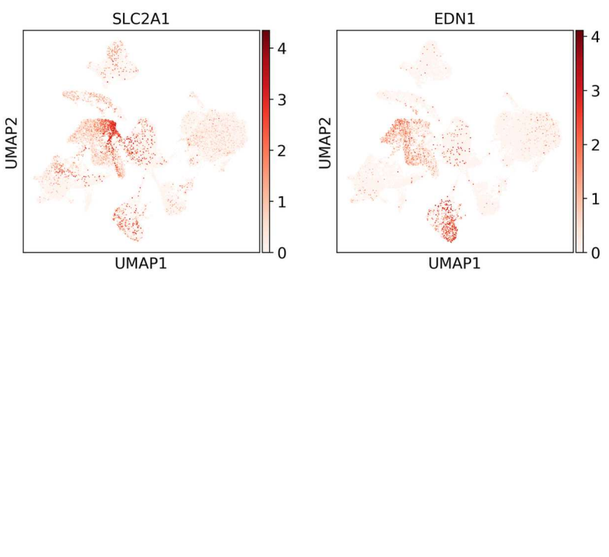
I



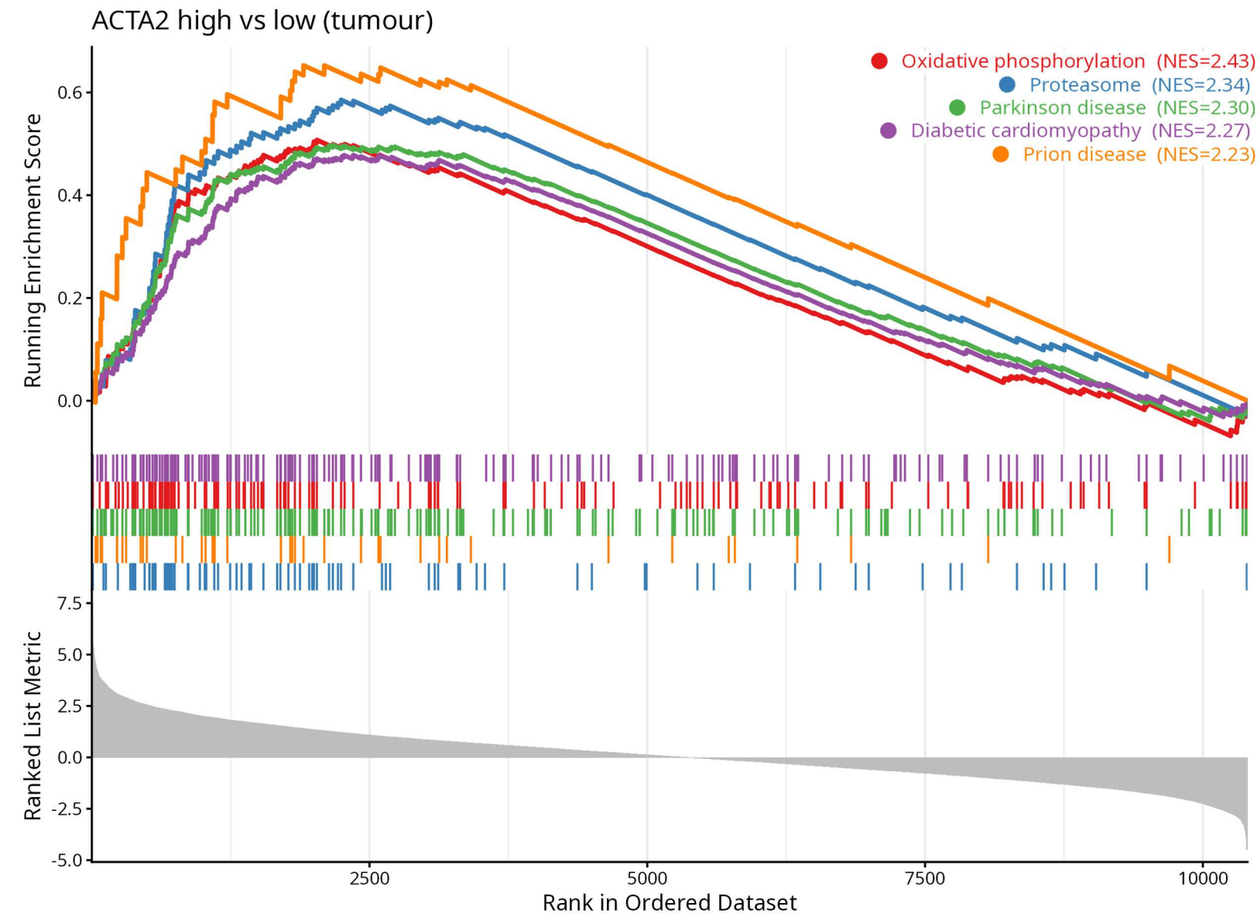
J



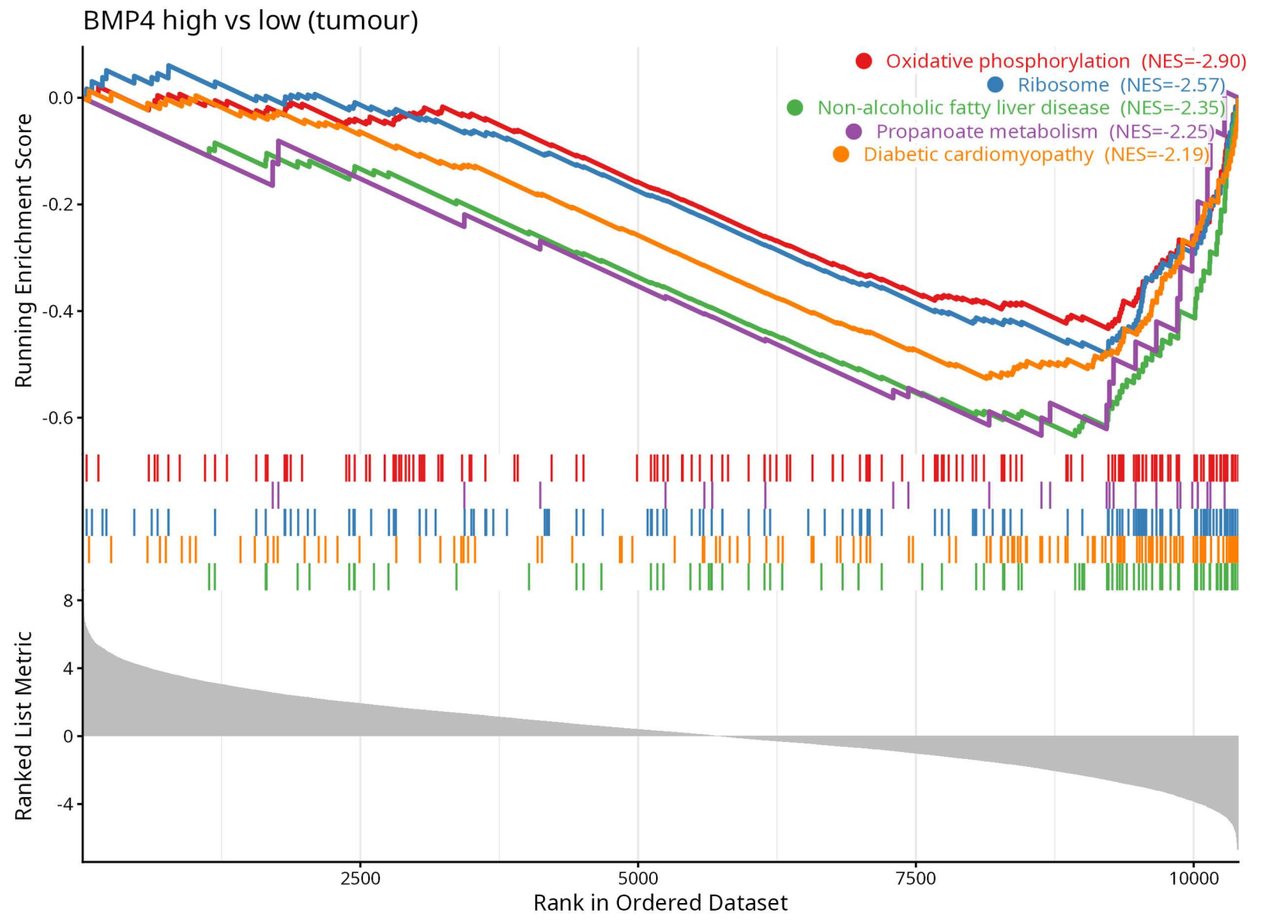
K



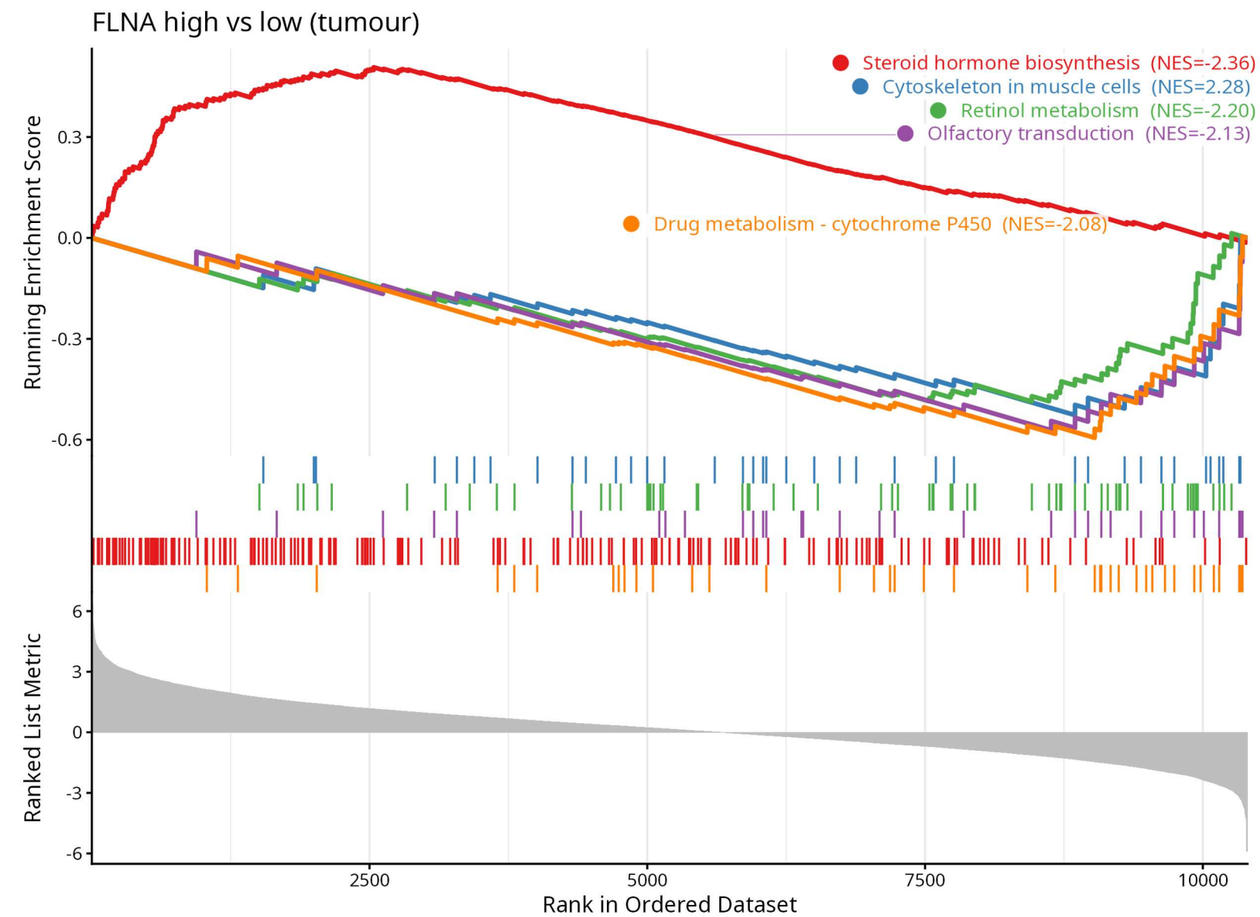
A


ACTA2 high vs low (tumour)
Oxidative phosphorylation (NES=2.43)
Proteasome (NES=2.34)
Parkinson disease (NES=2.30)
Diabetic cardiomyopathy (NES=2.27)
Prion disease (NES=2.23)
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset

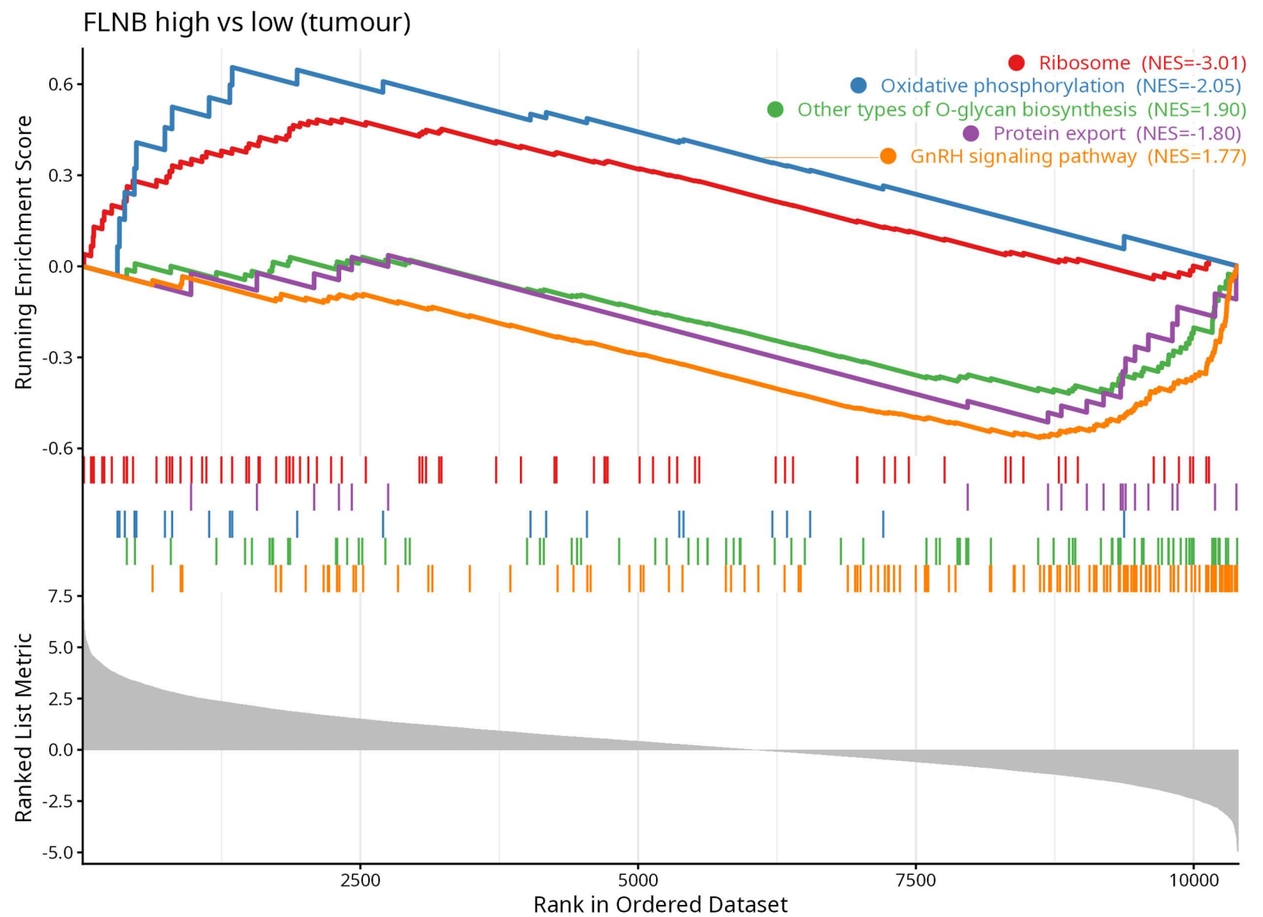
B


BMP4 high vs low (tumour)
Oxidative phosphorylation (NES=-2.90)
Ribosome (NES=-2.57)
Non-alcoholic fatty liver disease (NES=-2.35)
Propanoate metabolism (NES=-2.25)
Diabetic cardiomyopathy (NES=-2.19)
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset

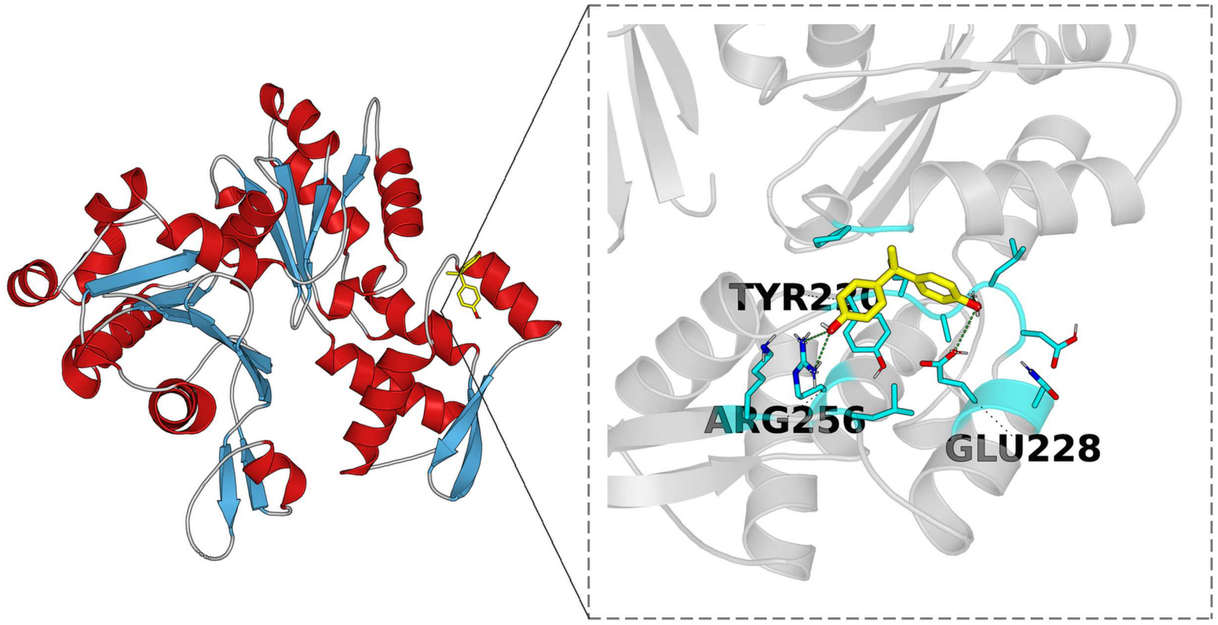
C


FLNA high vs low (tumour)
Steroid hormone biosynthesis (NES=-2.36)
Cytoskeleton in muscle cells (NES=2.28)
Retinol metabolism (NES=-2.20)
Olfactory transduction (NES=-2.13)
Drug metabolism - cytochrome P450 (NES=-2.08)
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset

D

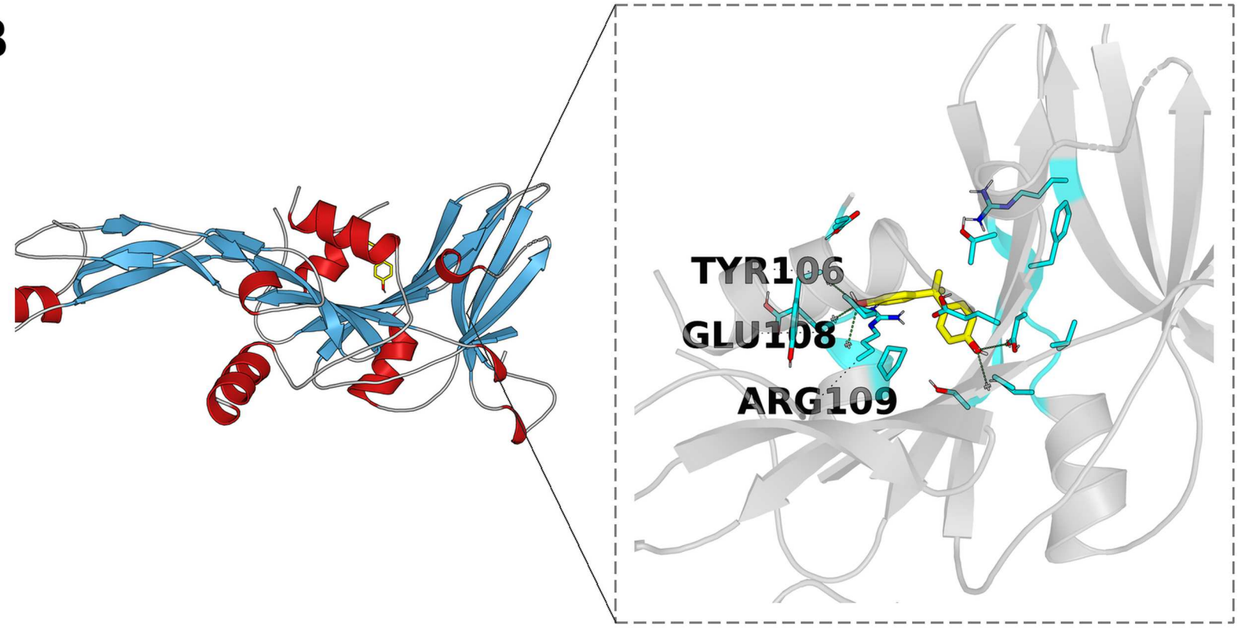

FLNB high vs low (tumour)
Ribosome (NES=-3.01)
Oxidative phosphorylation (NES=-2.05)
Other types of O-glycan biosynthesis (NES=1.90)
Protein export (NES=-1.80)
GnRH signaling pathway (NES=1.77)
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset

A



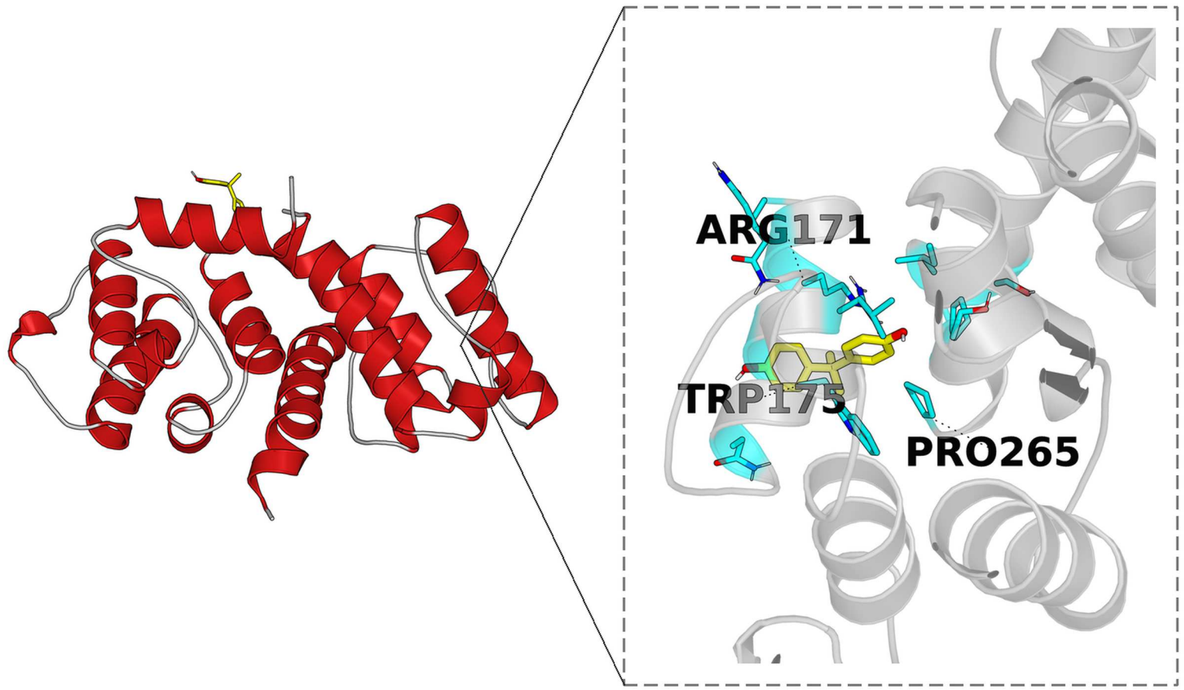
ACTA2-BPA
-5.70 kcal/mol

B



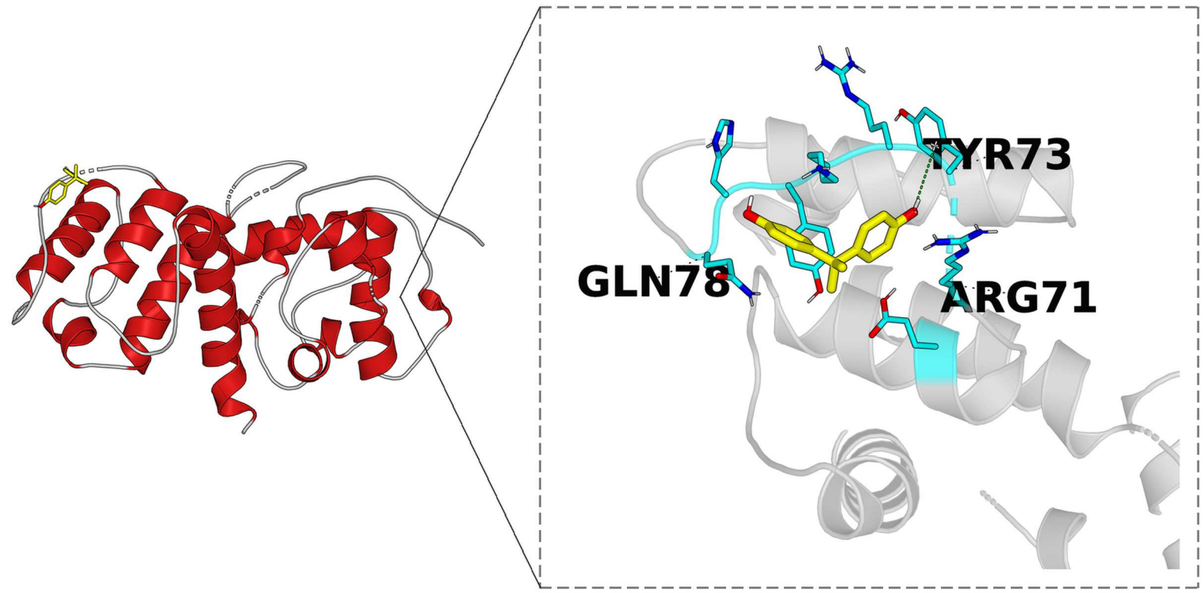
BMP4-BPA
-5.68 kcal/mol

C



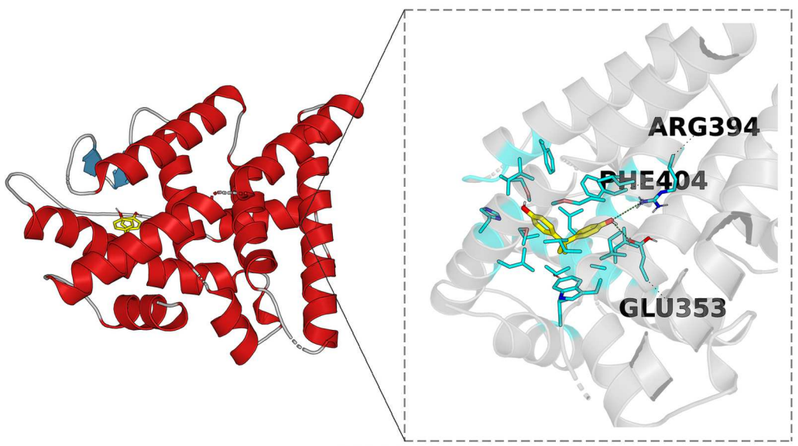
FLNA-BPA
-5.70 kcal/mol

D



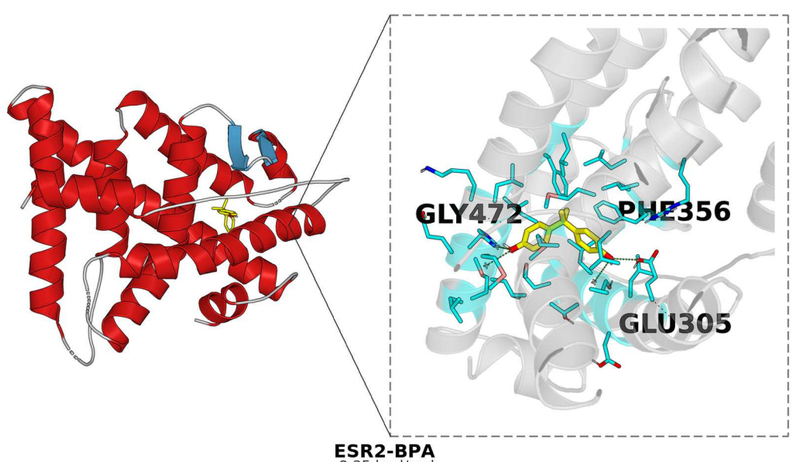
FLNB-BPA
-5.17 kcal/mol

E



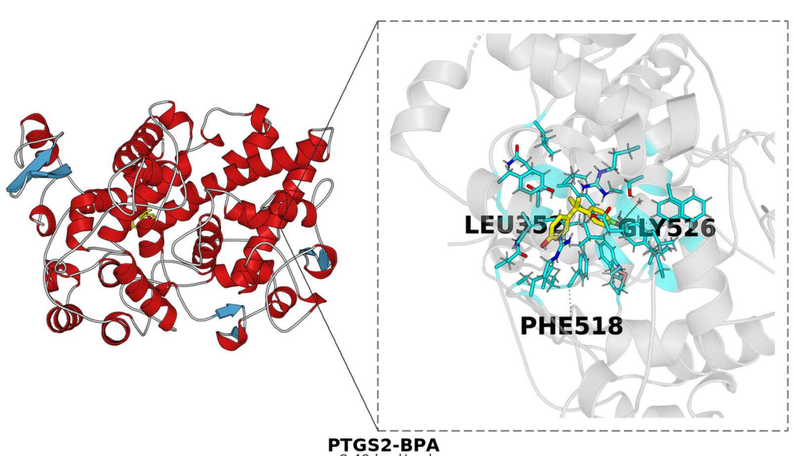
ESR1-BPA
-8.21 kcal/mol

F



ESR2-BPA
-8.35 kcal/mol

G



PTGS2-BPA
-8.49 kcal/mol