STDrug enables spatially informed personalized drug repurposing from spatial transcriptomics

Yiwen Yang[1,$], Thatchayut Unjitwattana[2,$], Shu Zhou[1], Suguru Kadomoto[6], Xiaotong Yang[1], Tianao Chen[4], Abdullah Karaaslanli[1], Yuheng Du[1], Wanxing Zhang[1], Haodong Liang[5], Xuhui Guo[1,7], Evan T. Keller[6], Lana X. Garmire[1,7#]

[1]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA

[2]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA

[3]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

[4]Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor, MI, USA

[5]Department of Statistics, University of Michigan, Ann Arbor, MI, USA

[6]Department of Urology, University of Michigan, Ann Arbor, MI, USA

[7]Department of Biomedical Informatics and Data Science, University of Alabama at Birmingham, Birmingham, AL, USA

[$]: These authors contribute equally to this work

[#]: Corresponding author email: lgarmire@uab.edu

Abstract

Drug repurposing offers a scalable route to accelerate therapeutic discovery, yet existing approaches based on single-cell RNA sequencing (scRNA-seq) often overlook spatial tissue context, limiting their ability to capture microenvironment-dependent drug responses. Here we present **STDrug**, a spatially informed computational framework that integrates spatial transcriptomics, graph-based modeling, and

multimodal learning to enable patient-specific therapeutic prioritization. STDrug identifies and aligns disease and control spatial domains using graph convolutional networks and coherent point drift, and prioritizes candidate drugs through an integrative scoring scheme combining tumor-reversible gene signatures, perturbation-based reversal scores, and knowledge-guided gene weighting within a machine learning framework. By modeling spatial domain interactions alongside predicted drug efficacy and toxicity, STDrug generates robust patient-level drug scores. Across hepatocellular carcinoma and prostate cancer datasets, STDrug outperforms existing single-cell and spatial transcriptomics–based drug repurposing methods, achieving signficantly improved predictive accuracy (AUCs=0.81-0.82) across patients. Validation using large-scale electronic health records and in vitro assays further supports the translational relevance of top-ranked candidates. Taking together, STDrug establishes a generalizable framework for incorporating spatial omics into therapeutic discovery, advancing spatially informed and personalized drug repurposing.

Introduction

Drug discovery remains a time-consuming and costly process, with most candidates failing before reaching the clinic[1]. Drug repurposing – identifying new therapeutic indications for approved or investigational compounds – offers a faster and less risky alternative by leveraging existing pharmacological and safety data[2]. Unlike de novo discovery, which relies on target identification and compound screening, drug repurposing strategies exploit diverse molecular, clinical, and computational resources to reposition known drugs toward unmet clinical needs.

Computation has become central to drug repurposing. Based on analytical approaches, it can be classified as network-based methods[3,4], machine learning and deep learning models[5,6] etc. Depending on the input data, drug repurposing can be categorized by drug-centric, disease/target-centric, or multi-modal approaches. Among them, omics-driven approaches, in particular transcriptomic-based approaches reversing disease-associated molecular signatures, have shown success[7,8]. However, the earlier

transcriptomics-based approach used bulk gene expression, overlooking the cellular heterogeneity within the tissue that may significantly bias the disease associated signatures[9].

To address this issue, some recent work was developed for drug repurposing, using single-cell RNA-seq (scRNA-Seq) data as the input. For example, Beyondcell[10] is a framework that utilizes scRNA-seq data to identify tumor subpopulations with distinct drug response signatures. It prioritizes cancer-specific therapeutic candidates based on enrichment of drug sensitivity profiles. DrugReSC[11] integrated single-cell transcriptional landscapes with compound-induced perturbation profiles to systematically identify repurposable drugs tailored to cellular phenotypes. We developed ASGARD, a framework that accounts for cell-type heterogeneity in scRNA-seq data to derive the comprehensive drug score holistically, by searching the candidate prepurposeful drugs through the L1000 pharmacogenomics database[12]. However, a major limitation of scRNA-Seq data is the lack of spatial context and organization of the single cells, which may include important clues on drug treatment effect[13].

To address this issue, we here propose STDrug, a new computational method dedicated to using spatial transcriptomics data to aid drug repurposing. STDrug exploits spatial information specifically in identifying spatial domain pairs between the disease and healthy tissue. It proposes potential drugs and compounds from large-scale pharmacogenomics perturbation data, using a machine-learning method that learns the importance of the disease-associated genes by referring to the ground truth obtained from pre-trained large language models (LLMs). It offers realistic estimation for repurposeful drugs by considering drug efficacy, toxicity, and potential side effects. In short, STDrug aims to bridge a major gap currently, in bringing the cutting-edge technology of spatial transcriptomics towards guiding therapeutic treatment decision-making.

Results

The architecture of STDrug

STDrug comprises three major steps, including 1) data preprocessing and integration, 2) spatial domain identification, and 3) drug repurposing **(Fig. 1)**. STDrug first takes the spatial transcriptomic data from paired diseased and normal tissues, ideally from the same patient, as the input. It then performs a batch effect correction and sample alignment to reduce variability between patients. Inspired by SpaGCN[14], we used batch-corrected Harmony embeddings rather than PCA to combine spatial coordinates and build a graph convolutional network (GCN) that learns spatially-aware low-dimensional embeddings. Different from SpaGCN, these embeddings are uniquely adjusted by a disease-control balanced clustering algorithm to identify paired spatial domains between the diseased and normal spatial transcriptomics slides for a patient.

The resulting paired spatial domains serve as the foundation for the drug repurposing module, which prioritizes therapeutic candidates that maximally reverse the disease-associated gene expression changes from the pharmacogenomics perturbation databases (eg. L1000 and Tahoe-100M). The drug repurposing module utilizes potential drug information obtained from GPT-4o[17], a pre-trained LLM by OpenAI, along with a ML method to assign weights to reversible disease-associated genes. Doing so prioritizes those drugs most likely to play key roles. In addition to reversible disease-associated genes, the drug repurposing also considers spatial domain interactions, as well as drug side effect information from the Side Effect Resource (SIDER) database (version 4.1)[19] and drug efficacy from Genomics of Drug Sensitivity in Cancer (GDSC) database[20]. The resulting comprehensive drug score is calculated by integrating the multiple factors stated above. The most promising candidate drugs are those with top-ranking drug scores at the patient-specific level.

To demonstrate the utility of STDrug, we validate predicted top drug results using a combination of (1) empirical evidence from literature, clinical trials, and LLM-based validation, (2) in-silico validation using

the real-world clinical data, and (3) in-vitro validation using multiple cell line experiments. Together, this study presents a spatially-informed, patient-specific, end-to-end drug repurposing framework that systematically integrates spatial transcriptomics, machine learning, and pharmacogenomics drug perturbation data, rigorously validated by real-world clinical and experimental evidence.

Spatial domain identification and benchmarking

To evaluate the clustering results generated by STDrug, we benchmarked STDrug against four popular spatial domain identification tools, including STAGATE[24], STAligner[25], SPACEL[26], and GraphST[27]. Due to the lack of ground truth for spatial domains from paired tumor and normal tissues which are crucial for the drug repurposing module of STDrug, we simulated paired tumor and adjacent normal dataset from real HCC data, where each spatial domain is occupied by a cells of the same type whose gene expression values are based on its scRNA-Seq datadisribution **(Fig. 2a)**. STDrug yielded the best averaged values of normalized mutual information (NMI) and adjusted rand index (ARI) that are significantly better than other methods as shown in **Fig. 2b** (NMI = 0.93, ARI = 0.94). STAGATES yielded the 2nd best ARI and NMI (NMI = 0.78, ARI = 0.79), followed by STAligner (NMI = 0.76, ARI = 0.75). Comparatively, SPACEL (NMI = 0.36, ARI = 0.30), and GraphST (NMI = 0.24, ARI = 0.19) have the lowest accuracies (**Fig. 2b)**. Detailed investigation on the slide level of simulated spatial transcriptomes of each patient shows that the spatial domains identified by STDrug is consistently better aligned with the spatial domain ground truth, compared to other spatial domain identification methods **(Fig. 2c)**. These results highlight the robustness and accuracy of STDrug in spatial domain identification which relies on precise contextual mapping between tumor and normal tissue, reinforcing its suitability for drug repurposing by STDrug.

Drug repurposing and benchmarking

Since interactions between spatial domains may also impact cellular processes within tissues[28], STDrug also exploits domain-domain interactions (DDI) for drug score calculation. It computes the ratio of

interaction strength between each pair of spatial domains by CellChat v2[29]. Domains exhibiting larger differences in DDI in diseased vs control conditions are considered more influential, when other parameters are equal. By integrating the proportions among all spatial domains, their DDIs, as well as the therapeutic connectivity FDR, a patient-level comprehensive therapeutic score is computed. To further take into account both drug efficacy and safety, STDrug can optionally include drug efficacy and safety scores in the comprehensive drug score, by leveraging the drug sensitivity from the GDSC database[20] and drug side effects information from the SIDER database[19]. To enable comparison among patients, the drug score is rank-scaled between 0 and 1.

We tested STDrug by predicting potential drugs for hepatocellular carcinoma (HCC) and prostate cancer (PCa) treatments. The HCC data are taken from four patients with paired tumor and adjacent normal tissues[30] and the PCa data are from two patients with paired tumor and adjacent normal tissue spatial transcriptomics data, where the non-tumor regions were annotated by pathologists[31]. Despite the existence of drug response prediction methods, the field currently generally lacks dedicated spatial-transcriptomic-based drug repurposing methods in which drug selection and ranking is the objective, except Beyondcell[10] which was extended to spatial transcriptomics data We also compared with ASGARD[12] in which scRNA-Seq is the input.We used the area under the ROC curve (AUC) to evaluate the prediction accuracy, where the “proxy truth data” of drug-disease association are obtained from clinical trials, and a pre-trained GPT-4o model using a hold-out set of literatures independent from those used to train the XGBoost model[17]. Note the hold-out set of literature from GPT-4o search is separate from those used to train the XGBoost model.

For HCC data, STDrug achieves an AUC of 0.81, significantly outperforming both ASGARD (AUC = 0.61) and Beyondcell (AUC = 0.59) in identifying potential drugs (**Fig. 3a)**. The significance of potential top drugs across four HCC patients are shown in **Fig. 3c**. STDrug consistently identifies promising drug

candidates across HCC patients, with bosutinib and sorafenib as top candidates in four patients. Sorafenib is the first-line chemotreatment for advanced HCC patients, acting by suppression of angiogenesis and tumor proliferation[32,33]. The top ranking of Sorafenib confirms the strong potential of STDrug in drug repurposing prediction for HCC. Bosutinib, an FDA-approved tyrosine kinase inhibitor for Philadelphia chromosome-positive chronic myeloid leukemia, has been reported to exhibit anti-tumor activity in preclinical models of hepatocellular carcinoma[34]. STdrug also predicted repurposeful drug used to treat other cancers such as daunorubicin (for AML and ALL), dactinomycin (for rare pediatric cancers such as Ewing sarcoma), trametinib (melanoma), panobinostat (multiple myeloma), and epirubicin (breast, colon and gastric cancers). STDrug additionally identified a diverse array of therapeutic candidates in at least one patient – including non-drug supplements niacin (vitamin B3) and caffeine, and medications for treating cardiovascular diseases including atorvastatin, digoxin, and losartan. Niacin has been reported to suppress liver tumor growth by enhancing antitumor immunity[35]. Caffeine has been shown to HCC tumor suppressing functions, ranging from cell lines[36], mouse models[37,38], to large-scale epidemiological studies[39–42]. Atorvastatin, commercial name as Lipitor, is used to prevent heart attack and stroke clinically. It has been clinically shown to improve HCC prognosis[43,44]. Similarly, digoxin (Lanoxin), a cardiac glycoside commonly used for heart failure and atrial fibrillation, has been reported to suppress angiogenic and tumor growth in HCC[45].

For PCa, STDrug also shows superior performance with an AUC of 0.82, compared to ASGARD (AUC = 0.57) and Beyondcell (AUC = 0.48) (**Fig. 3b)**. It identifies 13 common drug and chemical compound candidates for PCa in the two patients **(Fig. 3d).** Among them, top five drugs have shown anti-prostate cancer functions, including bortezomib[46], vorinostat[47], sirolimus[48], panobinostat[49], and doxorubicin[50]. Interestingly, STDrug also prioritized ouabain and digoxin-agents traditionally used for heart conditions- as potential therapeutic candidates for prostate cancer.

Validation of potential repurposeful drugs for HCC and PCa

We first conducted validations using the real-world patient clinical data from MarketScan (n=264 million patients). As patients diagnosed with cancers were seldomly given repurposeful new drug proposed here, we focused on the patients exposed to repurposeful drugs before cancer diagnosis. A subset of predicted repurposed drugs reached the patient size threshold (n=66) among the cancer of interest, after matching them with untreated controls by demographics and comorbidities. These drugs are atorvastatin, digoxin, and niacin for HCC, and digoxin, colchicine, pitavastatin, and sirolimus for PCa. As shown by Kaplan-Meier curves, atorvastatin, digoxin, and niacin all significantly delay the onset of HCC compared to those with no history of using these drugs **(Fig. 4a).** Digoxin, colchicine, pitavastatin, and sirolimus also exhibit notable delays in PCa development compared to the control group **(Fig. 4b)**. The Hazard Ratios (HR) of the Cox proportional hazards models are shown in **Fig. 4c-d,** further confirming the protective effects of these candidate drugs/compounds.

Additionally, we conducted in-vitro validation of the repurposed drug candidates for PCa by cell viability assays. As complement, we evaluated the effects of bortezomib and vorinostat, the candidates which lack sufficient clinical records among PCa patients. We tested five prostate cancer cell lines, including C4-2B, DU145, DU145-TxR, PC3, and PC3-TxR. **(Fig. 4e-f)**. Bortezomib, a protease inhibitor used in multiple myeloma[51], was effective against all prostate cancer cell lines (IC50: C4-2B = 8.57 nM, DU145 = 21.2 nM, DU145-TxR = 11.4 nM, PC3 = 16.3 nM, PC3-TxR = 8.55 nM; clinical blood concentration = 53 nM[52]). Vorinostat, a histone deacetylase (HDAC) inhibitor approved for the treatment of lymphoma[53], also shows selective efficacy against DU145-TxR, PC3 and PC3-TxR prostate cancer cell lines (IC50: DU145-TxR = 1.59 uM, PC3 = 0.963 uM, PC3-TxR = 0.508 uM; clinical blood concentration = 2uM[54]). Taxane-resistant cell lines (PC3-TxR and DU145-TxR) exhibit slightly higher sensitivity to vorinostat compared to their parental counterparts (PC3 and DU145). These findings highlight the potential of bortezomib and vorinostat for prostate cancer treatment, especially in the treatment-resistant prostate cancer subtypes.

Bioinformatics analysis of the effect of HCC candidate drugs

To interpret how candidate drugs may affect the cells in HCC ST data, we first performed analysis of spatial domains from HCC tumor and adjacent normal tissues (**Fig. 5a, b).** On each spatial domain, we conducted cell type deconvolution using cell2location[55] **(Fig 5c)**. As a result, five related but transitional spatial domains are detected optimally (**Fig 5a**), characterized as a proliferative adaptive domain C1, a fibrotic immune-evasive domain C2, a transitioning immunoactive domain C3, an immune-limited adaptive domain C4, and an immunosurveillance-disrupted domain C5.

To observe drug effects in these spatial domains, we visualized the drug scores stratified by spatial domains. As an example, **Fig. 5d-e** shows 10 top drugs on patient 2, by UMAP and 2D-geolocation plot respectively. The main targeted regions by candidate drugs vary and can be roughly divided into 2 classes. Class 1 includes sorafenib, niacin, and trametinib that main act on spatial domains 2, 3 and 4, particularly domain 2 which is enriched with hepatocytes but relatively low in fibroblasts and immune cells. Others (eg. bosutinib, caffeine and panobinostat) mainly act on spatial domains 3 and 4. In particular, caffeine exerts the most significant effect on domain 3- the transitioning immunoactive domain.

To explore the potential molecular mechanism of the top drug candidate bosutinib, we analyzed enriched pathways of the putative target genes (gene that have reversed expression under bosutinib, compared to the HCC condition) across the five spatial domains **(Fig. 5f)**. While a few pathways are activated across all spatial domains, such as PPAR, PI3K-Akt and apoptosis pathways, pathway dys-regulation is much more prevalent in spatial domains C3 and C4 as compared to C1 and C2. This is consistent with that C3 and C4 are the main spatial.domains affected by bosutinib. In domain C3 (the transitioning immunoactive domain) MAPK signaling is enriched (**Fig 5f**), supported by the evidence that overexpressed pro-oncogenic MAP2K2 (MEK2) and MAP4K4 in the MAPK/ERK cascade are predicted to be down-regulated by bosutinib indicated by blue edge (**Fig. 5g**). As bosutinib is an Src/Abl tyrosine kinase inhibitor, ITGB1, a gene in the canonical Src-associated signaling pathway, is predicted to be down-

regulated as expected. In addition, multiple cytoskeleton-associated genes that are part of Src-regulated adhesion and motility programs are predicted to be down-regulated by bosutinib, including FLNA and PFN1 in the actin-remodeling components, and ACTB and ACTA2 in the actin filaments. Additionally, molecular chaperones HSP90AA1 and HSP90AB1, which was linked to stability of the Src signaling axis[56,57], are also predicted to be downregulated by bosutinib, indicated by blue edges in **Fig. 5g**. Overall, Src-regulated biological functions show repression, being the targets of bosutinib.

Bioinformatics analysis of the effect of PCa candidate drugs

We conducted similar spatial domain analysis in the PCa patients. Interestingly, STDrug again identifies five optimally distinct domains after integrating the tumor and normal tissues (**Fig 6a, 6b**). We classify them as the vascular remodeling domain C1, fibrovascular stromal domain C2, adaptive immune-regulated domain C3, metabolic and immune-responsive domain C4, and epithelial-driven tumor expansion domain C5**.** As expected, the tumor samples have much higher presence of C5 but much lower C2 domains (**Fig 6b**). The cell-type deconvulation results show that tumor has attracted significantly more naive T-cell and B-cells to the site, however is reduced in fibrobrasts and pericytes (**Fig 6c**). Furthermore, we investigated the effect on the spatial domains from top common drugs, using patient 1 as an example **(Fig. 6d-e)**. Most drug candidates act on multiple domains. Despite significant spatial domain-level differences in the drug scores among these drugs, C3 domain (adaptive immune-regulated domain) appears to be a common targeted spatial domain. C3 domain is characterized by significant increase of B-cell and naive T-helper cell population but largest decrease of fibroblasts in tumor samples (**Fig 6c)**. C1 domain (vascular remodeling domain) is the other common, but less-strongly targeted spatial domain. Corroborating with **Fig 3e**, Bortezomib and Vorinostat, the top 2 most significant drugs, exert the strong effects on all domains except C5 (**Fig 6e**). Sirolimus targets C3 relatively the strongest, followed by panobinostat, niclosamide, everolimus, ouabain, azacitidine, and doxorubicin in the descending order of the drug score over C3.

To further elucidate the molecular mechanisms underlying the top candidate drug bortezomib, we analyzed the genes affected as well as their enriched pathways across the spatial domains. **(Fig. 6f,g)**. Corresponding to the domain level drug impact, the altered pathways in C5 are the least and weakest. Some most significant pathways include apoptosis in spatial domains C1-C4, which is a known consequential effect of bortezomib; cGMP-PKG signaling pathway and PI3K-Akt signaling pathway in domains C1-C3, and TNF and IL-17 signaling pathways in domains C1 and C3. Some of the genes previously known to be affected by bortezomib are among the putative target genes including the direct drug target PSMB5 and downstream genes TNFSF10, DCN, and WIF1 **(Fig. 6g)**. PSMB5 level is upregulated in prostate cancer (red node) and repressed by Bortezomib as indicated by the blue edge (**Fig. 6g)**. TNFSF10, encoding the pro-apoptotic ligand TRAIL, is downregulated in the tumor (blue node) and predicted to be restored by bortezomib (red edge). Two prostate cancer tumor suppressors, DCN and WIF1, are also downregulated and predicted to be indirectly rescued by bortezomib (red edge).

Discussion

We present STDrug (Spatial Transcriptomics aided Drug Repurposing), a computational framework that enhances drug repurposing by leveraging spatial transcriptomic data. Beyond data preprocessing and integration, STDrug is characterized by two unique and innovative modules: the spatial domain identification module and the drug repurposing module. The spatial domain identification module aligns matched spatial regions between diseased and normal tissues using a graph convolutional network (GCN) combined with coherent point drift (CPD) algorithm. The superior accuracy of spatial domain pair identification is confirmed by the benchmarking study with other tools, including STAGATE[24], STAligner[25], SPACEL[26], and GraphST[27].

We evaluated STDrug by various approaches, from literature/text-based validation, real-world EHR based validation, to in vitro experimental system. As no spatial drug-repurposing framework currently exists, we first compared STDrug to scRNA-seq-based approaches, including ASGARD[12] and Beyondcell[10], and evaluated by the literature and clinical trial data curated by Gen AI. STDrug yielded significantly higher AUCs in both hepatocellular carcinoma (HCC) and prostate cancer (PCa) patient data. Next, using real-world large-scale MarketScan clinical dataset for case-control studies, we confirmed the significant therapeutic effect of top drugs predicted by STDrug, including atorvastatin, digoxin and niacin for HCC, and digoxin, colchicine, pitavastatin and sirolimus for PCa. They were associated with delayed disease onset time and improved survival in propensity-adjusted Cox and Kaplan-Meier analyses. Lastly, we showed that in vitro cell culture across five PCa cell lines, including taxane-resistant PC3-TxR and DU145 models, confirmed potent cytotoxicity of bortezomib and vorinostat at clinically achievable concentrations[58–60].

In HCC, STDrug prioritized bosutinib and sorafenib, with sorafenib serving as an internal positive control as a first-line therapy for advanced HCC. Bosutinib-predicted signatures included suppression of MAPK and Src-associated signaling (MAP2K2, MAP4K4, ITGB1), suppression of cytoskeletal regulators (FLNA, PFN1, ACTB, ACTA2), and reduced HSP90 chaperone activity (HSP90AA1, HSP90AB1), consistent with destabilization of oncogenic kinase networks and disruption of mitochondrial respiration[34]. Beyond kinase inhibitors, STDrug also identified several non-oncology agents with potential relevance to HCC. Niacin (vitamin B3) was associated with enhanced antitumor immunity through reduced immunosuppressive myeloid polarization and restoration of CD8 T-cell activity[35], while caffeine has been linked to reduced HCC risk and suppression of proliferation via inhibition of adenosine signaling and PI3K-Akt activity[37,39–42]. Cardiovascular medicine including atorvastatin and digoxin were also highlighted, aligning with reported inhibition of MAPK signaling and hypoxia-inducible pathways, respectively[43–45,61].

In PCa, STDrug prioritized bortezomib and vorinostat as top candidates, which were validated by the cell line experiments. Bortezomib, a proteasome inhibitor, exerts its antitumor activity by selectively and reversibly blocking the 26S proteasome – a multiprotein complex responsible for degrading ubiquitinated proteins involved in cell cycle regulation and apoptosis[62,63]. On the other hand, vorinostat, a histone deacetylase (HDAC) inhibitor, promotes histone hyperacetylation, leading to chromatin remodeling and the re-expression of silenced tumor suppressor genes, which induces cell cycle arrest at the G1 and G2/M checkpoints[64,65]. Further analyses indicated that bortezomib is associated with perturbation of PI3K-Akt, cGMP-PKG, TNF, and IL-17 signaling pathways in multiple spatial domains. This is consistent with coordinated modulation of survival, inflammatory, and androgen receptor-driven transcriptional programs[46,66–76]. These predictions align with established bortezomib mechanisms, including proteotoxic stress induction, G0/G1 arrest, and antiangiogenic effects in xenograft models[46,72]. Predictions on the downregulation of PSMB5, a direct proteasome target of bortezomib is consistent with established mechanisms of bortezomib on proteasome inhibition. Additionally, restoration of pro-apoptotic TNFSF10 (TRAIL) signaling and tumor suppressors DCN and WIF1, inferred from the spatial transcriptomics data by bortezomib treatment, also confirm its roles in the proteotoxic stress-induced apoptosis and tumor-suppressive signaling activation in prostate cancer[77–80]. Interestingly, STDrug also identified multiple endocrine pathways regulated by hormones as the potential targets of bortezomib, including relaxin, oxytocin, estrogen, and prolactin signaling, supporting the cross-talk with androgen receptor (AR)-driven transcriptional programs, which are central to prostate cancer biology[81].

In summary, STDrug represents a comprehensive computational framework for drug repurposing using spatial transcriptomics information. By integrating spatial domain architecture, transcriptional reversibility, pharmacological constraints, and machine-learning and generative AI-based marker prioritization, it identifies clinically validated therapies. STDrug is expected to inform precision therapeutic discoveries, supported by mechanistic insight for repurposing from drugs and compounds.

Methods

Spatial transcriptomics (ST) dataset

To demonstrate the utility of STDrug, we applied it to two Visium ST datasets for two cancer types, HCC and PCa. The HCC Visium dataset consists of four patients, each with one tumor tissue and one matched adjacent normal tissue. The accession number for the raw sequencing data deposited in Genome Sequence Archive (GSA) is HRA000437 (https://ngdc.cncb.ac.cn/gsa-human/browse/HRA000437)[30]. The PCa Visium dataset consists of two patients, one with seven tumor tissues and two adjacent normal tissues, the other with four tumor tissues and 11 adjacent normal tissues. The spatial count matrices and high-resolution images are available on Mendeley Data (https://doi.org/10.17632/svw96g68dv.1)[31].

Processing of spatial transcriptomics data

We initiate the spatial transcriptomics data analysis by performing normalization and log transformation. The *normalize_total* function from Scanpy is used to scale the total counts per cell, followed by the application of the log1p function to stabilize the variance. Next, we reduce the data's dimensionality through principal component analysis (PCA), where 50 principal components are computed using the *pp.pca* function. Following PCA, we apply the Harmony algorithm to correct for batch effects across different experimental conditions. This batch correction ensures the data are harmonized across patients, reducing unwanted technical variation while preserving biological signals.

Functional downstream analysis of spatial transcriptomic data

Differential expression analysis was performed using the Seurat package (version 5.0.1)[82]. We performed GSEA using *gseKEGG()* in the package ClusterProfiler (4.8.2) with the gene set information from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[83,84]. The spatial domain interaction inference was performed using the package CellChat v2[29]. Cell-type deconvolution was performed using

the Cell2location package[55]. Cell-type deconvolution of HCC tissues was performed using the integrated dataset of GSE189903[85] and GSE149614[86] as a reference. Cell-type deconvolution of prostate cancer tissues was performed using GSE181294[87] as a reference. Cell-type dropout experiments were conducted by subtracting the average expression of the target cell type in the single-cell reference data from the spatial expression deconvolution results.

Simulation of spatial transcriptomics data

Due to the lack of ground truth for spatial domains from paired diseased and normal tissues, we've developed a data simulation method to generate spatial transcriptomics (ST) data while ensuring paired spatial domains across conditions. For each tissue, we derived k spatial domains by jointly clustering cells based on their original gene expression profiles and the Euclidean distances between their spatial locations. Spatial domains between normal and tumor tissues were then matched, where a single cell-type cluster was assigned to a spatial domain (e.g. hepatocyte, macrophage, monocyte, NK cell, and T-cell). The cell type-specific gene expression distribution is derived from scRNA-seq reference data, following the Poisson distribution.

Spatial Domain Identification

Spatially Aware Embedding

We construct a spatially aware representation by integrating batch-corrected low-dimensional gene expression embeddings with spatial neighborhood structure. Specifically, spatial graphs are built for each sample and incorporated into a unified graph learning framework, where expression features and spatial connectivity jointly inform representation learning. We employ a graph neural network architecture adapted for multi-sample inputs to capture shared and sample-specific spatial patterns. The model is optimized to learn latent embeddings that resolve spatial domains. These embeddings are subsequently used to identify conserved spatial structures across patients.

Disease-Control Balanced Clustering

The spatially aware embeddings were further processed using a nonlinear dimensionality reduction approach that preserves local neighborhood structure. To facilitate within-patient cross-condition comparison, embeddings derived from diseased and control tissues were aligned using a probabilistic non-rigid registration framework that enforces smooth spatial correspondences. Based on the aligned embeddings, we implemented a condition-balanced clustering strategy to jointly partition cells or spots across samples. Clustering was performed through iterative community detection under multiple resolution parameters, with constraints to ensure representation from both conditions and across patients within each cluster. An initial partition was obtained using a data-driven estimate of domain complexity, followed by iterative refinement steps that adjusted cluster granularity and reassigned underrepresented elements to achieve balanced partitions. The resulting clusters define spatial domains that are comparable across disease states and individuals.

Drug Score Calculation

Construction of Integrated Drug Perturbation Reference

To construct a comprehensive reference for drug response, we integrated large-scale bulk and single-cell perturbation datasets into a unified framework. Single-cell perturbation profiles were aggregated into pseudo-bulk representations at the cell line level to ensure compatibility with bulk expression signatures. Expression profiles were standardized using within-sample normalization followed by robust scaling to enable cross-dataset comparability. The processed pseudo-bulk data were then harmonized with bulk perturbation data on a shared gene set and overlapping compound space. The resulting integrated reference comprises a large collection of perturbational signatures spanning multiple cell lines and tissue contexts.

DEG and CEG Z-score Calculation

Using the paired spatial domains identified from the STDrug clustering module, we compare the differentially expressed genes (DEGs) between tumor tissue and its adjacent normal tissue for each domain. Integrating gene expression profiles from the integrated drug perturbation reference, we calculate a standardized Z-score based on the probability of each gene being a concordantly expressed gene (CEG).

LLM-Enhanced Perturbed Gene Weights

To quantify the potential contribution of each CEG to the drug repurposing process, we introduce a perturbed gene weight by training the machine-learning classifier. To assist drug power labeling, we use gpt-4o (version 2024-08-06), a pre-trained large language model (LLM) from OpenAI to convert the base knowledge of the drug-disease relationship extracted from PubMed articles. We then calculate the feature gain for each gene.

Patient Therapeutic Drug Score

For each drug, a patient-specific therapeutic score was computed by integrating signals across spatial domains. This score accounts for domain composition, relative differences between conditions, and interaction patterns among spatial regions, resulting in a composite measure of predicted therapeutic relevance at the individual level. In addition, we incorporated complementary measures of drug efficacy and safety derived from external pharmacological and clinical knowledge sources. These measures were standardized and integrated to refine drug prioritization.

Repurposed Drug Ranking

The personalized drug score is a combination of the individual patient therapeutic score, drug efficacy score, and drug safety score. To facilitate comparison across different patients, we also provide a rank-based standardized drug score (percentile).

Evaluation of potential repurposed drugs and benchmarking

Drug predictions from STDrug were evaluated using the area under the ROC curve (AUC), comparing predicted drugs against a reference set of known or validated drugs for each cancer type. To construct the validated drug reference list, we used FDA-approved cancer drug databases, clinical trial results from the NIH clinical trial registry (clinicaltrials.gov), DrugBank, and peer-reviewed literature from PubMed with supporting evidence from experimental or pharmacological studies. To benchmark STDrug against single-cell RNA-seq-based drug repurposing methods, including ASGARD and Beyondcell, we selected the top 80 HCC drugs and top 50 PCa drugs predicted by each method and compared their AUC scores against the validated drug reference list.

In-silico validation of potential repurposed drugs

Merative MarketScan is a large private research database of de-identified US patient data, covering 264 million employees and their dependents from 2009 to 2022. It includes claims data, medical, drug, dental records, lab results, and hospital discharges, sourced from insurance providers, Blue Cross Blue Shield, and third-party administrators, offering insights into healthcare utilization and costs.

Our approach to validate STDrug repurposed drugs for the cancer types of interest using EHR data involved identifying treatment and control cohorts, and comparing their risk of developing cancer over a defined observation period. First, we designated patients who were administered or prescribed the drugs as the treated group. We then matched four patients from the untreated control group to each individual in the treated group based on birth year, sex, geographic location, and baseline comorbidities, including hypertension, type 2 diabetes (T2D), and coronary artery disease (CAD). Corresponding International Classification of Diseases (ICD) codes were used to identify the comorbidities diseases from individual records.

The observation period for the treated group was defined as the time between the first drug order and 60 days following the last drug order, provided there were no subsequent orders. For the control group, the observation period was matched to that of the treated group based on patient birth year, sex, geographic location and medical condition. For both the treated and control cohorts, we collected available medical records and identified individuals diagnosed with HCC or PCa based on ICD codes. Patients diagnosed within the observation period were marked as 1, while those who were not diagnosed were censored as 0.

We employed the Cox proportional hazards (CoxPH) model, stratified by propensity score, to analyze the relative risk of disease development between the treated and control cohorts. The propensity score was estimated using the same covariates applied in the case–control matching. Specifically, a logistic regression model was fitted to estimate each individual's propensity score, representing the probability of disease development conditional on observed covariates. The stratification by propensity score allows for the adjustment of baseline confounders and ensures that the treated and control groups are comparable in terms of their likelihood of developing the disease. The p-value derived from the CoxPH model indicates whether the treated cohort has a statistically significant lower risk of developing the disease compared to the control cohort within the observation period, after adjusting for the propensity score. Kaplan-Meier (KM) curves were fitted to compare the disease-free survival probabilities over time between the treated and control cohorts.

In-vitro validation of potential repurposed drugs for prostate cancer

The prostate cancer cell lines C4-2B, DU145, and PC3 were obtained from American Type Culture Collection (ATCC). Taxane-resistant DU145 (DU145-TxR) and taxane-resistant PC3 (PC3-TxR) were generated via paclitaxel exposure[91]. All cell lines were cultured in RPMI medium containing 10% FBS and penicillin-streptomycin (15140-122, Invitrogen). Cells were cultured at 37°C in 5% CO2. Candidate

repurposed drugs, including bortezomib (S1013) and vorinostat (S1047) were purchased from Selleck Chemicals. The cell proliferation reagent WST-1 (11644807001) was purchased from Roche for cytotoxicity assessment.

Each STDrug repurposed drug was diluted in DMSO and added to each well so that the DMSO concentration in the medium was equal and less than 1%. Cells were seeded at a density of 5.0×104 cells/well in 96-well cell culture plates and cultured in 100uL of medium containing multiple concentration of drugs. The effect of the drugs were evaluated by a cytotoxicity assay using WST-1 reagen. Cells were cultured in a CO_2 incubator at 37°C. After 48 hours of incubation, 10 µl of WST-1 reagent was added to 100 µl of medium in each well, and the plates were incubated at 37°C in a CO_2 incubator for 1-2 hours. Using a Synergy H1 multi-detection reader (BioTek), the background absorption

$$\text{Cell Viability } (\%) = 100 * \frac{As - Ab}{Ac - Ab}$$

at a wavelength of 650 nm was measured and subtracted from the actual absorption value at a wavelength of 450 nm. Absorbance was measured for the drug-exposed group, negative control group, and blank group. Cell viability was calculated as where As is absorbance of sample, Ab is absorbance of blank (culture medium without cells), and Ac is absorbance of blank (culture medium without cells). The IC50 was then calculated using GraphPad Prism[92].

Acknowledgments

The authors acknowledge all lab members of Garmire Group for helpful discussions.

Author contributions statement

L.X.G conceived this project and supervised the study. A.K. and S.Z. developed the first version of the tool. Y.Y. and T.U. modified the pipeline, carried the analysis, and wrote the manuscript. S.K., supervised

by E.K, performed in-vitro drug validation in prostate cancer cell lines. X.Y. performed in-silico drug validation. Y.D. assisted the functional downstream analysis. T.C. assisted the disease-control balanced clustering implementation. H.L assisted the data simulation. X.G. assisted with technical troubleshooting and packaging STDrug. All authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Data availability

All pre-processed data will be made publicly available upon the publication of this manuscript.

Funding Information

LXG is supported by grants from NLM R01 LM012373 and NIH R03 OD039978. TU is supported by a fellowship from the Royal Thai government. EK is supported by NIH P01CA093900.

References

1. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B. 2022 Jul;12(7):3049–62. doi:10.1016/j.apsb.2022.02.002

2. Kulkarni VS, Alagarsamy V, Solomon VR, Jose PA, Murugesan S. Drug Repurposing: An Effective Tool in Modern Drug Discovery. Russ J Bioorganic Chem. 2023 Apr;49(2):157–66. doi:10.1134/S1068162023020139

3. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabási AL, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. Nat Commun. 2018 Jul 12;9(1):2691. doi:10.1038/s41467-018-05116-5

4. Sadegh S, Skelton J, Anastasi E, Bernett J, Blumenthal DB, Galindez G, et al. Network medicine for disease module identification and drug repurposing with the NeDRex platform. Nat Commun. 2021 Nov 25;12(1):6848. doi:10.1038/s41467-021-27138-2

5. Kang H, Hou L, Gu Y, Lu X, Li J, Li Q. Drug–disease association prediction with literature based multi-feature fusion. Front Pharmacol. 2023 May 22;14:1205144. doi:10.3389/fphar.2023.1205144

6. Jin S, Niu Z, Jiang C, Huang W, Xia F, Jin X, et al. HeTDR: Drug repositioning based on heterogeneous networks and text mining. Patterns. 2021 Aug;2(8):100307. doi:10.1016/j.patter.2021.100307

7. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, et al. Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data. Sci Transl Med. 2011 Aug 17;3(96). doi:10.1126/scitranslmed.3001318

8. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua MS, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. Nat Commun. 2017 Jul 12;8(1):16022. doi:10.1038/ncomms16022

9. Garmire LX, Li Y, Huang Q, Xu C, Teichmann SA, Kaminski N, et al. Challenges and perspectives in computational deconvolution of genomics data. Nat Methods. 2024 Mar;21(3):391–400. doi:10.1038/s41592-023-02166-6

10. Fustero-Torre C, Jiménez-Santos MJ, García-Martín S, Carretero-Puche C, García-Jimeno L, Ivanchuk V, et al. Beyondcell: targeting cancer therapeutic heterogeneity in single-cell RNA-seq data. Genome Med. 2021 Dec 16;13(1):1. doi:10.1186/s13073-021-01001-x PubMed PMID: 34911571; PubMed Central PMCID: PMC8675493.

11. Liu C, Zhang Y, Liang Y, Zhang T, Wang G. DrugReSC: targeting disease-critical cell subpopulations with single-cell transcriptomic data for drug repurposing in cancer. Brief Bioinform. 2024 Sep 23;25(6):bbae490. doi:10.1093/bib/bbae490

12. He B, Xiao Y, Liang H, Huang Q, Du Y, Li Y, et al. ASGARD is A Single-cell Guided Pipeline to Aid Repurposing of Drugs. Nat Commun. 2023 Feb 22;14(1):993. doi:10.1038/s41467-023-36637-3

13. Cao J, Li C, Cui Z, Deng S, Lei T, Liu W, et al. Spatial Transcriptomics: A Powerful Tool in Disease Understanding and Drug Discovery. Theranostics. 2024;14(7):2946–68. doi:10.7150/thno.95908

14. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods. 2021 Nov;18(11):11. doi:10.1038/s41592-021-01255-8 PubMed PMID: 34711970.

15. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017 Nov;171(6):1437-1452.e17. doi:10.1016/j.cell.2017.10.049

16. Zhang J, Ubas AA, de Borja R, Svensson V, Thomas N, Thakar N, et al. *Tahoe-100M*: A Giga-Scale Single-Cell Perturbation Atlas for Context-Dependent Gene Function and Cellular Modeling [Internet]. 2025 [cited 2025 Aug 28]. Available from: http://biorxiv.org/lookup/doi/10.1101/2025.02.20.639398 doi:10.1101/2025.02.20.639398

17. OpenAI, Hurst A, Lerer A, Goucher AP, Perelman A, Ramesh A, et al. GPT-4o System Card [Internet]. arXiv; 2024 [cited 2025 May 29]. Available from: http://arxiv.org/abs/2410.21276 doi:10.48550/arXiv.2410.21276

18. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. 2016 [cited 2025 May 29]. p. 785–94. Available from: http://arxiv.org/abs/1603.02754 doi:10.1145/2939672.2939785

19. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug Target Identification Using Side-Effect Similarity. Science. 2008 Jul 11;321(5886):263–6. doi:10.1126/science.1158140

20. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2012 Nov 22;41(D1):D955–61. doi:10.1093/nar/gks1111

21. 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 Dec;16(12):12. doi:10.1038/s41592-019-0619-0 PubMed PMID: 31740819; PubMed Central PMCID: PMC6884693.

22. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. GigaScience. 2018 Jul 1;7(7):giy083. doi:10.1093/gigascience/giy083

23. Myronenko A, Song X. Point-Set Registration: Coherent Point Drift. IEEE Trans Pattern Anal Mach Intell. 2010 Dec;32(12):2262–75. doi:10.1109/TPAMI.2010.46

24. Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nat Commun. 2022 Apr 1;13(1):1739. doi:10.1038/s41467-022-29439-6

25. Zhou X, Dong K, Zhang S. Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. Nat Comput Sci. 2023 Oct 12;3(10):894–906. doi:10.1038/s43588-023-00528-w

26. Xu H, Wang S, Fang M, Luo S, Chen C, Wan S, et al. SPACEL: deep learning-based characterization of spatial transcriptome architectures. Nat Commun. 2023 Nov 22;14(1):7603. doi:10.1038/s41467-023-43220-3

27. Long Y, Ang KS, Li M, Chong KLK, Sethi R, Zhong C, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. Nat Commun. 2023 Mar 1;14(1):1155. doi:10.1038/s41467-023-36796-3

28. Torres A, Darrigrand F, Herrera G, Goss G, Willnow D, Salowka A, et al. Spatially organized cellular communities shape functional tissue architecture in the pancreas. C E Nc E V Anc E S. 2025.

29. Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics [Internet]. 2023 [cited 2024 Aug 27]. Available from: http://biorxiv.org/lookup/doi/10.1101/2023.11.05.565674 doi:10.1101/2023.11.05.565674

30. Wu R, Guo W, Qiu X, Wang S, Sui C, Lian Q, et al. Comprehensive analysis of spatial architecture in primary liver cancer. Sci Adv. 2021 Dec 17;7(51):51. doi:10.1126/sciadv.abg3750 PubMed PMID: 34919432; PubMed Central PMCID: PMC8683021.

31. Erickson A, He M, Berglund E, Marklund M, Mirzazadeh R, Schultz N, et al. Spatially resolved clonal copy number alterations in benign and malignant tissue. Nature. 2022 Aug 11;608(7922):360–7. doi:10.1038/s41586-022-05023-2

32. Mousa A. Sorafenib in the treatment of advanced hepatocellular carcinoma. Saudi J Gastroenterol. 2008;14(1):40. doi:10.4103/1319-3767.37808

33. Nakano M, Tanaka M, Kuromatsu R, Nagamatsu H, Tajiri N, Satani M, et al. Sorafenib for the treatment of advanced hepatocellular carcinoma with extrahepatic metastasis: a prospective multicenter cohort study. Cancer Med. 2015 Dec;4(12):1836–43. doi:10.1002/cam4.548

34. Zhong C, Lu C, Zhang X, Zhang E, Lai F, Zhong X, et al. Bosutinib Targets Hepatocellular Carcinoma as a Mitochondrial Complex I Inhibitor [Internet]. 2024. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4978152

35. Yang Y, Pei T, Hu X, Lu Y, Huang Y, Wan T, et al. Dietary vitamin B3 supplementation induces the antitumor immunity against liver cancer via biased GPR109A signaling in myeloid cell. Cell Rep Med. 2024 Sep;5(9):101718. doi:10.1016/j.xcrm.2024.101718

36. Moreno-Ceballos M, Cortes-Mancera FM, Moshage H, Arroyave-Ospina JC. Coffee Extracts and Chlorogenic Acid Inhibit the Proliferation of HepG2 Cells and c-Myc Expression Without Significant Modulation of Wnt/β-Catenin Signaling. Livers. 2025 Oct 15;5(4):49. doi:10.3390/livers5040049

37. Shan L, Zhao N, Wang F, Zhai D, Liu J, Lv X. Caffeine in Hepatocellular Carcinoma: Cellular Assays, Animal Experiments, and Epidemiological Investigation. J Inflamm Res. 2024 Mar;Volume 17:1589–605. doi:10.2147/JIR.S424384

38. Okano J, Nagahara T, Matsumoto K, Murawaki Y. Caffeine Inhibits the Proliferation of Liver Cancer Cells and Activates the MEK/ERK/EGFR Signalling Pathway. Basic Clin Pharmacol Toxicol. 2008 Jun;102(6):543–51. doi:10.1111/j.1742-7843.2008.00231.x

39. Bravi F, Bosetti C, Tavani A, Gallus S, La Vecchia C. Coffee Reduces Risk for Hepatocellular Carcinoma: An Updated Meta-analysis. Clin Gastroenterol Hepatol. 2013 Nov;11(11):1413-1421.e1. doi:10.1016/j.cgh.2013.04.039

40. Yu C, Cao Q, Chen P, Yang S, Deng M, Wang Y, et al. An updated dose–response meta-analysis of coffee consumption and liver cancer risk. Sci Rep. 2016 Dec 2;6(1):37488. doi:10.1038/srep37488

41. Bhurwal A, Ratta P, Yoshitake S, Pioppo L, Reja D, Dellatore P, et al. Inverse Association of Coffee with Liver Cancer Development: An Updated Systematic Review and Meta-analysis. J Gastrointestin Liver Dis. 2020 Sep 9;29(3):421–8. doi:10.15403/jgld-805

42. Yu J, Liang D, Li J, Liu Z, Zhou F, Wang T, et al. Coffee, Green Tea Intake, and the Risk of Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis of Observational Studies. Nutr Cancer. 2023 May 28;75(5):1295–308. doi:10.1080/01635581.2023.2178949

43. Simon TG, Bonilla H, Yan P, Chung RT, Butt AA. Atorvastatin and fluvastatin are associated with dose-dependent reductions in cirrhosis and hepatocellular carcinoma, among patients with hepatitis C virus: Results from ERCHIVES. Hepatology. 2016 Jul;64(1):47–57. doi:10.1002/hep.28506

44. Wang Y, Wang W, Wang M, Shi J, Jia X, Dang S. A Meta-Analysis of Statin Use and Risk of Hepatocellular Carcinoma. Granito A, editor. Can J Gastroenterol Hepatol. 2022 Mar 20;2022:1–15. doi:10.1155/2022/5389044

45. Chiu DK, Xu IM, Lai RK, Tse AP, Wei LL, Koh H, et al. Hypoxia induces myeloid-derived suppressor cell recruitment to hepatocellular carcinoma through chemokine (C-C motif) ligand 26. Hepatology. 2016 Sep;64(3):797–813. doi:10.1002/hep.28655

46. Zheng RP, Wang W, Wei CD. Bortezomib inhibits cell proliferation in prostate cancer. Exp Ther Med. 2015 Sep;10(3):1219–23. doi:10.3892/etm.2015.2617

47. Schweizer M. Vorinostat to Augment Response to 177Lutetium-PSMA-617 in the Treatment of Patients With PSMA-Low Metastatic Castration-Resistant Prostate Cancer [Internet]. 2023 Nov [cited 2024 Sep 18]. Report. Located at: NCT06145633. Available from: https://clinicaltrials.gov/study/NCT06145633

48. Imrali A, Mao X, Yeste-Velasco M, Shamash J, Lu Y. Rapamycin inhibits prostate cancer cell growth through cyclin D1 and enhances the cytotoxic efficacy of cisplatin. 2016 Aug 1.

49. Chen E, Liu N, Zhao Y, Tang M, Ou L, Wu X, et al. Panobinostat reverses HepaCAM gene expression and suppresses proliferation by increasing histone acetylation in prostate cancer. Gene. 2022 Jan;808:145977. doi:10.1016/j.gene.2021.145977

50. El-Zawahry A, McKillop J, Voelkel-Johnson C. Doxorubicin increases the effectiveness of Apo2L/TRAIL for tumor growth inhibition of prostate cancer xenografts. BMC Cancer. 2005 Jan 7;5(1):2. doi:10.1186/1471-2407-5-2

51. Chen D, Frezza M, Schmitt S, Kanwar J, P. Dou Q. Bortezomib as the First Proteasome Inhibitor Anticancer Drug: Current Status and Future Perspectives. Curr Cancer Drug Targets. 2011 Mar 1;11(3):239–53. doi:10.2174/156800911794519752

52. Moreau P, Pylypenko H, Grosicki S, Karamanesht I, Leleu X, Grishunina M, et al. Subcutaneous versus intravenous administration of bortezomib in patients with relapsed multiple myeloma: a randomised, phase 3, non-inferiority study. Lancet Oncol. 2011 May;12(5):431–40. doi:10.1016/s1470-2045(11)70081-x

53. Xue K, Gu JJ, Zhang Q, Mavis C, Hernandez-Ilizaliturri FJ, Czuczman MS, et al. Vorinostat, a histone deacetylase (HDAC) inhibitor, promotes cell cycle arrest and re-sensitizes rituximab- and chemo-resistant lymphoma cells to chemotherapy agents. J Cancer Res Clin Oncol. 2016 Feb;142(2):379–87. doi:10.1007/s00432-015-2026-y

54. Kelly WK, O'Connor OA, Krug LM, Chiao JH, Heaney M, Curley T, et al. Phase I Study of an Oral Histone Deacetylase Inhibitor, Suberoylanilide Hydroxamic Acid, in Patients With Advanced Cancer. J Clin Oncol. 2005 Jun 10;23(17):3923–31. doi:10.1200/jco.2005.14.167

55. Kleshchevnikov V, Shmatko A, Dann E, Aivazidis A, King HW, Li T, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. Nat Biotechnol. 2022 May;40(5):5. doi:10.1038/s41587-021-01139-4

56. Nguyen V, Ahler E, Sitko KA, Stephany JJ, Maly DJ, Fowler DM. Molecular determinants of Hsp90 dependence of Src kinase revealed by deep mutational scanning. Protein Sci. 2023 Jul;32(7):e4656. doi:10.1002/pro.4656

57. Mshaik R, Simonet J, Georgievski A, Jamal L, Bechoua S, Ballerini P, et al. HSP90 inhibitor NVP-BEP800 affects stability of SRC kinases and growth of T-cell and B-cell acute lymphoblastic leukemias. Blood Cancer J. 2021 Mar 18;11(3):61. doi:10.1038/s41408-021-00450-2

58. Sogbein O, Paul P, Umar M, Chaari A, Batuman V, Upadhyay R. Bortezomib in cancer therapy: Mechanisms, side effects, and future proteasome inhibitors. Life Sci. 2024 Dec;358:123125. doi:10.1016/j.lfs.2024.123125

59. Hideshima T, Anderson KC. Biologic Impact of Proteasome Inhibition in Multiple Myeloma Cells—From the Aspects of Preclinical Studies. Semin Hematol. 2012 Jul;49(3):223–7. doi:10.1053/j.seminhematol.2012.04.006

60. Jin R, Yamashita H, Yu X, Wang J, Franco OE, Wang Y, et al. Inhibition of NF-kappa B signaling restores responsiveness of castrate-resistant prostate cancer cells to anti-androgen treatment by decreasing androgen receptor-variant expression. Oncogene. 2015 Jul;34(28):3700–10. doi:10.1038/onc.2014.302

61. Meng Y, Mo L, Liang T, Mo S, Han C, He Y. Atorvastatin negatively regulates MAPK pathway in vitro to inhibit proliferation, migration, and invasion of hepatocellular carcinoma cells. Sci Rep. 2025 Nov 21;15(1):41279. doi:10.1038/s41598-025-25040-1

62. Dou Q, Zonder J. Overview of Proteasome Inhibitor-Based Anti-cancer Therapies: Perspective on Bortezomib and Second Generation Proteasome Inhibitors versus Future Generation Inhibitors of Ubiquitin-Proteasome System. Curr Cancer Drug Targets. 2014 Aug 31;14(6):517–36. doi:10.2174/1568009614666140804154511

63. Kuroda K, Liu H. The proteasome inhibitor, bortezomib, induces prostate cancer cell death by suppressing the expression of prostate-specific membrane antigen, as well as androgen receptor. Int J Oncol. 2019 Feb 1. doi:10.3892/ijo.2019.4706

64. Parveen R, Harihar D, Chatterji BP. Recent histone deacetylase inhibitors in cancer therapy. Cancer. 2023 Nov;129(21):3372–80. doi:10.1002/cncr.34974

65. Richon VM. Cancer biology: mechanism of antitumour action of vorinostat (suberoylanilide hydroxamic acid), a novel histone deacetylase inhibitor. Br J Cancer. 2006 Dec;95(S1):S2–6. doi:10.1038/sj.bjc.6603463

66. Shi CX, Zhu YX, Bruins LA, Bonolo de Campos C, Stewart W, Braggio E, et al. Proteasome Subunits Differentially Control Myeloma Cell Viability and Proteasome Inhibitor Sensitivity. Mol Cancer Res. 2020 Oct 1;18(10):1453–64. doi:10.1158/1541-7786.MCR-19-1026

67. Hideshima T, Ikeda H, Chauhan D, Okawa Y, Raje N, Podar K, et al. Bortezomib induces canonical nuclear factor-κB activation in multiple myeloma cells. Blood. 2009 Jul 30;114(5):1046–52. doi:10.1182/blood-2009-01-199604

68. Podar K, Gouill SL, Zhang J, Opferman JT, Zorn E, Tai YT, et al. A pivotal role for Mcl-1 in Bortezomib-induced apoptosis. Oncogene. 2008 Jan 31;27(6):721–31. doi:10.1038/sj.onc.1210679

69. Pakos-Zebrucka K, Koryga I, Mnich K, Ljujic M, Samali A, Gorman AM. The integrated stress response. EMBO Rep. 2016 Oct;17(10):1374–95. doi:10.15252/embr.201642195

70. Mahboubi H, Moujaber O, Kodiha M, Stochaj U. The Co-Chaperone HspBP1 Is a Novel Component of Stress Granules that Regulates Their Formation. Cells. 2020 Mar 29;9(4):825. doi:10.3390/cells9040825

71. Schitcu VH, Raduly L, Zanoaga O, Jurj A, Munteanu VC, Budisan L, et al. TP53 gene implications in prostate cancer evolution: potential role in tumor classification. Med Pharm Rep. 2023 Jul 14;96(4):384–91. doi:10.15386/mpr-2639

72. Williams S, Pettaway C, Song R, Papandreou C, Logothetis C, McConkey DJ. Differential effects of the proteasome inhibitor bortezomib on apoptosis and angiogenesis in human prostate tumor xenografts. 2003 Oct 10.

73. Glaviano A, Foo ASC, Lam HY, Yap KCH, Jacot W, Jones RH, et al. PI3K/AKT/mTOR signaling transduction pathway and targeted therapies in cancer. Mol Cancer. 2023 Aug 18;22(1):138. doi:10.1186/s12943-023-01827-6

74. Fajardo A, Piazza G, Tinsley H. The Role of Cyclic Nucleotide Signaling Pathways in Cancer: Targets for Prevention and Treatment. Cancers. 2014 Feb 26;6(1):436–58. doi:10.3390/cancers6010436

75. Xu J. The role of tumor necrosis factor receptor superfamily in cancer: insights into oncogenesis, progression, and therapeutic strategies. Npj Precis Oncol. 2025 Aug 6;9(1):275. doi:10.1038/s41698-025-00990-x

76. Zhang X, Li B, Lan T, Chiari C, Ye X, Wang K, et al. The role of interleukin-17 in inflammation-related cancers. Front Immunol. 2025 Jan 21;15:1479505. doi:10.3389/fimmu.2024.1479505

77. Oerlemans R, Franke NE, Assaraf YG, Cloos J, Van Zantwijk I, Berkers CR, et al. Molecular basis of bortezomib resistance: proteasome subunit β5 (PSMB5) gene mutation and overexpression of PSMB5 protein. Blood. 2008 Sep 15;112(6):2489–99. doi:10.1182/blood-2007-08-104950

78. Lashinger LM, Zhu K, Williams SA, Shrader M, Dinney CPN, McConkey DJ. Bortezomib Abolishes Tumor Necrosis Factor–Related Apoptosis-Inducing Ligand Resistance via a p21-Dependent Mechanism in Human Bladder and Prostate Cancer Cells. Cancer Res. 2005 Jun 1;65(11):4902–8. doi:10.1158/0008-5472.CAN-04-3701

79. Hu Y, Sun H, Owens RT, Wu J, Chen YQ, Berquin IM, et al. Decorin Suppresses Prostate Tumor Growth through Inhibition of Epidermal Growth Factor and Androgen Receptor Pathways. Neoplasia. 2009 Oct;11(10):1042–53. doi:10.1593/neo.09760

80. Xia Z, Du D, Zhang Z, Liu Z, Hu Z, Li X, et al. WIF1 and DKK3 in prostate cancer: from molecular pathways to therapeutic targets: a narrative review. Transl Androl Urol. 2024 Nov;13(11):2601–16. doi:10.21037/tau-24-304

81. Truong TH, Lange CA. Deciphering Steroid Receptor Crosstalk in Hormone-Driven Cancers. Endocrinology. 2018 Dec 1;159(12):3897–907. doi:10.1210/en.2018-00831

82. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024 Feb;42(2):293–304. doi:10.1038/s41587-023-01767-y

83. 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. The Innovation. 2021 Aug;2(3):100141. doi:10.1016/j.xinn.2021.100141

84. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30. doi:10.1093/nar/28.1.27

85. Ma L, Heinrich S, Wang L, Keggenhoff FL, Khatib S, Forgues M, et al. Multiregional single-cell dissection of tumor and immune cells reveals stable lock-and-key features in liver cancer. Nat Commun. 2022 Dec 7;13(1):7533. doi:10.1038/s41467-022-35291-5

86. Lu Y, Yang A, Quan C, Pan Y, Zhang H, Li Y, et al. A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. Nat Commun. 2022 Aug 6;13(1):4594. doi:10.1038/s41467-022-32283-3

87. Hirz T, Mei S, Sarkar H, Kfoury Y, Wu S, Verhoeven BM, et al. Dissecting the immune suppressive human prostate tumor microenvironment via integrated single-cell and spatial transcriptomic analyses. Nat Commun. 2023 Feb 7;14(1):663. doi:10.1038/s41467-023-36325-2

88. Gatti AA, Khallaghi S. PyCPD: Pure NumPy Implementation of the Coherent PointDrift Algorithm. J Open Source Softw. 2022 Dec 15;7(80):4681. doi:10.21105/joss.04681

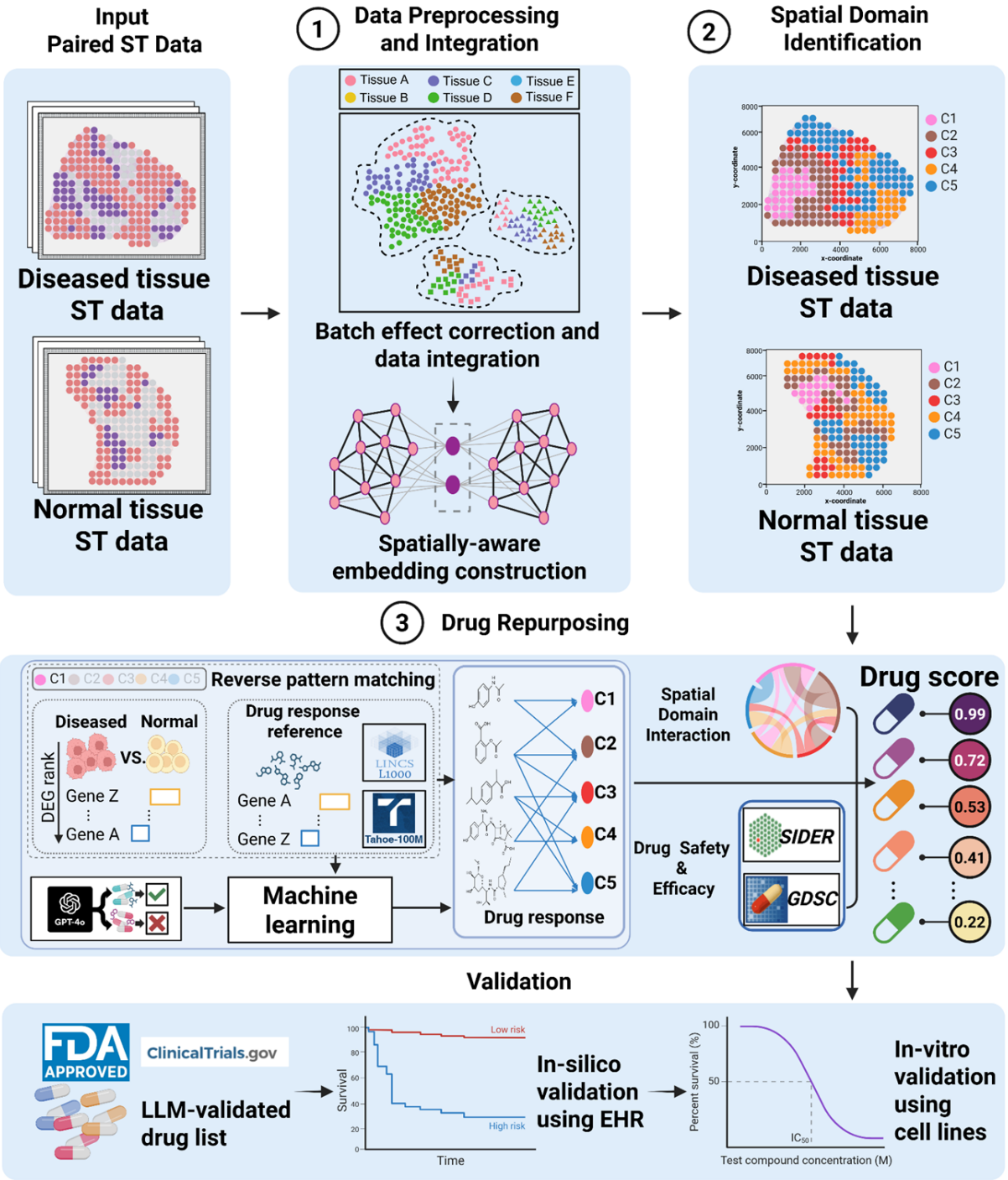
89. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47–e47. doi:10.1093/nar/gkv007

90. Chan J, Wang X, Turner JA, Baldwin NE, Gu J. Breaking the paradigm: Dr Insight empowers signature-free, enhanced drug repurposing. Bioinforma Oxf Engl. 2019 Aug 15;35(16):16. doi:10.1093/bioinformatics/btz006 PubMed PMID: 30624606; PubMed Central PMCID: PMC6691331.

91. Takeda M, Mizokami A, Mamiya K, Li YQ, Zhang J, Keller ET, et al. The establishment of two paclitaxel-resistant prostate cancer cell lines and the mechanisms of paclitaxel resistance with two cell lines. The Prostate. 2007 Jun 15;67(9):955–67. doi:10.1002/pros.20581

92. IC50 calculation using GraphPad Prism version 10.0.0 for Windows [Internet]. GraphPad Software; [cited 2025 Jul 11]. Available from: www.graphpad.com

Figures



771 **Figure 1: Illustration of STDrug architecture.** STDrug utilizes paired diseased and normal tissues as the
772 input. Data preprocessing, batch correction, and sample alignment are performed initially, before spatially-
773 aware embeddings are constructed to identify corresponding spatial domain pairs between the diseased and
774 normal conditions. These paired spatial domains then serve as inputs for the drug repurposing module,
775 which identifies potentially reversible disease-associated genes determined as differentially expressed
776 genes (DEGs) between spatial domains in diseased vs normal conditions. These disease-associated genes
777 are used as the "queries" to search candidate drugs that may reverse these genes expression in the
778 pharmacogenomics perturbation data from the L1000 and Tahoe-100M single-cell RNA-seq datasets. To
779 prioritize key reversible disease-associated genes, STDrug assigns them weights learned from a machine
780 learning model, where the gene-disease association was obtained from a pre-trained GPT-4o model on
781 existing literature. Additionally, STDrug accounts for spatial domain interactions, drug side effect profiles
782 from SIDER and sensitivity data from GDSC in calculating the comprehensive drug score for each
783 candidate drug. The top-ranked drugs are then validated using a combination of evidence, including existing
784 evidence from literature and clinical trials, in-silico validation using large-scale clinical data, and in-vitro
785 validation from cell line experiments.
786

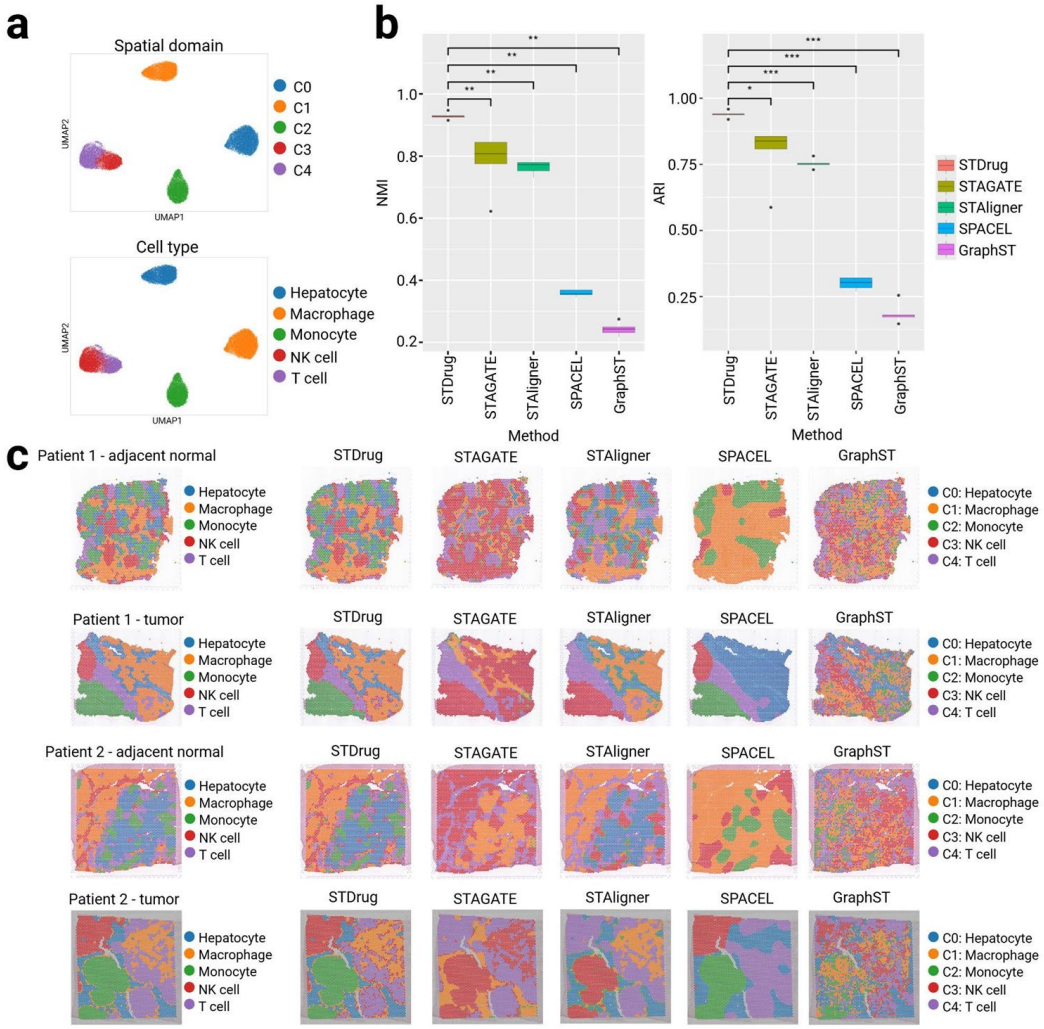


Figure 2: Spatial domain identification and benchmarking. a, UMAP plots of STDrug-identified domains (top) and cell types in the domains (bottom) in the simulated ST dataset. b, NMI and ARI of different spatial domain identification methods. Asterisks represent the significance of the statistical test (p-value), *** = 0.001, ** = 0.01, * = 0.05. c, Visualization of spatial domains estimated by different spatial domain identification methods on simulated spatial transcriptomics data from tumor and tumor-adjacent normal tissues from patients 1 and 2.

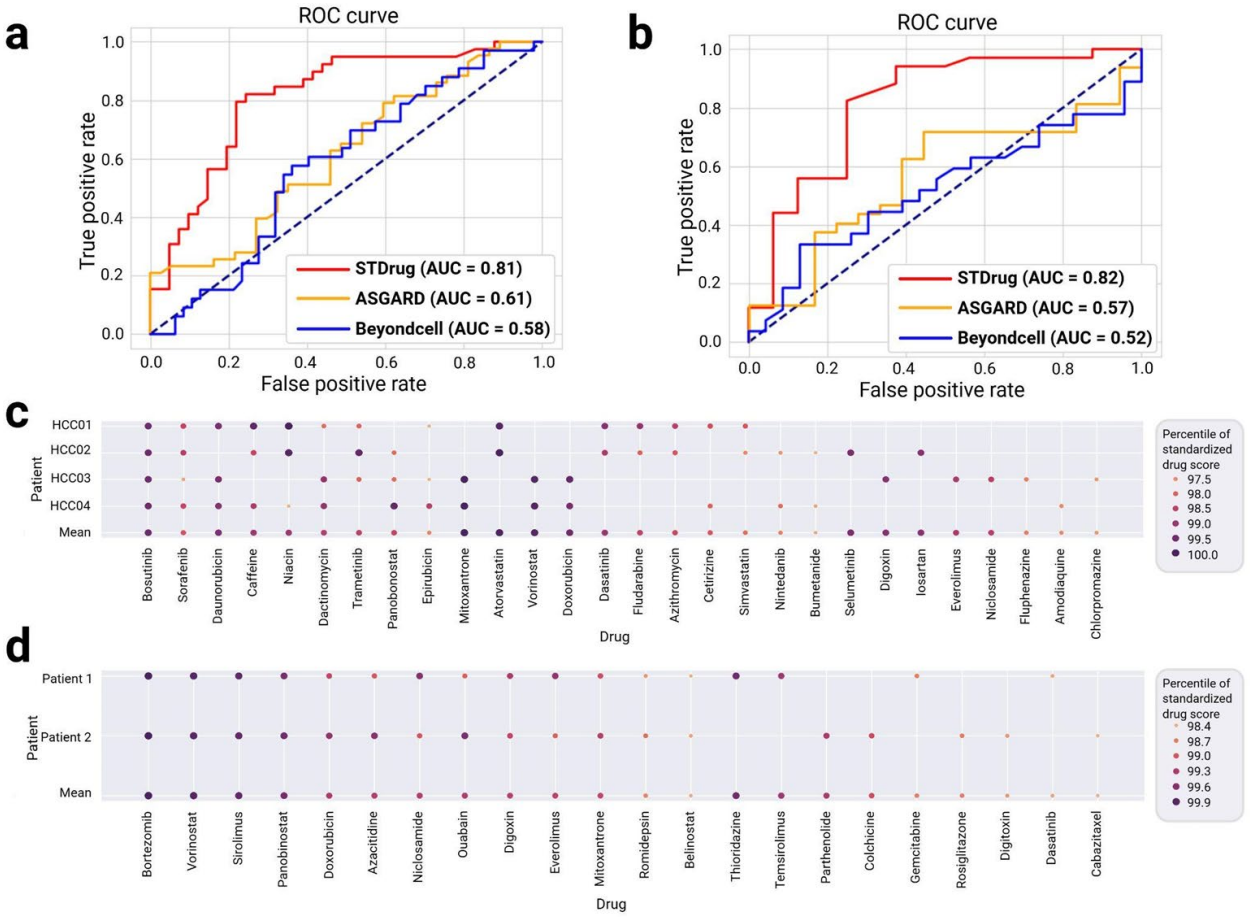
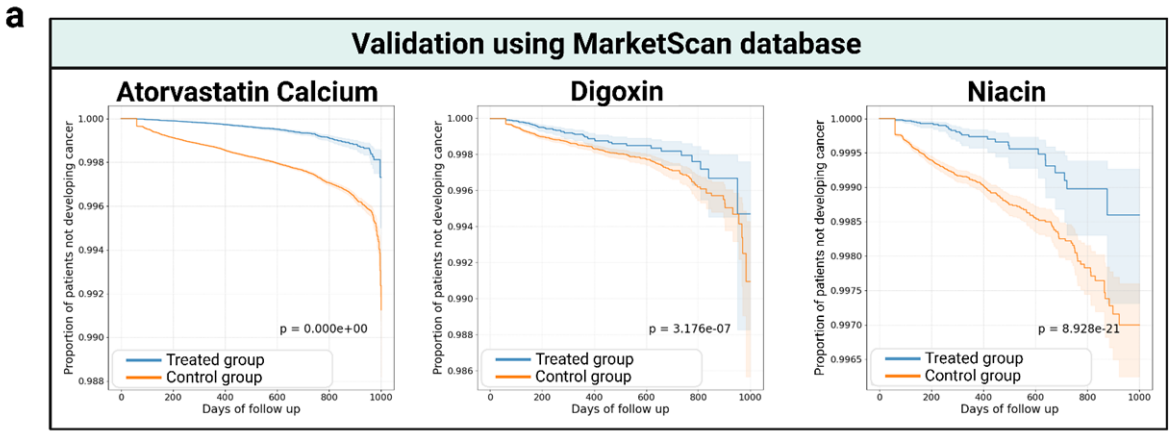
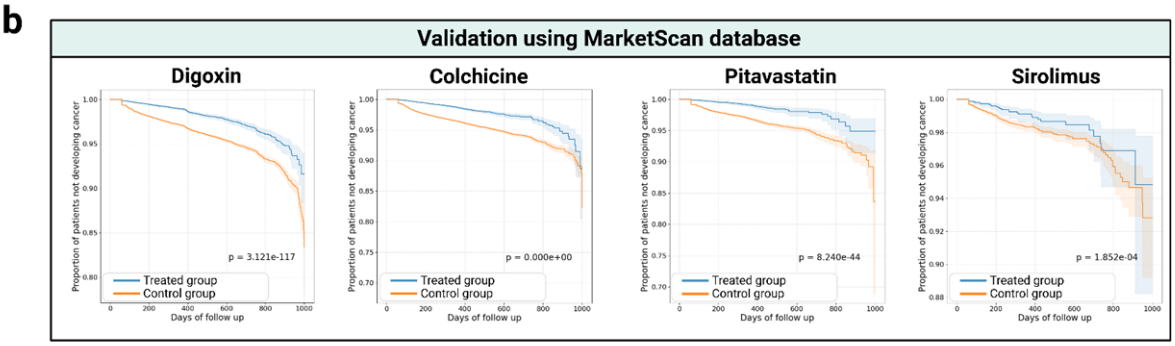
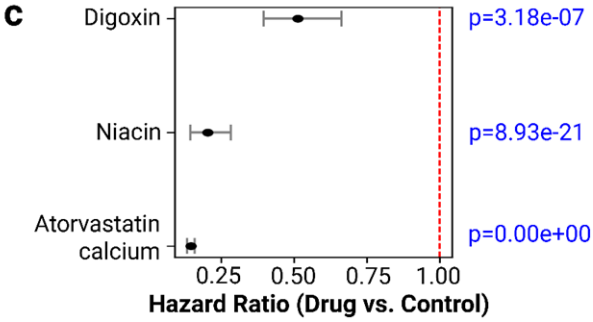
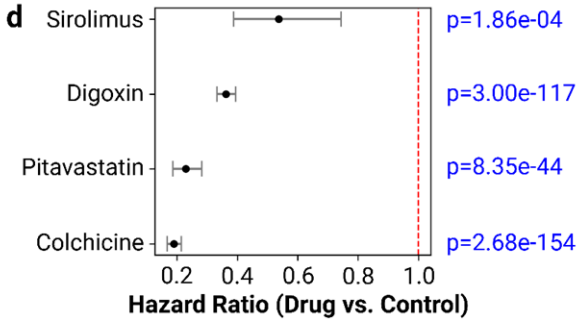


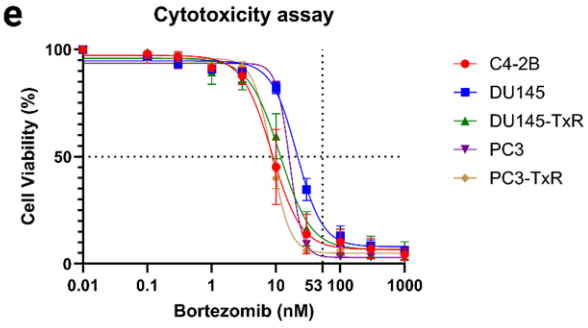
Figure 3: Drug repurposing and benchmarking. a-b, Evaluation of drug repurposing tools on HCC tissues (a) and PCa tissues (b). For each patient, only top 20 drugs are selected and ranked by normalized drug score percentages (%). For the averaged ranking across patients, drugs are ordered in descending order based on their frequency of occurrence among patients, followed by the mean percentile of normalized drug scores among only the patients who have the drugs enlisted as top 20. c, STDrug-identified top candidate drugs and compounds for HCC patients. d, STDrug-identified top candidate drugs and compounds for PCa patients.

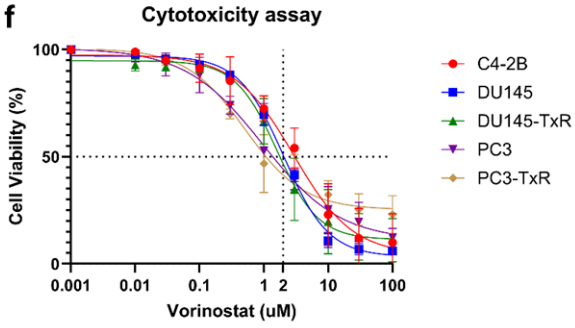












804

Figure 4: Real-world and in-vitro validation of top candidate drugs. a, Real-world clinical data validation of potential HCC drugs using Marketscan database. b, Real-world clinical data validation of potential Pca drugs from MarketScan database. c-d, Hazard Ratios (HR) of Cox proportional hazards models for the drug/compound candidates for (c) HCC and (d) PCa from Marketscan database. e-f, In-vitro validation of (e) bortezomib and (f) vorinostat in prostate cancer cell lines. The dotted horizontal line marks 50% reduction in cell viability, and the dotted vertical line represents clinical blood concentration of the drug.

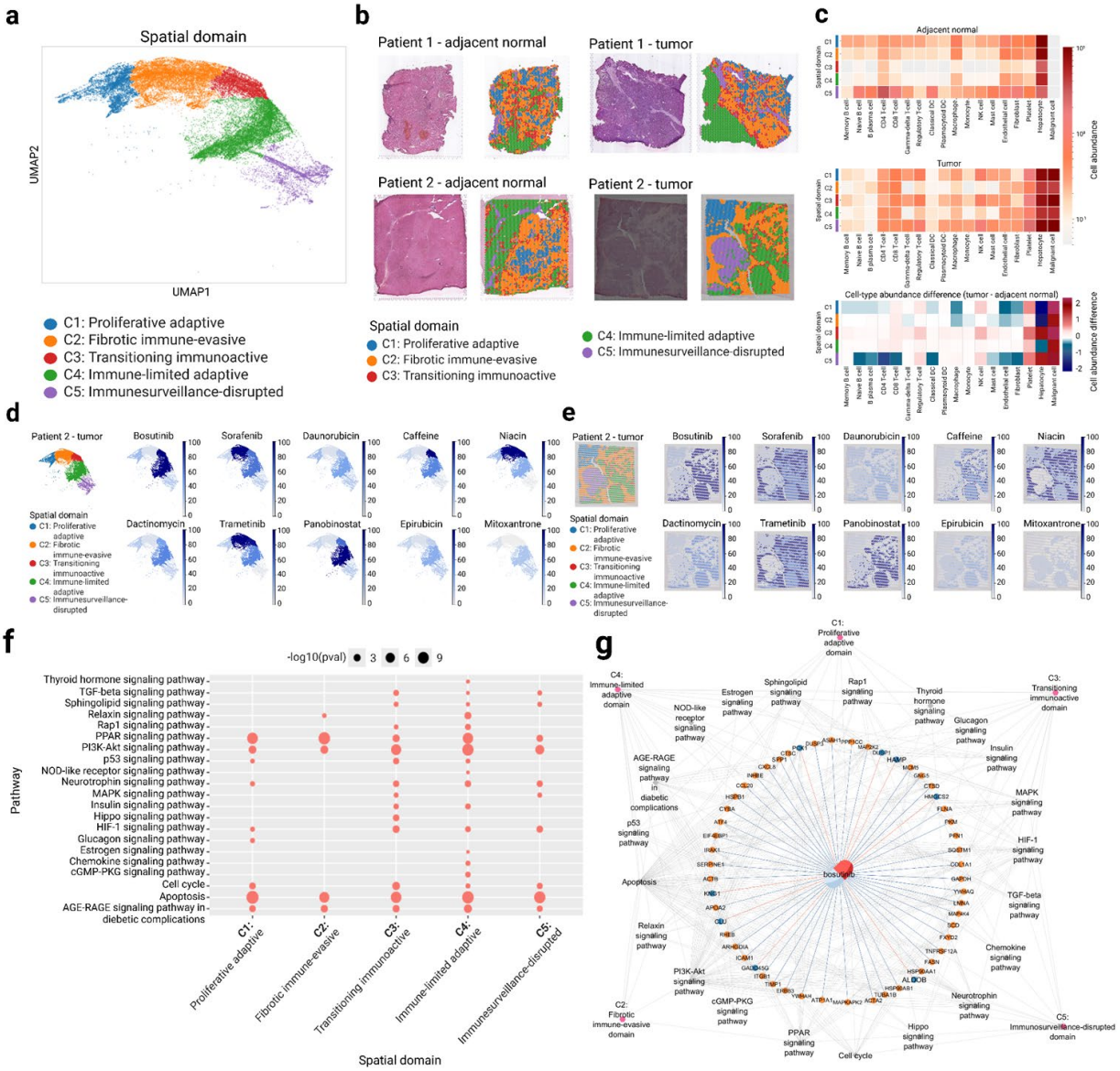


Figure 5: HCC functional downstream analysis. a, UMAP plot of HCC tissues colored by STDrug domains. b, Visualization of STDrug domains on tissues from patients 1 and 2. c, cell type deconvolution of HCC tumor and adjacent normal tissues. d-e, drug score visualization of top 10 drugs from Figure 3d on (d) UMAP plot and (e) spatial plot of patient 2. f. Target pathways of bosutinib in patient 2. g, The top drug candidate, bosutinib, its target genes, pathways, and domains. Orange node: up-regulated gene (logFC>1 and adjusted P-value < 0.05). Blue node: down-regulated gene (logFC < −1 and adjusted P-value < 0.05). Orange solid edge: drug stimulates gene expression. Blue solid edge: drug inhibits gene expression. The width of the edge is proportional to the strength of the drug effect. Gray dotted edge: gene belonging to a pathway. Gray backward slash: pathway significant in a corresponding domain.

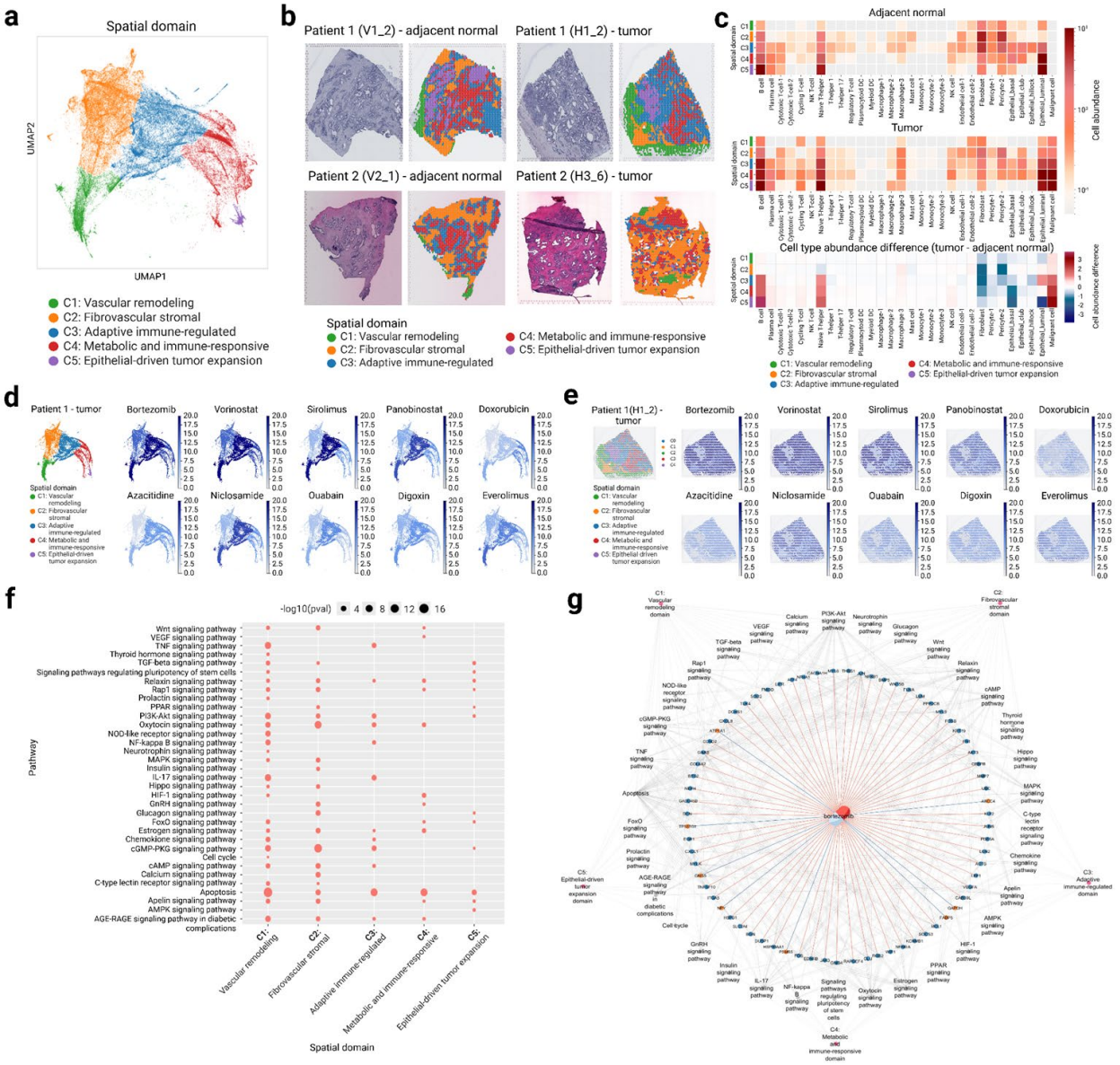


Figure 6: Prostate cancer functional downstream analysis. a, UMAP plot of prostate cancer tissues colored by STDrug domains. b, Visualization of STDrug domains on tissues from patients 1 and 2. c, cell type deconvolution of prostate tumor and adjacent normal tissues. d-e, drug score visualization on (d) UMAP plot and (e) spatial plot of patient 1. f, Target pathways of bortezomib in patient 1. g, The top drug candidate, bortezomib, its target genes, pathways, and domains. Orange node: up-regulated gene (logFC>1 and adjusted *P*-value < 0.05). Blue node: down-regulated gene (logFC < -1 and adjusted *P*-value < 0.05). Orange solid edge: drug stimulates gene expression. Blue solid edge: drug inhibits gene expression. The width of the edge is proportional to the strength of the drug effect. Gray dotted edge: gene belonging to a pathway. Gray backward slash: pathway significant in a corresponding domain.