Am J Cancer Res 2025;15(3):1096-1108
www.ajcr.us /ISSN:2156-6976/ajcr0159111

Original Article

Potential therapeutic targets for bladder cancer: a proteome-wide Mendelian randomization study

Jia-Hao Liu*, Yuan-Zhuo Du*, Fang Liu*, Lin Yang, Xiao-Qiang Liu, Bin Fu*, Xiao-Rong Yang*

*The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China. *Equal contributors.*

Received July 10, 2024; Accepted February 13, 2025; Epub March 15, 2025; Published March 30, 2025

Abstract: The incidence of bladder cancer (BCa) is increasing worldwide and the development of drug targets for BCa is necessary. We conducted a proteome-wide association study (PWAS) mainly using mendelian randomization (MR) to explore the causal proteins associated with BCa. Protein quantitative trait locis (pQTLs) were derived from two large proteome genome-wide association studies. After validation by multiple sensitivity analysis and two replication analyses, we identified five plasma proteins showed significant causal associations with BCa. Our study indicated that GSTM4 (OR = 0.81 (0.74-0.89), $P = 5.14 \times 10^{-6}$, PPH4 = 0.89) emerged as the most reliable target. Besides, PSCA, LY6D, SLURP1 and GSTM1 also showed clear causal association but only failed in colocalization. We also performed several downstream analyses. Protein-protein interactions analysis found these causal targets came from glutathione S-transferase family or lymphocyte antigen-6 family. Phenome-wide MR analysis revealed PSCA may lead to peptic ulcer and local infections of skin and subcutaneous tissue. We then employed single-cell analysis, protein-protein interactions, and druggability evaluation. Phenome-wide MR analysis was to assess the possible side effects of these drug targets. Finally, the reliability of GSTM4 in BCa was confirmed via colony formation assay.

Keywords: Bladder cancer, mendelian randomization, drug targets, biomarker

Introduction

Bladder cancer (BCa) is a common malignant tumor of the urinary system. In 2020, there were 573,278 new cases of BCa globally, with 212,536 deaths [1]. After the initial treatment of BCa, if not diagnosed and monitored in a timely manner, it often recurs and may progress. There is an urgent need for further evidence on non-invasive early diagnostic biomarkers and the development of new therapeutic targets for BCa.

Human proteins represent the primary category of drug targets [2]. Due to active secretion or cellular leakage, proteins may appear in circulation, providing the potential for a more precise understanding of current human health status [3]. From a genetic perspective, conducting in-depth exploration of the relationship between potential protein drug targets and diseases can help reduce the failure rate and cost of drug research [4]. With advancements in high-throughput genomic and proteomic technologies in plasma, the identification of potential therapeutic targets for diseases based on Mendelian randomization (MR) is becoming feasible across an increasing range of phenotypes.

MR analysis is a tool for causal inference in the context of the rapid development of large-scale GWAS and MR was used as the main research method in this study [5]. There are three crucial assumptions underlying MR analysis. Firstly, the genetic variants designated as instrumental variables need to exhibit a strong and reliable association with the exposure under study. Secondly, these genetic variants must not have any associations with potential confounding factors. Thirdly, the chosen genetic variants should influence the risk of the outcome strictly through the target risk factor and not by any alternative routes [6]. Using genetic variables as instrumental variables for exposure, MR can strengthen causal inference [7]. Compared to observational studies, the MR analysis can minimize residual confounding because the genetic variants will be randomly assorted at concep-

https://doi.org/10.62347/UBEJ3345

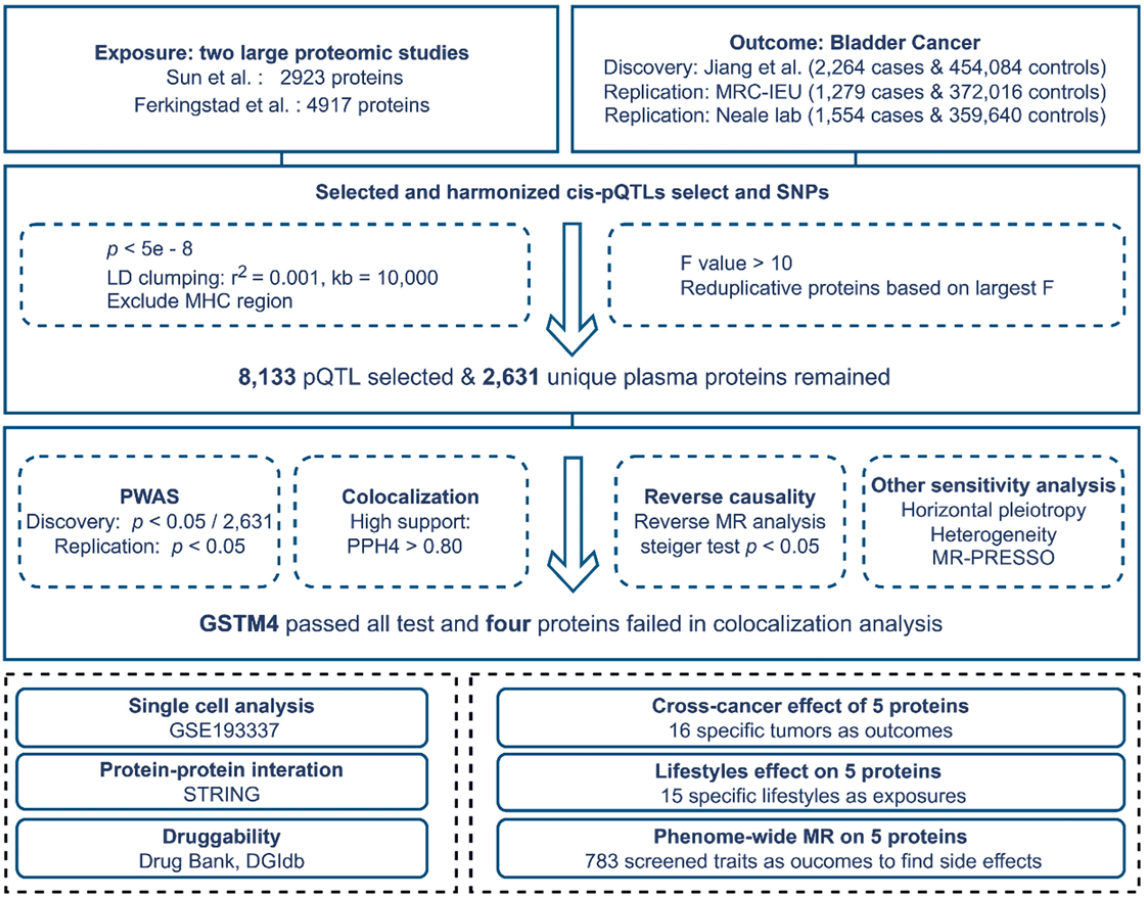


Figure 1. Overview of our study. MR: Mendelian randomization; PRESSO: Pleiotropy residual sum and outlier; PPH4: Posterior probability of hypothesis 4; PWAS: Proteome-wide association study.

tion and therefore uncorrelated with important confounders [8]. However, MR studies integrating GWAS and protein quantitative trait loci (pQTL) data on BCa are scarcely reported.

In this study, we combined two large human plasma proteomic studies with proteome-wide MR analysis to systematically identify circulating protein biomarkers associated with BCa risk.

Materials and methods

The overall design of this study is shown in **Figure 1**. Firstly, we assessed the causal association between plasma proteins and BCa using proteome-wide MR analysis and Bonferroni correction. We then check the reliability by two replication analyses. Bayesian colocalization test, MR pleiotropy residual sum and outlier (MR PRESSO) test, heterogeneity analysis, pleiotropy analysis, reverse MR, and steiger filtering test will further serve as sensitivity analysis methods. Besides, single-cell analysis, protein-protein interaction (PPI) analysis and druggability for these significant targets will be performed. We also investigated the causal associations between these targets and other phenotypes.

Study population and data source

All data used in this study were of European ethnic origin. As this study was based on publicly available summary statistics, ethical approval was not required. The GWAS data of BCa was obtained from Jiang L et al. (PheCode 189.21), which contains 2,264 cases and 454,084 controls performed by fast genome-wide association (GWA) methods, generalized linear mixed model (GLMM)-based methods model [9]. GWASs for replication were obtained from MRCIEU Consortium (1,279 cases and 372,016 controls) and Neale lab Consortium

(1,554 cases and 359,640 controls, ICD10: C67) [10, 11]. Two large proteomic association studies were obtained from Sun et al. (54,219 samples) and Ferkingstad et al. (35,559 samples), containing 2,940 and 4,917 proteins GWAS data, respectively [12, 13]. Plasma protein levels of Ferkingstad et al. were measured using the SomaScan multiplex aptamer assay and expressed in standardized units of standard deviation (SD) [13]. Plasma protein levels of Sun et al. were measured using the antibody-based proximity extension assay (PEA) on the Olink Explore 3072 platform and expressed in normalized protein expression (NPX) units [12]. The Ensembl Gene ID for each plasma protein can be found in Supplementary Table 2. Moreover, the GWAS studies of 16 types of pan-cancer analysis and 15 types of lifestyle influencing factors were shown in Supplementary Table 1.

Proteome-wide MR analysis

First, we mapped SNPs to human genome Build 38 (NCBI GRCh38) for further analysis. We used the following criteria to select pQTLs: (i) Showed genome-wide significant association ($P < 5 \times 10^{-8}$). (ii) The LD clumping was set ($r^2 < 0.001$) to identify independent pQTLs for each protein. (iii) For the pQTL, the SNPs and proteins were located outside the major histocompatibility complex (MHC) region (chr6, 26-34 Mb) and the pQTL was a cis-acting pQTL. (iv) To ensure the statistical effectiveness of IV, we only retained IVs with F-statistic > 10 ($F = R^2 \times (N-2)/(1-R^2)$, $R^2 = 2 \times EAF \times (1 - EAF) \times beta^2$) [14]. F-statistics can evaluate the statistical power of each genetic instrument and R^2 was the proportion of the variability of the protein levels explained by each genetic instrument. For duplicate proteins, the protein with the largest sum of F-statistics was selected. In our research, we defined cis-acting pQTL as the leading SNP in the region that was located within 1 Mb of the transcription start site of the protein-coding gene. Finally, a total of 8,133 instruments (cis pQTLs) and 2,631 unique plasma proteins were included in the analysis. Instrument variables were presented in Supplementary Table 2.

In this study, the 'TwoSampleMR' package (version 0.5.6) and R software (version 4.3.1) were used to perform proteome MR analysis [10]. We used the default parameters of 'TwoSampleMR' package to process GWAS data. Odds ratio (OR) with a 95% confidence interval (95% CI) was used to evaluate the risk effect. We mainly performed the inverse-variance weighted (IVW) method to evaluate the causal association between plasma proteins and BCa. Besides, The Wald ratio method was also used for plasma proteins containing only one pQTL as a genetic instrument. Multiple testing correction was corrected by the Bonferroni method. We also performed additional analyses including weighted median, MR-Egger, simple mode, and weighted mode to account for horizontal pleiotropy. These analyses can account for heterogeneity and horizontal pleiotropy to an extent [15]. Replication MR analysis was performed for the significantly identified proteins based on the GWAS from MRC-IEU consortium and Neale Lab consortium, respectively. *P*-value < 0.05 was defined as the significance level for replication.

Sensitivity analysis for proteome MR

If the number of IVs is sufficient, a heterogeneity test and horizontal pleiotropy test will be performed. $P > 0.05$ was considered to pass the heterogeneity test and horizontal pleiotropy test, respectively. MR-PRESSO global test was also used to evaluate the horizontal pleiotropy. If find evidence of horizontal pleiotropy, we will conduct the MR-PRESSO outlier test and provide the estimate of causal effect again after the removal of outliers by IVW analysis.

Bidirectional MR analysis can detect potential reverse causality. Following the criteria of pQTLs, 5 genetic instruments for BCa were selected from GWAS summary data of Jiang L et al. Genetic instruments for BCa are presented in Supplementary Table 3. We mainly used the IVW method to assess the causal association between BCa and plasma proteins. Besides, steiger test was used to assess the directionality of each genetic instrument and $P < 0.05$ was considered a stronger effect of instrumental variables on exposure than on outcomes.

Bayesian colocalization analysis

Bayesian colocalization analysis was employed to test whether plasma proteins and BCa share the same causal variant [16]. Colocalization was executed with default parameters ($p1 = 1$

$\times 10^{-4}$, $p2 = 1 \times 10^{-4}$, $p12 = 1 \times 10^{-5}$), and SNPs within ±500 kb of the pQTL were selected for analysis. When multiple pQTLs existed for a plasma protein, colocalization analyses were performed on each pQTL separately and the pQTL with the strongest colocalization analyses was displayed. There were 5 types of colocalization analysis scenarios. In our study, we only analyzed the posterior probability of hypothesis 3 (PPH3) and the posterior probability of hypothesis 4 (PPH4). PPH3 was defined as plasma proteins and BCa is significantly associated with SNP loci in a given genomic region but is driven by different causal variant loci. PPH4 was defined as plasma proteins and BCa are significantly associated with SNP loci in a genomic region and driven by the same causal variant locus. PPH4 ≥ 80% was considered strong evidence of causal associations and a shared causal variant. We also defined proteins with PPH3 + PPH4 ≥ 80% as potential evidence [17]. The colocalization analysis was conducted using the coloc package (version 5.2.3) and the visualization of colocalized regions was facilitated by the 'LocusCompareR' package.

Single-cell analysis

The significant proteins discovered by Proteome MR were assessed by the cell-specific enrichment single-cell analysis. We used five BCa tissues from Gene Expression Omnibus (GEO) high throughput expression sequencing data as well as four adjacent normal tissues from GSE222315. The sequencing data from cancer and normal tissues was separately integrated and batch effects. The Seurat data of BCa tissues included 26,498 genes in 42,654 cells. We selected cells and genes with the following criteria: a gene detected in at least 3 cells was retained; a cell detected more than 200 but less than 5000 genes were retained; a cell with < 10% mitochondrial gene and with < 3% hemoglobin gene retained. The 'SingleR' package was used to annotate each cell type. Genes with an average Log_2 fold change (Log_2FC) more than 0.25 and a false discovery rate (FDR) adjusted *p*-value < 0.05 were considered as enrichment genes in a cell type.

Protein-protein interaction (PPI) and druggability evaluation

We used the STRING database to assess protein-protein interactions (https://string-db.org/). Minimum required interaction score was set at medium confidence (0.400). All other settings were set to default. The sources of associations were also shown in the result and can be divided into known interactions, predicted interactions, and others.

We assessed whether the identified genes or proteins could serve as potential therapeutic targets by DGIdb and DrugBank databases [18, 19]. We searched for drugs targeting these selected proteins. The information of drugs that targeted these causal proteins was documented, including "drug or component name", "drug groups", "action", and "indication". We prioritized the potential druggable targets by integrating the above information and focus on whether drugs have the potential to treat cancer patients.

Cross-cancer effect analysis and influence of lifestyles for the potential drug targets

We conducted two-sample MR analyses to evaluate the association between significant circulating proteins and 16 site-specific cancers (Supplementary Table 1). Besides, we conducted two-sample MR analyses of healthy lifestyle factors and significant circulating proteins to determine which BCa-associated proteins could be modulated by healthy lifestyle interventions. A total of 15 healthy lifestyle factors (Supplementary Table 1) were used to assess their association with five causal proteins. The analytical approach of MR was consistent with the description provided above.

Phenome-wide MR analysis

To investigate the potential side effects of these five significant druggable targets, we used summary statistics of diseases in the UK Biobank cohort (N ≤ 408961) as outcomes, and still used the pQTL of these five plasma proteins as IVs to perform phenome-wide MR via the IVW and Wald ratio method. We considered FDR < 0.05 as statistically significant causal effect. Scalable and Accurate Implementation of Generalised Mixed Model (SAIGE) method can adjust for sample correlation and case-control imbalance and produce accurate *p*-values in very large sample data [20]. Summary statistics of disease-associated SNPs were downloaded from the SAIGE GWAS (https://www.leelabsg.org/resources) [20]. To

ensure effective statistical power, we chose 783 traits with more than 500 cases for phenome-MR analyses. Further details are provided in Supplementary Table 8.

Cell lines and transfaction

BCa cell lines T24 cells and BIU cells were grown in Dulbecco's Modified Eagle Medium (DMEM) (Gibco, USA) and RPMI-1640 (Gibco, USA) containing 10% fetal bovine serum, 1% penicillin, and streptomycin. Cells grew in the environment at 37°C with 5% CO_2. T24 cells were sourced from Procell Life Science and Technology Co., Ltd. (Wuhan, China) and BIU were sourced from the cell bank of the Type Culture Collection of the Chinese Academy of Sciences (Shanghai, China). pCMV-GSTM4 (human)-3×HA-Neo was purchased from Miao-Ling Biology Co., Ltd. (Wuhan, China) and cell transfection was performed using Lipofectamine 2000 (Invitrogen). Western blotting assessment was used to validate the transfection. Protein extracts were separated using gradient sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and then transferred to polyvinylidene fluoride (PVDF) membranes (Cytiva, catalog number: 10600021) through electroblotting and incubated with primary antibodies (Proteintech Group, catalog number: 81290-1-RR & 66009-1-Ig) at 4°C overnight. Then a 1 h incubation with the appropriate HRP-conjugated secondary antibody at room temperature was performed.

Colony formation assay

T24 or BIU cells were seeded into a 6-well plate (approximately 1.0×10^3 cells/well) and cultured with the corresponding medium for 10 days. Then the cells were washed with phosphate-buffered saline and then treated with 4% paraformaldehyde solution and fixed for 20 min. The cells were then stained with 1% crystal violet solution for 30 min. Finally, the staining solution was slowly washed away with running water.

Results

Proteome-wide MR analysis identified five circulating proteins for BCa

All genetic instruments had F-statistics above 10, indicating sufficient statistical power (Supplementary Table 2). The Wald ratio and IVW method were employed to identify proteins significantly associated with BCa risk. After Bonferroni correction ($P < 1.90 \times 10^{-5}$ (0.05/2,631)), a total of five proteins were identified (**Figure 2**). We found that genetically predicted higher levels of PSCA (OR = 1.16 (1.09-1.23), $P = 1.30 \times 10^{-6}$), LY6D (OR = 1.35 (1.18-1.54), $P = 1.26 \times 10^{-5}$) and SLURP1 (OR = 1.86 (1.24-1.80), $P = 5.03 \times 10^{-5}$) were associated with an increased risk of BCa, while lower levels of the GSTM1 (OR = 0.80 (0.72-0.88), $P = 3.27 \times 10^{-6}$) and GSTM4 (OR = 0.81 (0.74-0.89), $P = 5.14 \times 10^{-6}$) were associated with a higher risk of BCa. Other analytical models also support the directionality of our findings. All results of the discovery of proteome-wide MR are shown in Supplementary Table 4. Based on the Wald ratio or IVW method, all of the circulating proteins were successfully validated in the replication stage of MRC-IEU and Neale Lab and were all orientated in the same direction as the discovery.

Sensitivity analysis for BCa causal proteins

Sensitivity analysis showed that there was no heterogeneity and pleiotropy in five circulating protein results. MR PRESSO analysis also denied the existence of pleiotropy in five proteins. Besides, bidirectional MR analysis did not discover any causal effect of BCa on the level of five identified significant proteins (Supplementary Table 5). All steiger filtering test further ensured the directionality. Finally, Bayesian colocalization strongly suggested that GSTM4 (PPH3 = 0.023, PPH4 = 0.893) shared the same variant with BCa (Supplementary Figure 1). However, we cannot find strong evidence of Bayesian colocalization in the other four proteins (PSCA (PPH3 = 0.308, PPH4 = 0.692); LY6D (PPH3 = 0.999, PPH4 = 0.001); SLURP1 (PPH3 = 0.985, PPH4 = 0.015); GSTM1 (PPH3 = 0.202, PPH4 = 0.183)) (Supplementary Table 6; Supplementary Figures 2, 3, 4 and 5). All results of sensitivity analysis are briefly shown in **Table 1**.

Cell-type specificity expression in the bladder tumor tissue

The sequencing data from five tumor tissues were classified into six cell types, including Epithelial cell, T cell, Fibroblast cell, Endothelial cell, Monocyte cell, B cell, respectively (**Figure 3A**). GSTM1, GSTM4, PSCA, and LY6D could be detected in bladder tumor tissues (**Figure 3C**).

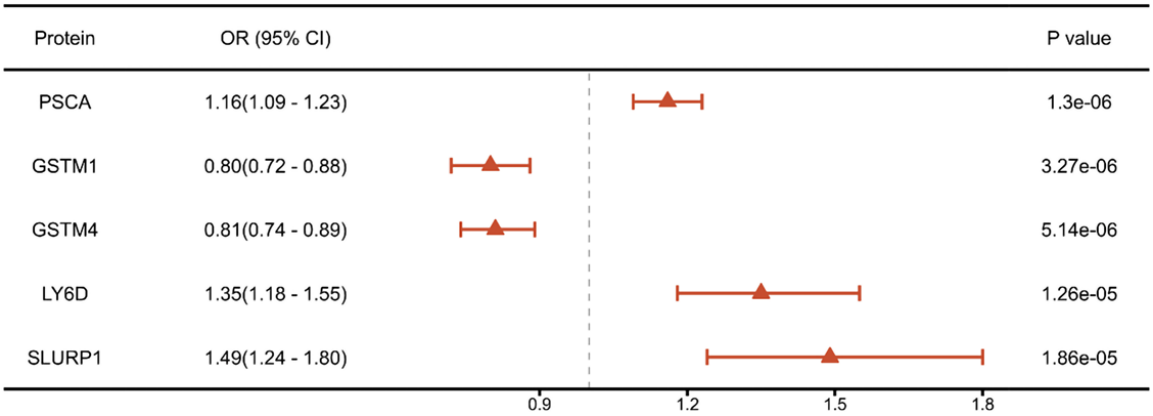


Figure 2. Results of five significant proteins in PWAS. OR: Odds ratio; 95% CI: 95% confidence interval.

Table 1. Sensitivity analysis of significant plasma proteins

Protein	$P_{heterogeneity}$	$P_{pleiotropy}$	$P_{MR\ PRESSO}$	$P_{Reverse\ MR}$	$P_{Steiger\ filtering}$	PPH4
GSTM1	0.489	0.604	0.773	0.824	1.79E-83	0.183
GSTM4	0.641	0.553	0.632	0.724	6.65E-11	0.893
PSCA	0.735	0.784	0.798	0.546	4.85E-03	0.692
LY6D	0.815	0.487	0.837	0.320	3.72E-02	9.07E-04
SLURP1	0.286	0.889		0.385	4.74E-03	0.015

MR: Mendelian randomization; PRESSO: Pleiotropy residual sum and outlier; PPH4: Posterior probability of hypothesis 4.

Among them, GSTM4, PSCA and LY6D were able to show cell type-specific enrichment at $Log_2FC > 0.25$, FDR < 0.05 (**Figure 3B**). GSTM4, PSCA and LY6D were mainly enriched specifically in epithelial cells. Besides, GSTM1, PSCA and LY6D were also able to show cell type-specific enrichment at $Log_2FC > 0.25$, FDR < 0.05 in epithelial cells (**Figure 3D**). In tumor adjacent tissues we also classified six cell types as the tumor tissue. However, GSTM4 cannot be detected at $Log_2FC > 0.25$, FDR < 0.05 (Supplementary Figure 6). We also found GSTM1 can mainly enrich specifically in epithelial cells in normal tissues.

PPI and druggability evaluation on the potentials of therapeutic targets

PPI analysis of five significant protein-coding genes identified four pairs of associations. Among them, GSTM1 and GSTM4 were known interactions from curated databases; The interactions of PSCA, LY6D and SLURP1 were from textmining and co-expression (Supplementary Figure 7). We further identified 3 available drug targets including PSCA, GSTM1 and GSTM4 in druggability assessment (Supplementary Table 7). The drug cisplatin can both target GSTM1 and GSTM4 and be used in the treatment of advanced BCa. We also found numerous drugs targeting GSTM1 were related to inflammation, autoimmune diseases, and cancers. Besides, several unapproved anti-tumor drugs can target GSTM4 and PSCA. Further development of these targets and drugs may provide new solutions for the treatment of BCa.

Four circulating proteins have cross-cancer effect

We further performed a two-sample MR analysis of these five significant circulating proteins in 16 types of cancers. As shown in **Figure 4**, PSCA, LY6D, SLURP1 and GSTM4 were observed to have cross-cancer effect. Same with the trend we found in BCa, PSCA, LY6D, SLURP1 frequently acted as cancer promoting factors, while GSTM1, GSTM4 frequently acted as cancer inhibiting factors. Among them, PSCA showed the most potent cross-cancer effect. Overexpression of PSCA significantly increases the risk of gastric cancer (OR = 1.29 (1.19-1.39), $P = 2.23 \times 10^{-10}$). Significance remains even after Bonferroni correction. High level of

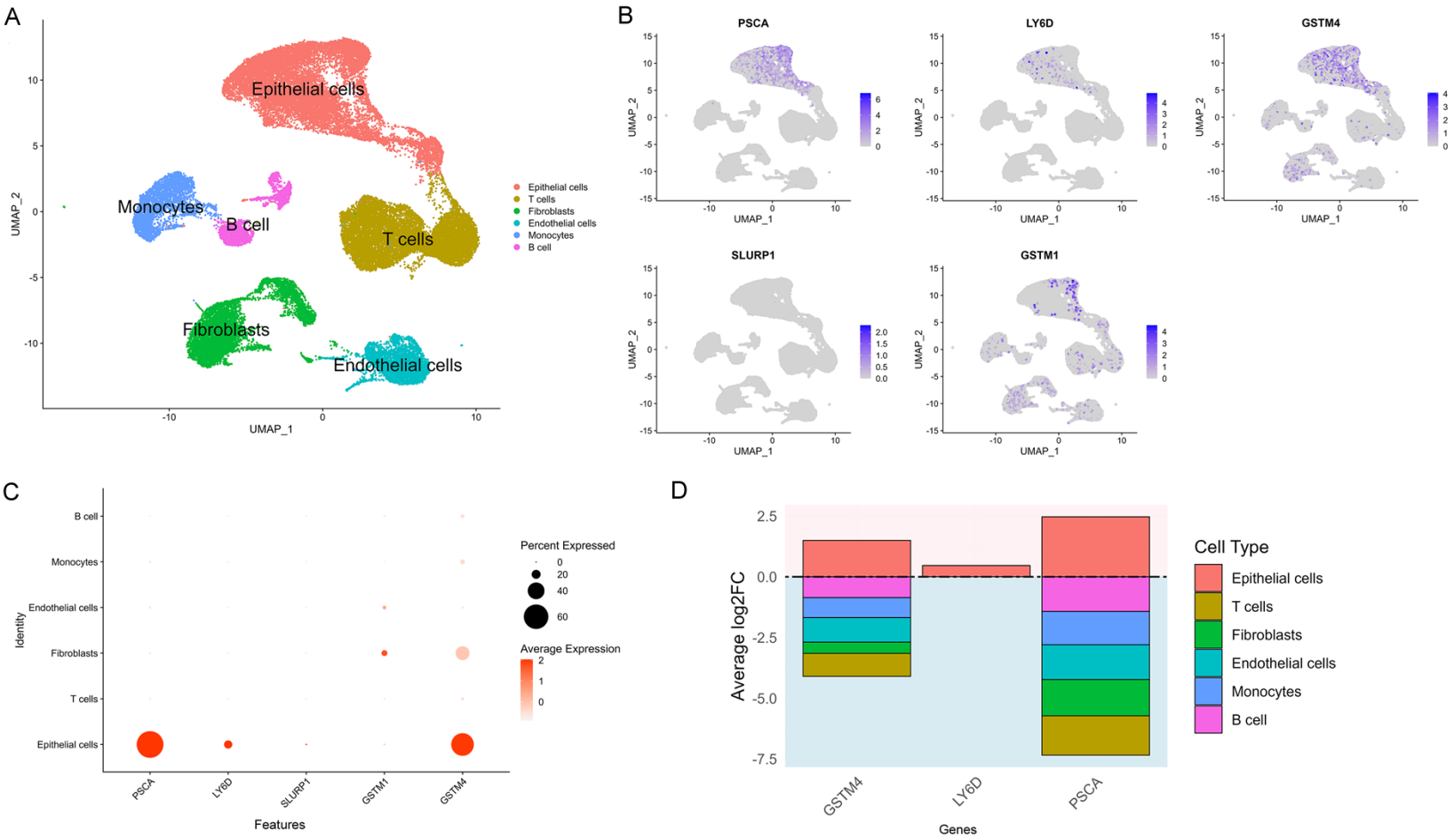


Figure 3. Single-cell type expression in bladder tumor tissues for the five significant targets identified by PWAS. A: A total of six cell types were identified. B, C: The expression of protein-coding genes in each cell type. D: GSTM4, LY6D and PSCA protein coding genes had evidence of enrichment in a cell type at average $Log_2FC > 0.25$ and FDR < 0.05.

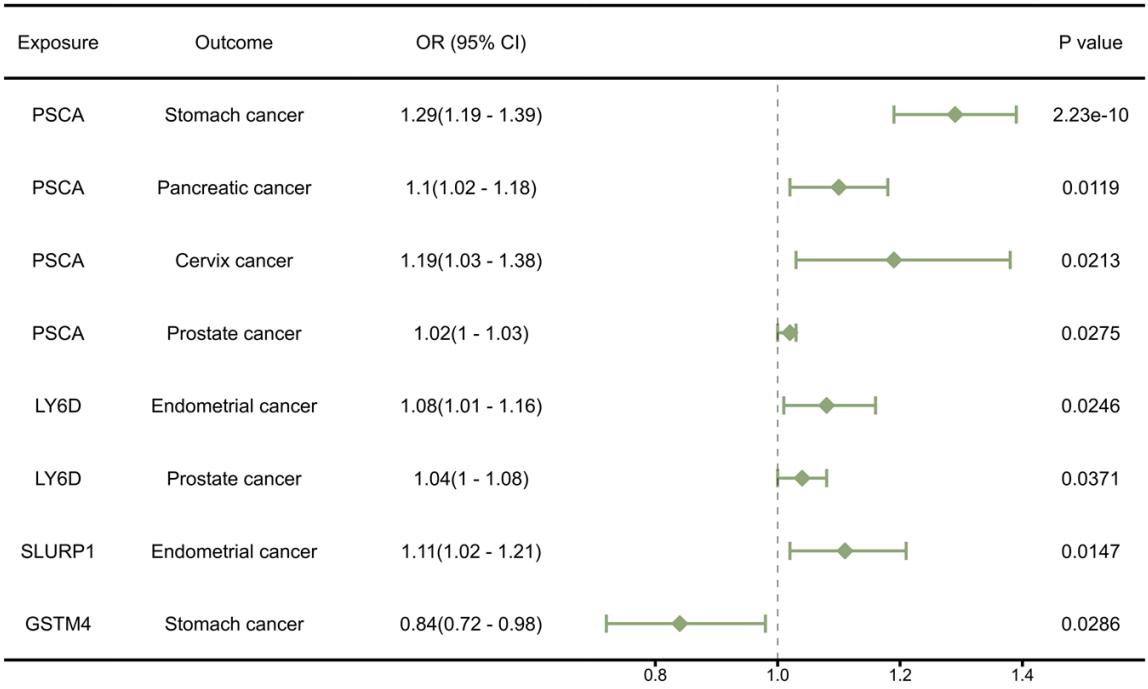


Figure 4. Results of cross-cancer effect of these circulating targets.

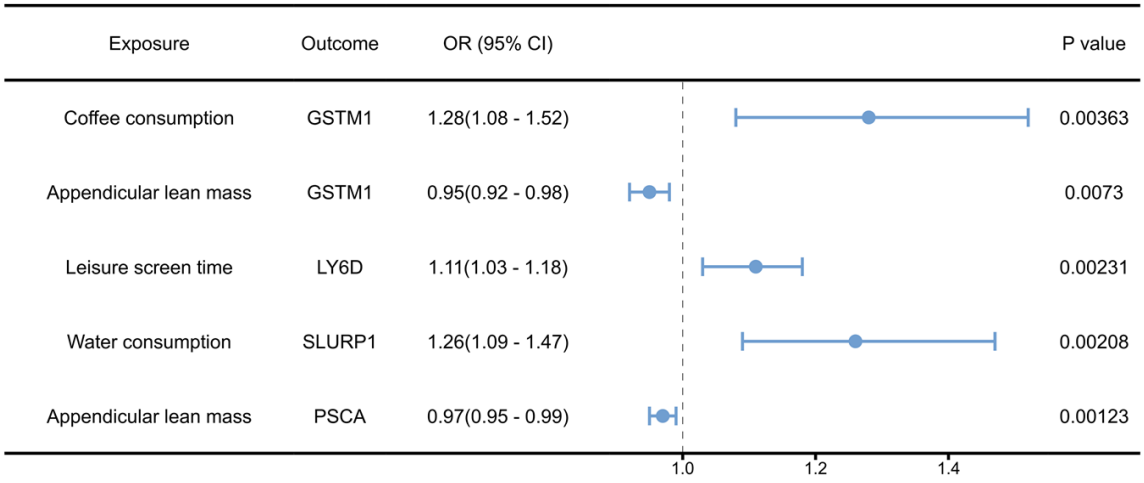


Figure 5. Results of lifestyles that can affect selected circulating protein targets.

PSCA can also increase the risk of pancreatic cancer (OR = 1.09 (1.02-1.18), *P* = 0.012), cervix cancer (OR = 1.19 (1.02-1.37), *P* = 0.021), prostate cancer (OR = 1.02 (1.00-1.03), *P* = 0.027). Besides, high level of LY6D can increase the risk of endometrial cancer (OR = 1.08 (1.01-1.16), *P* = 0.025) and prostate cancer (OR = 1.04 (1.00-1.08), *P* = 0.037). High level of SLURP1 can increase the risk of endometrial cancer (OR = 1.11 (1.02-1.21), *P* = 0.015). Finally, we found high level of GSTM4 may decrease the risk of stomach cancer (OR = 0.84 (0.72-0.98), *P* = 0.029).

Four lifestyles were related to circulating proteins

After analyzing the effects of 15 lifestyles on five circulating proteins, we identified four lifestyles that had a significant effect on these cir-

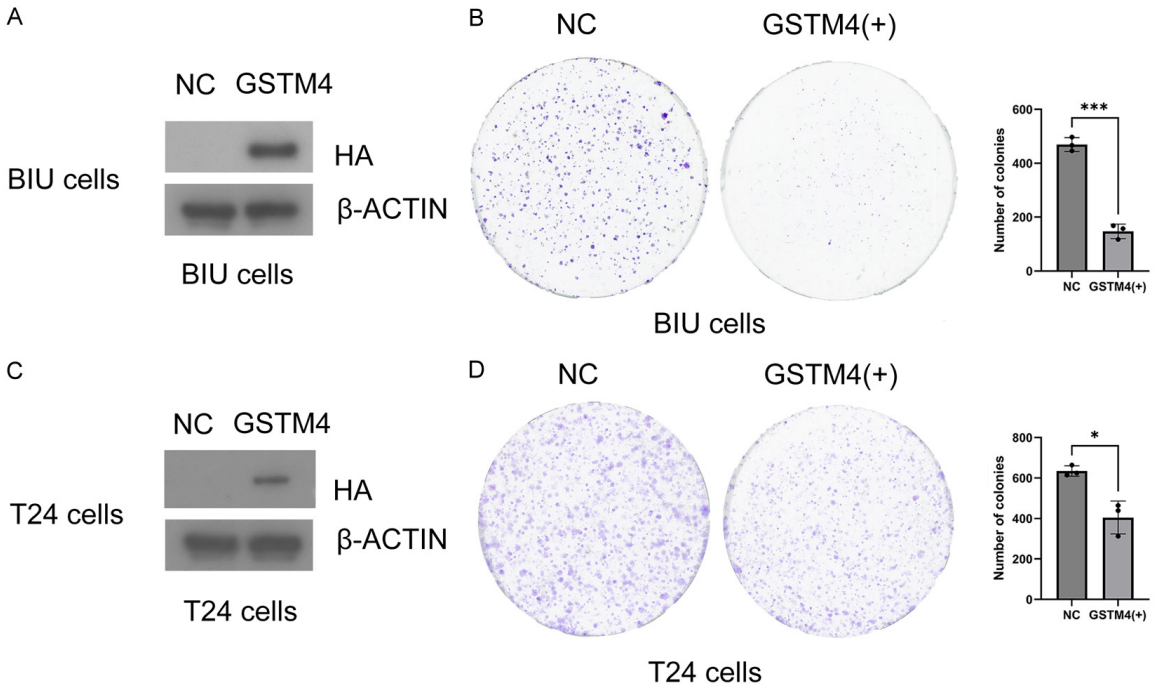


Figure 6. Colony formation assay showed that GSTM4 inhibited proliferation of bladder cancer cells. Colony numbers was counted and plotted. A, C: GSTM4 transfection in bladder cancer cells was detected using western blotting. B, D: The effects of overexpression of GSTM4 on proliferation were measured by colony formation assay in T24 and BIU cells. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

culating proteins. As shown in **Figure 5**, we found that excessive leisure screen time increased circulating levels of LY6D (OR = 1.11 (1.04-1.19), *P* = 0.00231), which can increase the risk of BCa. Moreover, increased coffee consumption enhanced circulating levels of GSTM1 (OR = 1.29 (1.09-1.53), *P* = 0.00363), which can decrease the risk of BCa. We also found that water consumption can act as a cancer promoter by increasing the level of SLURP1 (OR = 1.26 (1.09-1.47), *P* = 0.00208). However, the elevation of appendicular lean mass was able to reduce both PSCA (OR = 0.97 (0.95-0.99), *P* = 0.00123) and GSTM1 (OR = 0.95 (0.92-0.99), *P* = 0.00730), thus demonstrating an opposite effect on cancer risk.

Phenome-wide MR analysis of five druggable proteins

Most drugs work through the bloodstream. We assessed whether these five BCa-associated plasma proteins are beneficial or harmful to other traits. Therefore, we performed a phenome-wide MR screening of 783 diseases or traits in the UK Biobank. The phenome-wide MR results are shown in Supplementary Tables 9-13. Overall, we found a low level of PSCA (OR = 0.90 (0.87-0.94), FDR = 1.49 × 10^{-5}) and LY6D (OR = 0.95 (0.92-0.99), FDR = 0.037) may lead to duodenitis and similar phenomena was also found in SLURP1 (OR = 0.84 (0.76-0.93), *P* = 8.03 × 10^{-4}). Lower PSCA was significantly associated with peptic ulcer (excl. esophageal) (OR = 0.90 (0.88-0.94), FDR = 2.93 × 10^{-5}), especially duodenal ulcer (OR = 0.79 (0.84-0.89), FDR = 6.74 × 10^{-7}). Besides, lower PSCA was associated with other local infections of skin and subcutaneous tissue (OR = 0.93 (0.89-0.97), FDR = 0.04). Finally, the five causal proteins we discovered once again showed a strong correlation with BCa in phenome-wide MR.

Plate colony formation assay verified the inhibitory action of GSTM4 in BCa cells

We have also carried out in vitro experiments to test the reliability of our results. As shown in **Figure 6A**, **6C**, western blotting validated the successful transfection of GSTM4 in T24 and BIU cells. After treatment with the same conditions, the number of BCa cells in the overexpressing GSTM4 group was generally lower

than that in the control group (**Figure 6B**, **6D**). These results illustrated GSTM4 could suppress the proliferation of BCa cells.

Discussion

In this study, we conducted a comprehensive investigation of the causal relationship between 2,631 plasma proteins and BCa risk. Whole proteome MR identified five protein targets where genetically determined higher levels of PSCA, LY6D and SLURP1 could increase the risk of BCa, and genetically determined higher levels of GSTM1 and GSTM4 were associated with reduced risk of BCa. Validation of five plasma proteins in two replication datasets (MRC-IEU & UKBioBank) all passed. In addition, these five proteins passed all other sensitivity analyses including the heterogeneity test, horizontal pleiotropy test, MR PRESSO test, reverse two-sample MR test, and steiger test. However, Bayesian colocalization only identified causal effect for GSTM4. Overall, we identified GSTM4 as the protein with the most convincing evidence, while PSCA, LY6D, SLURP1 and GSTM1 had slightly weaker evidence for causal association with BCa. Among them, GSTM4 and SLURP1 are novel plasma protein markers associated with BCa. Several follow-up analyses were performed for these five causal proteins. We verified the expression of these protein-coding genes in different cell types of BCa tissues and adjacent tissues. Meanwhile, PPI analysis identified protein-protein interactions between PSCA, LY6D, and SLURP1, and protein-protein interactions between GSTM1 and GSTM4. The drugability assessment prioritized GSTM1 and GSTM4. Drugs targeting the other three proteins were not approved or explored. Besides, we found GSTM4, PSCA, LY6D, and SLURP1 had cross-cancer effects. Four types of daily lifestyles including coffee consumption, water consumption, leisure screen time and appendicular lean mass can affect the causal proteins related to BCa. In addition, phenome-wide MR showed potential side effects of PSCA and LY6D, which can serve as a reference for the pharmaceutical safety. Importantly, we also found that overexpression of GSTM4 can inhibit the proliferation of BCa cells through colony formation assay.

The five proteins we identified can be categorized into two groups based on PPI analysis. The first group is PSCA, LY6D, and SLURP1, whose high expression increases the risk of BCa. PSCA, LY6D and SLURP1 are all belong to lymphocyte antigen-6 (LY6) gene family and located on human chromosome 8q24.3 [21]. The second group is GSTM1 and GSTM4, whose high expression reduces the risk of BCa. GSTM1 and GSTM4 are both belong to glutathione S-transferase (GST) gene family μ class and located on human chromosome 1p13.3 [22]. The proteins within these two groups have a similar locus, so it is valuable to explore the effect of these regions on BCa.

PSCA encodes a glycosylphosphatidylinositol-anchored cell membrane glycoprotein and lowly expressed in bladder and stomach tissue [23]. Several genome-wide association studies indicated that SNPs on 8q24.3 in the PSCA gene can act as a risk factor for BCa [23, 24]. However, antineoplastic drugs targeting PSCA mostly in not approved and investigated states. Interestingly, as past studies have found and research has shown in our study, PSCA showed strong cross-cancer effect such as stomach cancer [25-27]. This implies that PSCA may play an important role in tumor formation and that it may be more beneficial to develop drugs that target PSCA. LY6D is a transmembranous protein widely spread across diverse cell types. Previous study have shown that LY6D can be a biomarker for urothelial cancer [21]. The expression of LY6D was also found to be higher in tumors with squamous cell differentiation than in tumors with classic urothelial histology [28]. In patients who had not received neoadjuvant chemotherapy, LY6D was lower expressed in resected tumors compared to transurethral resection of the bladder specimens [28]. SLURP1 can regulate the epithelial cell growth, program cell death, and tumor development, involve epithelial-mesenchymal transition, and serve as a biomarker for late epithelial differentiation [21, 29, 30]. However, the association between SLURP1 and BCa has never been studied. To confirm our findings, further epidemiologic studies and experimental research are needed.

Our research indicated that circulating GSTM4 can act as an inhibitor in BCa. Similar to the trend of BCa, the protein levels of GSTM4 were reduced in breast cancer compared to normal tissue [31, 32]. In breast cancer and ovarian

cancer, researchers also indicated that low expression of GSTM4 was associated with a worse prognosis. However, direct evidence for GSTM4 protein and BCa risk has not been reported. In our research, GSTM4 can inhibit the proliferation of T24 and BIU cells. Further exploration of the molecular mechanism of GSTM4 in BCa will help us to better develop related drugs. Circulating GSTM1 also act as an inhibitor in our study. The association between GSTM1 polymorphisms and the risk of cancer has been studied extensively in a variety of cancer types, including BCa [33-36]. GSTM1 tends to exhibit tumor suppression in the previous studies. For example, deletion of the GSTM1 can increase the risk of cancer, including BCa, gastric cancer and liver cancer [37, 38]. Besides, according to our druggability evaluation, drugs targeting GSTM1 have been widely used in the treatment of cancer in the clinic. Our study also showed that drug development for BCa targeting GSTM1 is valuable.

This study has strong advantages. With the development of GWAS and MR analysis, we comprehensively examined the causal associations between plasma protein biomarkers and BCa risk by performing proteome-wide MR. Through replications and multiple sensitivity analyses, we have largely ensured the reliability of the results. Meanwhile, evidence from single-cell type expression analyses, PPI, druggability assessments, cross-cancer analyses, and phenome-wide MR provided additional insights into the potential pathogenicity and druggability of candidate proteins for BCa. Assessing potential lifestyle effects on these target proteins can provide new insights into cancer prevention.

Our study has several limitations. Firstly, we only used cis-pQTL as an IV to study protein drug targets related to BCa. Using trans-pQTL can expand the scope of our search for protein targets. However, due to the lack of research between trans-pQTL and proteins, it is often difficult to interpret the causal relationship between trans-pQTL and BCa [39]. Secondly, although we divided significant circulating proteins into two groups based on gene locus and protein-protein interactions, some PPI results were not supported by basic experiments. Thirdly, Bonferroni correction allows us to maximize the reliability but may lead to underestimating the persuasive power of the omitted results. Finally, our study was performed only on data from European ancestry, further research is needed on non-European ancestry to confirm our findings.

Conclusion

Our study identified 5 plasma proteins associated with BCa risk and provided promising targets for developing biomarkers and therapeutic agents for BCa. Further experimental and clinical studies are needed to evaluate the role of these targets.

Acknowledgements

We want to acknowledge the participants and investigators of two proteomic studies UKBPPP and DECODE, three bladder cancer GWASs and all of other data involved in this article. This study was supported by Jiang Xi Natural Science Foundation (20202BABL206023), Jiangxi Provincial "Double Thousand Plan" Fund Project (Grant No. jxsq2019201027), the Key Project of the Natural Science Foundation of Jiangxi Province (20212ACB206013), and the Youth Project of Natural Science Foundation of Jiangxi Province (20212BAB216037).

Disclosure of conflict of interest

None.

Abbreviations

BCa, Bladder cancer; GEO, Gene Expression Omnibus; GWASs, Genome-wide association studies; IV, Instrumental variable; IVW, Inverse-variance weighted; Log_2FC, Log_2 fold change; MR, Mendelian randomization; PPH4, Posterior probability of hypothesis 4; PPI, Protein-protein interaction; Pqtl, Protein quantitative trait loci; PWAS, Proteome-wide association study; MR PRESSO, MR pleiotropy residual sum and outlier.

Address correspondence to: Bin Fu and Xiao-Rong Yang, The First Affiliated Hospital of Nanchang University, No. 17 Yongwai Zheng Street, Donghu District, Nanchang, Jiangxi, China. Tel: +86-138-79103861; E-mail: urodoc@ncu.edu.cn (BF); Tel: +86-1357614128; E-mail: ndyfy02349@ncu.edu.cn (XRY)

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[2] Joshi SK, Piehowski P, Liu T, Gosline SJC, McDermott JE, Druker BJ, Traer E, Tyner JW, Agarwal A, Tognon CE and Rodland KD. Mass spectrometry-based proteogenomics: new therapeutic opportunities for precision medicine. Annu Rev Pharmacol Toxicol 2024; 64: 455-479.

[3] Kaur G, Poljak A, Ali SA, Zhong L, Raftery MJ and Sachdev P. Extending the depth of human plasma proteome coverage using simple fractionation techniques. J Proteome Res 2021; 20: 1261-1279.

[4] Zhang X, Yu W, Li Y, Wang A, Cao H and Fu Y. Drug development advances in human genetics-based targets. MedComm (2020) 2024; 5: e481.

[5] Reay WR and Cairns MJ. Advancing the use of genome-wide association studies for drug repurposing. Nat Rev Genet 2021; 22: 658-671.

[6] Burgess S, Butterworth A and Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013; 37: 658-665.

[7] Tin A and Köttgen A. Mendelian randomization analysis as a tool to gain insights into causes of diseases: a primer. J Am Soc Nephrol 2021; 32: 2400-2407.

[8] Yuan S, Mason AM, Carter P, Vithayathil M, Kar S, Burgess S and Larsson SC. Selenium and cancer risk: wide-angled Mendelian randomization analysis. Int J Cancer 2022; 150: 1134-1140.

[9] Jiang L, Zheng Z, Fang H and Yang J. A generalized linear mixed model association tool for biobank-scale data. Nat Genet 2021; 53: 1616-1621.

[10] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR and Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. Elife 2018; 7: e34408.

[11] Lyon MS, Andrews SJ, Elsworth B, Gaunt TR, Hemani G and Marcora E. The variant call format provides efficient and robust storage of GWAS summary statistics. Genome Biol 2021; 22: 32.

[12] Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, Surendran P, Mahajan A, Robins C, Vasquez-Grinnell SG, Hou L, Kvikstad EM, Burren OS, Davitte J, Ferber KL, Gillies CE, Hedman ÅK, Hu S, Lin T, Mikkilineni R, Pendergrass RK, Pickering C, Prins B, Baird D, Chen CY, Ward LD, Deaton AM, Welsh S, Willis CM, Lehner N, Arnold M, Wörheide MA, Suhre K, Kastenmüller G, Sethi A, Cule M, Raj A; Alnylam Human Genetics; AstraZeneca Genomics Initiative; Biogen Biobank Team; Bristol Myers Squibb; Genentech Human Genetics; GlaxoSmithKline Genomic Sciences; Pfizer Integrative Biology; Population Analytics of Janssen Data Sciences; Regeneron Genetics Center, Burkitt-Gray L, Melamud E, Black MH, Fauman EB, Howson JMM, Kang HM, McCarthy MI, Nioi P, Petrovski S, Scott RA, Smith EN, Szalma S, Waterworth DM, Mitnaul LJ, Szustakowski JD, Gibson BW, Miller MR and Whelan CD. Plasma proteomic associations with genetics and health in the UK Biobank. Nature 2023; 622: 329-338.

[13] Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, Gunnarsdottir K, Helgason A, Oddsson A, Halldorsson BV, Jensson BO, Zink F, Halldorsson GH, Masson G, Arnadottir GA, Katrinardottir H, Juliusson K, Magnusson MK, Magnusson OT, Fridriksdottir R, Saevarsdottir S, Gudjonsson SA, Stacey SN, Rognvaldsson S, Eiriksdottir T, Olafsdottir TA, Steinthorsdottir V, Tragante V, Ulfarsson MO, Stefansson H, Jonsdottir I, Holm H, Rafnar T, Melsted P, Saemundsdottir J, Norddahl GL, Lund SH, Gudbjartsson DF, Thorsteinsdottir U and Stefansson K. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet 2021; 53: 1712-1721.

[14] Rashkin SR, Graff RE, Kachuri L, Thai KK, Alexeeff SE, Blatchins MA, Cavazos TB, Corley DA, Emami NC, Hoffman JD, Jorgenson E, Kushi LH, Meyers TJ, Van Den Eeden SK, Ziv E, Habel LA, Hoffmann TJ, Sakoda LC and Witte JS. Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. Nat Commun 2020; 11: 4423.

[15] Bowden J, Davey Smith G, Haycock PC and Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol 2016; 40: 304-314.

[16] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C and Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014; 10: e1004383.

[17] Su WM, Gu XJ, Dou M, Duan QQ, Jiang Z, Yin KF, Cai WC, Cao B, Wang Y and Chen YP. Systematic druggable genome-wide Mendelian randomisation identifies therapeutic targets for Alzheimer's disease. J Neurol Neurosurg Psychiatry 2023; 94: 954-961.

[18] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C and Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2018; 46: D1074-D1082.

[19] Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, Griffith M, Griffith OL and Wagner AH. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res 2021; 49: D1144-D1151.

[20] Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, LeFaive J, VandeHaar P, Gagliano SA, Gifford A, Bastarache LA, Wei WQ, Denny JC, Lin M, Hveem K, Kang HM, Abecasis GR, Willer CJ and Lee S. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 2018; 50: 1335-1341.

[21] Upadhyay G. Emerging role of lymphocyte antigen-6 family of genes in cancer and immune cells. Front Immunol 2019; 10: 819.

[22] Nebert DW and Vasiliou V. Analysis of the glutathione S-transferase (GST) gene family. Hum Genomics 2004; 1: 460-464.

[23] Kadhum T, Selinski S, Blaszkewicz M, Reinders J, Roth E, Volkert F, Ovsiannikov D, Moormann O, Gerullis H, Barski D, Otto T, Höhne S, Hengstler JG and Golka K. Bladder cancer course, four genetic high-risk variants, and histopathological findings. EXCLI J 2023; 22: 867-879.

[24] Gakis G, Perner S, Stenzl A and Renninger M. The role of single-nucleotide polymorphisms of the 8q24 chromosome region in patients with concomitant bladder and prostate cancer. Scand J Urol 2022; 56: 126-130.

[25] Pyun H, Gunathilake M, Lee J, Choi IJ, Kim YI, Sung J and Kim J. Functional annotation and gene set analysis of gastric cancer risk loci in a Korean population. Cancer Res Treat 2024; 56: 191-198.

[26] Shah Z, Tian L, Li Z, Jin L, Zhang J, Li Z, Barr T, Tang H, Feng M, Caligiuri MA and Yu J. Human anti-PSCA CAR macrophages possess potent antitumor activity against pancreatic cancer. Cell Stem Cell 2024; 31: 803-817.e6.

[27] Nayerpour Dizaj T, Doustmihan A, Sadeghzadeh Oskouei B, Akbari M, Jaymand M, Mazloomi M and Jahanban-Esfahlan R. Significance of PSCA as a novel prognostic marker and therapeutic target for cancer. Cancer Cell Int 2024; 24: 135.

[28] Andersson N, Ohlsson J, Wahlin S, Nodin B, Boman K, Lundgren S and Jirström K. Lymphocyte antigen 6 superfamily member D is a marker of urothelial and squamous differentiation: implications for risk stratification of bladder cancer. Biomark Res 2020; 8: 51.

[29] Kalantari-Dehaghi M, Bernard HU and Grando SA. Reciprocal effects of NNK and SLURP-1 on oncogene expression in target epithelial cells. Life Sci 2012; 91: 1122-1125.

[30] Bychkov ML, Shulepko MA, Shlepova OV, Kulbatskii DS, Chulina IA, Paramonov AS, Baidakova LK, Azev VN, Koshelev SG, Kirpichnikov MP, Shenkarev ZO and Lyukmanova EN. SLURP-1 controls growth and migration of lung adenocarcinoma cells, forming a complex with α7-nAChR and PDGFR/EGFR heterodimer. Front Cell Dev Biol 2021; 9: 739391.

[31] Gohari N, Abbasi E and Akrami H. Comprehensive analysis of the prognostic value of glutathione S-transferases Mu family members in breast cancer. Cell Biol Int 2024; 48: 1313-1325.

[32] Zhang J, Li Y, Zou J, Lai CT, Zeng T, Peng J, Zou WD, Cao B, Liu D, Zhu LY, Li H and Li YK. Comprehensive analysis of the glutathione S-transferase Mu (GSTM) gene family in ovarian cancer identifies prognostic and expression significance. Front Oncol 2022; 12: 968547.

[33] Yu Y, Zhang Z, Dong X, Yang R, Duan Z, Xiang Z, Li J, Li G, Yan F, Xue H, Jiao D, Lu J, Lu H, Zhang W, Wei Y, Fan S, Li J, Jia J, Zhang J, Ji J, Liu P, Lu H, Zhao H, Chen S, Wei C, Chen H and Zhu Z. Pangenomic analysis of Chinese gastric cancer. Nat Commun 2022; 13: 5412.

[34] Huang YC, Lai JC, Peng PH, Wei KC and Wu KJ. Chromatin accessibility analysis identifies GSTM1 as a prognostic marker in human glioblastoma patients. Clin Epigenetics 2021; 13: 201.

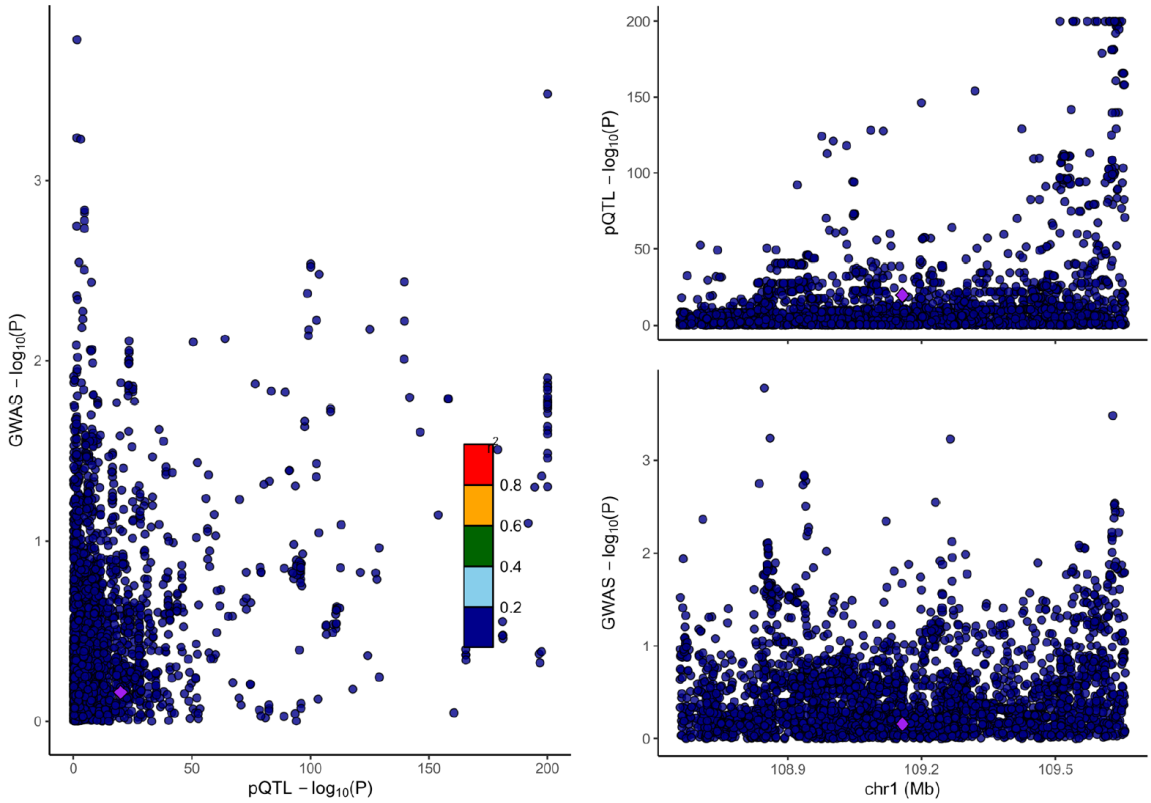
[35] Song L, Yang C and He XF. Individual and combined effects of GSTM1 and GSTT1 polymorphisms on colorectal cancer risk: an updated meta-analysis. Biosci Rep 2020; 40: BSR20201927.

[36] Albarakati N, Khayyat D, Dallol A, Al-Maghrabi J and Nedjadi T. The prognostic impact of GSTM1/GSTP1 genetic variants in bladder Cancer. BMC Cancer 2019; 19: 991.

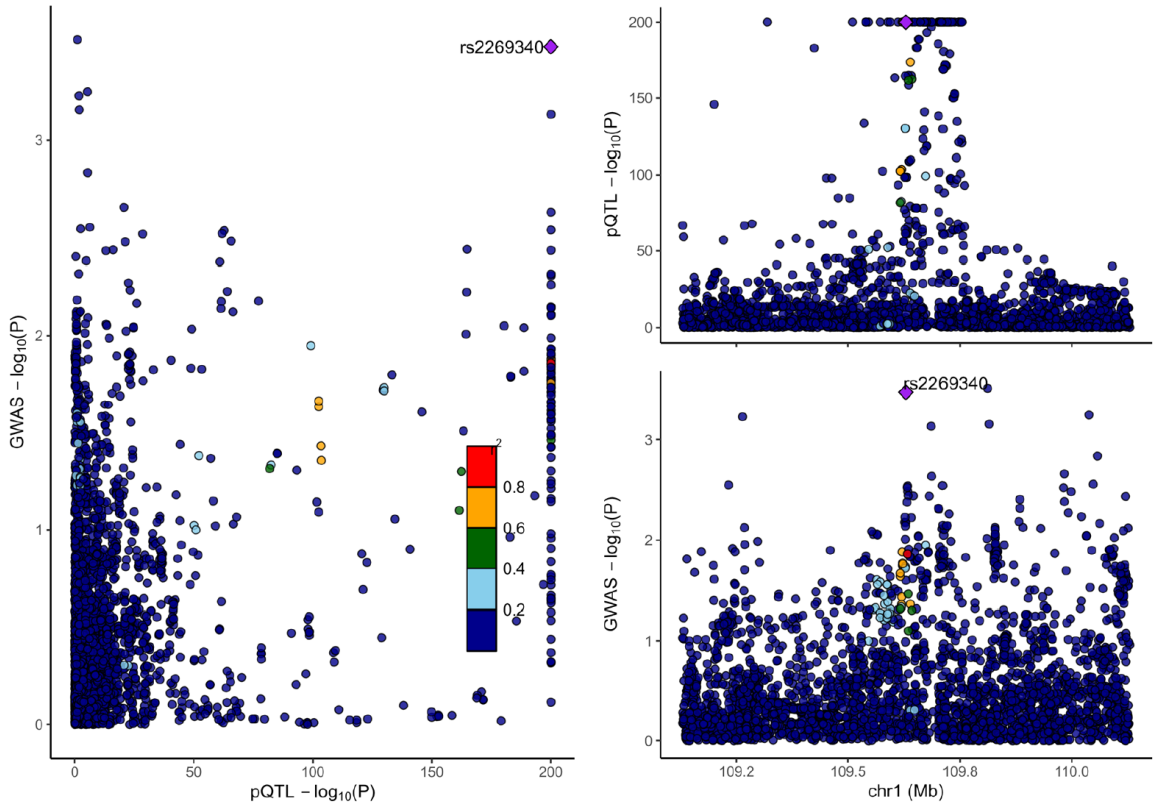
[37] Lu Y, Zhou J, Zhang J, Wang Z, Yu Y, Miao M and Yao Q. Dual roles of glutathione S-transferase mu 1 in the development and metastasis of hepatocellular carcinoma. Biomed Pharmacother 2019; 120: 109532.

[38] Chen Y, Li B, Wang J, Liu J, Wang Z, Mao Y, Liu S, Liao X and Chen J. Identification and verification of the prognostic value of the glutathione S-transferase Mu genes in gastric cancer. Oncol Lett 2020; 20: 100.

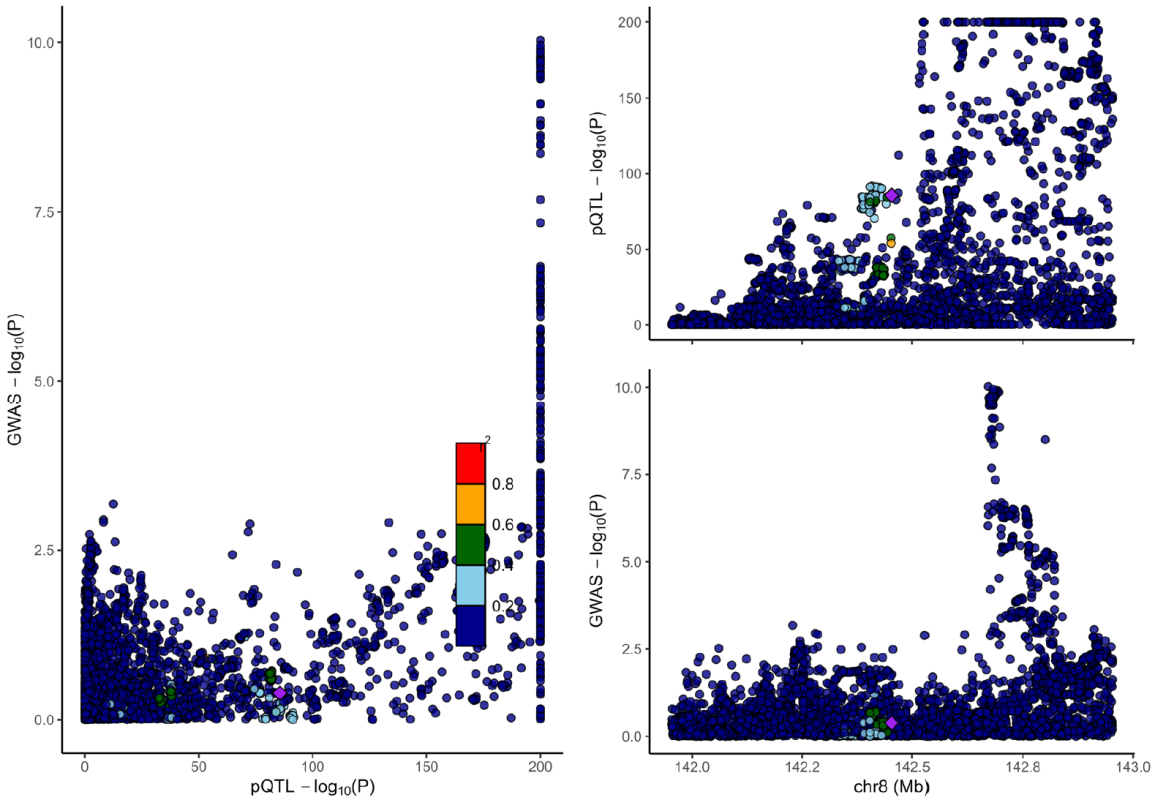
[39] Surapaneni A, Schlosser P, Zhou L, Liu C, Chatterjee N, Arking DE, Dutta D, Coresh J, Rhee EP and Grams ME. Identification of 969 protein quantitative trait loci in an African American population with kidney disease attributed to hypertension. Kidney Int 2022; 102: 1167-1177.



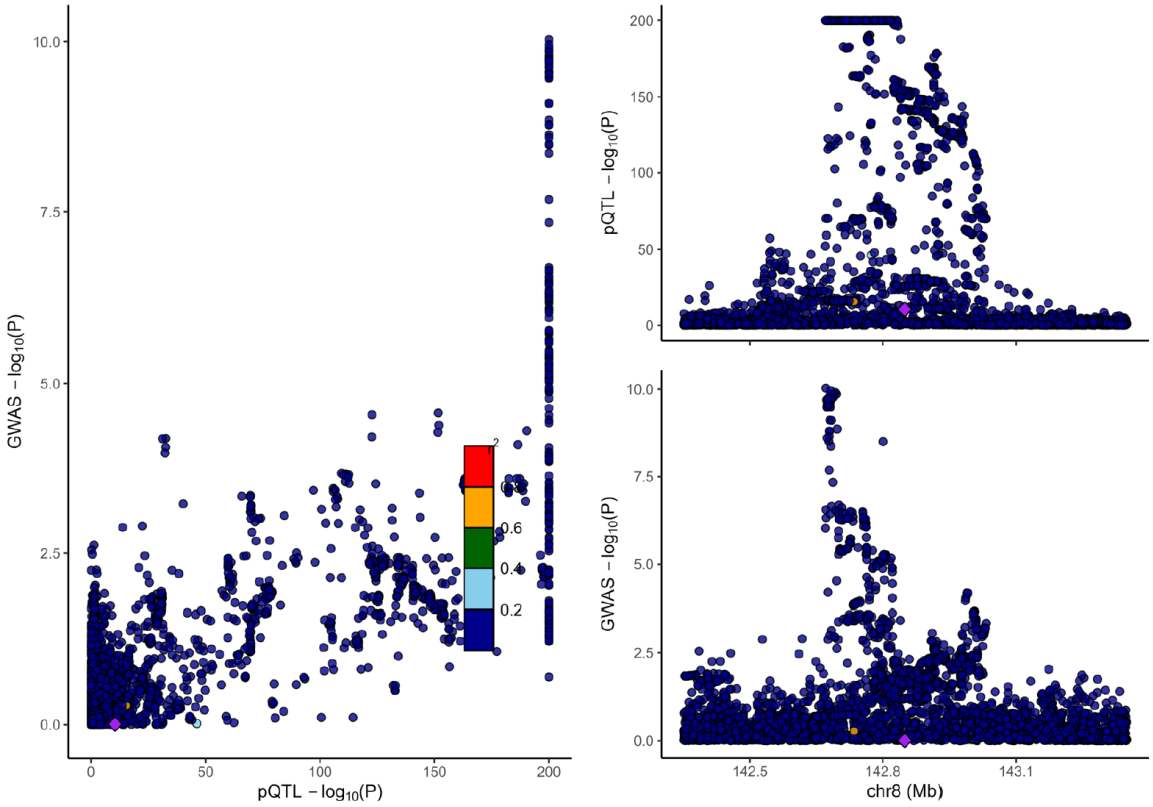
Supplementary Figure 1. Locus comparing plots for colocalization analysis of GSTM4 protein with bladder cancer (BCa) risk. The lead SNP rs17641881 is shown as a purple diamond.



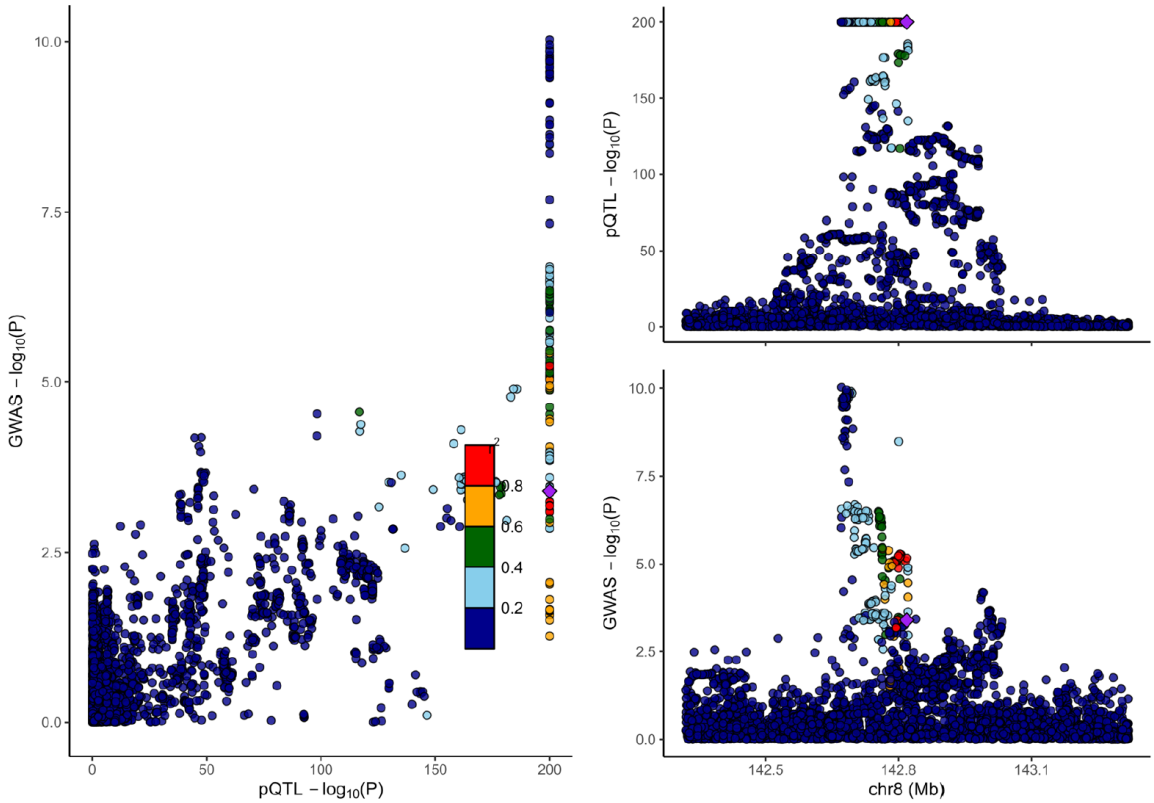
Supplementary Figure 2. Locus comparing plots for colocalization analysis of GSTM1 protein with BCa risk. The lead SNP rs2269340 is shown as a purple diamond.



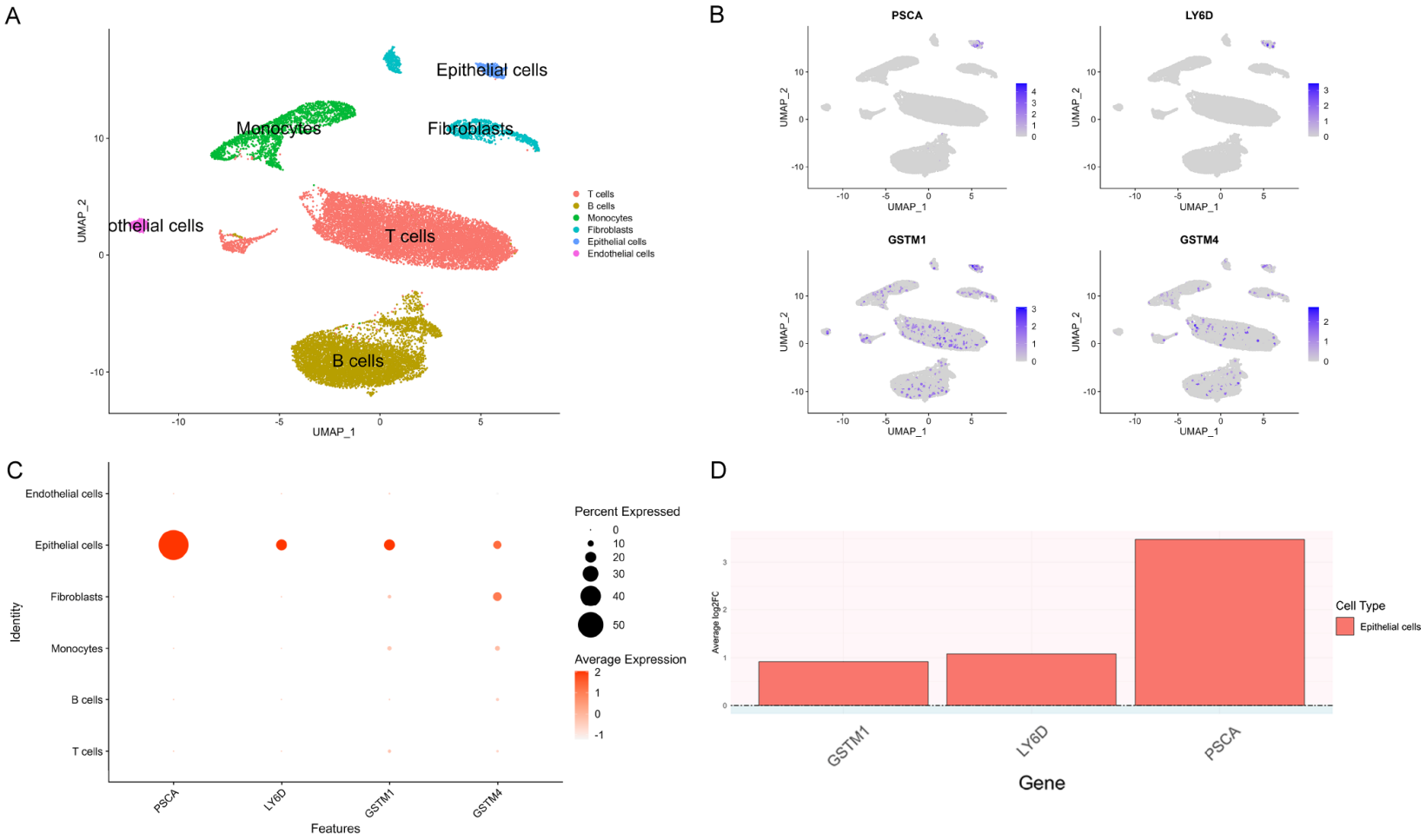
Supplementary Figure 3. Locus comparing plots for colocalization analysis of PSCA protein with BCa risk. The lead SNP rs7460106 is shown as a purple diamond.



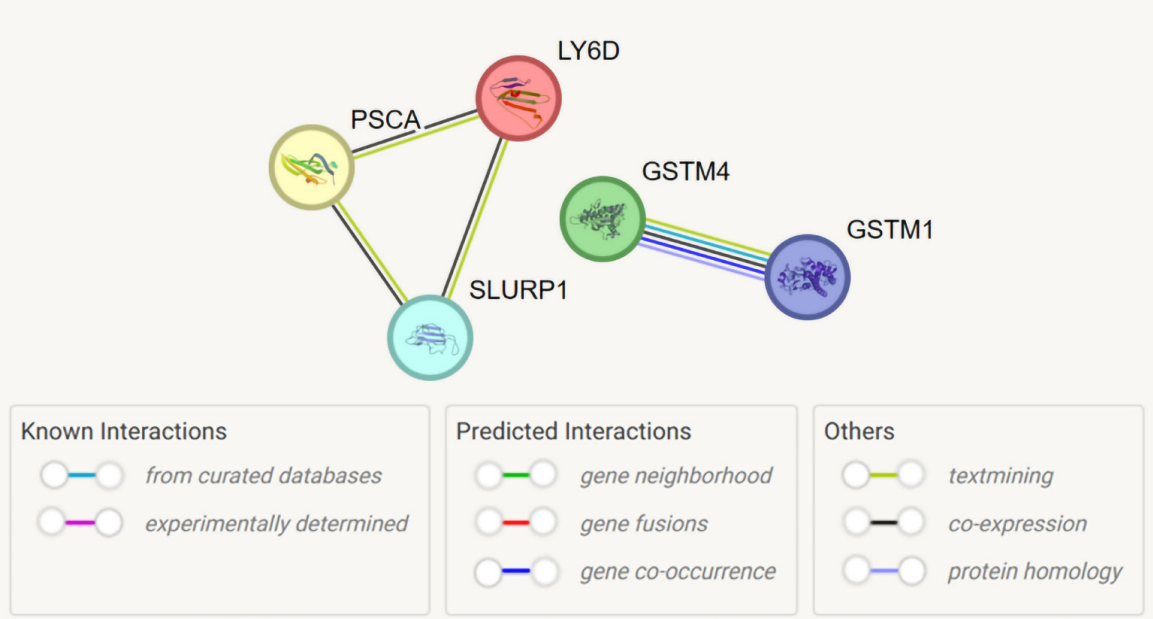
Supplementary Figure 4. Locus comparing plots for colocalization analysis of LY6D protein with BCa risk. The lead SNP rs118142585 is shown as a purple diamond.



Supplementary Figure 5. Locus comparing plots for colocalization analysis of SLURP1 protein with BCa risk. The lead SNP rs2516381 is shown as a purple diamond.



Supplementary Figure 6. Single-cell type expression in bladder tumor adjacent tissues for the significant targets identified by proteome-wide MR. A. A total of six cell types were identified. B, C. The expression of protein-coding genes in each cell type. D. Three protein-coding genes had evidence of enrichment in a cell type at average $Log_2FC > 0.25$ and FDR < 0.05.



Supplementary Figure 7. Protein-protein interaction results.