#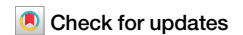 Mendelian randomization reveals potential causal relationships between cellular senescence-related genes and multiple cancer risks

Check for updates

Xunan Qiu[1,2,3,4], Rui Guo[1,2,3,4], Yingying Wang[1,2,3], Shuwen Zheng[1,2,3], Bengang Wang [1,2,3] ✉ & Yuehua Gong [1,2,3] ✉

Cellular senescence is widely acknowledged as having strong associations with cancer. However, the intricate relationships between cellular senescence-related (CSR) genes and cancer risk remain poorly explored, with insights on causality remaining elusive. In this study, Mendelian Randomization (MR) analyses were used to draw causal inferences from 866 CSR genes as exposures and summary statistics for 18 common cancers as outcomes. We focused on genetic variants affecting gene expression, DNA methylation, and protein expression quantitative trait loci (cis-eQTL, cis-mQTL, and cis-pQTL, respectively), which were strongly linked to CSR genes alterations. Variants were selected as instrumental variables (IVs) and analyzed for causality with cancer using both summary-data-based MR (SMR) and two-sample MR (TSMR) approaches. Bayesian colocalization was used to unravel potential regulatory mechanisms underpinning risk variants in cancer, and further validate the robustness of MR results. We identified five CSR genes (*CNOT6, DNMT3B, MAP2K1, TBPL1*, and *SREBF1)*, 18 DNA methylation genes, and LAYN protein expression which were all causally associated with different cancer types. Beyond causality, a comprehensive analysis of gene function, pathways, and druggability values was also conducted. These findings provide a robust foundation for unravelling CSR genes molecular mechanisms and promoting clinical drug development for cancer.

Cellular senescence is a biological response to different external stimuli, encompassing microenvironmental stress, nutrient deprivation, DNA damage, cellular and organelle impairment, and cellular signaling network abnormalities. The process manifests with four characteristics: cell cycle arrest, a senescence-associated secretory phenotype (SASP), macromolecular damage, and metabolic disorder[1,2]. Cellular senescence is a recently hypothesized molecular cancer hallmark and has important implications for age-related illnesses, carcinogenesis, and tissue repair[3,4]. In terms of carcinogenesis, cellular senescence functions as a "double-edged sword"; on the one hand, senescence exerts inhibitory effects toward tumorigenesis through cell-intrinsic mechanisms, fostering local anti-tumor immunity, contributing to wound healing, and enhancing host immunity[5]. Senescent cells, on the other hand, generate and secrete cytokines that stimulate tumor cell growth, ultimately contributing to tumorigenesis by facilitating chronic inflammation and associated degenerative and hyperplastic lesions[6]. Therefore, elucidating cellular senescence roles in tumorigenesis is critical.

Currently, based on traditional observational studies, associations between specific cellular senescence-related (CSR) genes and diverse tumor risks have been identified. For example, *EZH2* and *TOP2A* knockdown inhibit hepatocarcinogenesis by inducing cellular senescence[7]. IL-1 is a key SASP component and promotes the malignant transformation of oral precursor lesions[7]. *SOX6* suppresses cervical cell proliferation by inducing cellular senescence[8]. Additionally, *FBXW7* likely functions as a tumor suppressor in esophageal squamous carcinoma by regulating cellular senescence and DNA damage repair, among other mechanisms[9].

[1]Tumor Etiology and Screening Department of Cancer Institute and General Surgery, the First Hospital of China Medical University, Shenyang, 110001, China. [2]Key Laboratory of Cancer Etiology and Prevention in Liaoning Education Department, the First Hospital of China Medical University, Shenyang, 110001, China. [3]Key Laboratory of GI Cancer Etiology and Prevention in Liaoning Province, the First Hospital of China Medical University, Shenyang, 110001, China. [4]These authors contributed equally: Xunan Qiu, Rui Guo. ✉e-mail: bgwang@cmu.edu.cn; yhgong@cmu.edu.cn

Nevertheless, current studies have only delineated the function of a restricted set of CSR genes in tumors, which lacks comprehensiveness. Furthermore, traditional observational studies only delineate associations between genes and tumors; they do not identify potential causal links between the two and fail to account for the impact of other confounding factors, such as environmental influences, on tumor risk. Furthermore, randomized controlled trials (RCTs) on CSR genes in different tumors are challenging due to heavy workloads, time constraints, and cost considerations. Therefore, a robust methodology is required to comprehensively analyze the causal effects of CSR genes in multiple tumors.

Mendelian Randomization (MR), positioned at the intersection between observational studies and RCTs, is gaining prominence in bolstering causal evidence in observational trials[10,11]. MR uses genetic variants as IVs to investigate potential causal associations between lifetime exposures and outcomes[12]. Compared to statistical methods in traditional observational studies, MR mitigates the impact of confounders and reverse causation, given the random assignment of genetic variants at birth (prior to disease development)[13]. To the best of our knowledge, MR has been extensively used to scrutinize causal connections between various traits and diseases, such as fatty acids and schizophrenia, and lung function and cardiovascular disease[14,15]. Both summary-data-based MR (SMR) and two-sample MR (TSMR) enable to estimate pleiotropic or potentially causal associations between genetically determined traits (e.g., gene expression, DNA methylation and protein expression) and complex traits of interest (disease phenotypes as outcomes, e.g., tumors)[16–19]. Additionally, heterogeneity independent instrument (HEIDI) tests have been used to differentiate potential causal associations between extensive linkage disequilibrium (LD) in the genome. Genome-wide association studies (GWAS) also leverage genetic associations with traits by analyzing single nucleotide polymorphisms (SNPs) and integrating GWAS data with gene expression and methylation data. However, MR has yet to be used to explore potential causal relationships between CSR genes and the risk of some common cancer types.

In this study, we first analyzed potential causal associations between CSR genes expression, DNA methylation, and protein expression levels in 18 cancers using SMR or TSMR. We then gained insights into the potential regulatory mechanisms underpinning cancer risk variants through Bayesian colocalization. Finally, we explored the druggability and functional roles of identified CSR genes. We adhered to reporting guidelines outlined in the Strengthening the Reporting of Observational Studies in Epidemiology (Supplementary Table 1)[20,21].

Results

Instrumental variables selection for MR analysis

We initially selected SNPs associated with the expression of 748 CSR genes transcripts from cis-eQTL data in the eQTLGen database. Subsequently, SNPs corresponding to 3515 CSR genes DNA methylation CpG sites were chosen from cis-mQTL pooled data. Additionally, SNPs extracted from the deCODE cis-pQTL database were closely linked to the expression of 147 CSR proteins. Genetic variants exhibiting strong associations with CSR genes expression, DNA methylation, and protein expression were selected for analyses (Fig. 1, Supplementary Table 2). The study design and workflow are outlined (Fig. 1).

MR and colocalization analysis of cellular senescence genome-wide cis-eQTL with multiple cancer risks

Associations between SNPs representing CSR genes expression and cancer outcomes, as determined by SMR testing, are shown (Fig. 2, Supplementary Table 3). Following the correction of P-values (FDR < 0.05) and HEIDI tests (P_{HEIDI} > 0.01), we identified four association signals for four unique genetic loci in breast cancer, two association signals for two unique genetic loci in colorectal cancer, one association signal for one unique genetic locus in endometrial cancer, two association signals for two unique genetic loci in lung cancer, two association signals for two unique genetic loci in melanoma, and sixteen association signals for sixteen unique genetic loci in

prostate cancer. We did not find any significant genetic connections between CSR genes expression and 12 other cancer types except mentioned above.

Subsequent Bayesian colocalization analysis revealed that the expression of five genes exhibited shared causal variants with certain cancers. Specifically, those were CNOT6 with lung cancer, and DNMT3B, MAP2K1, TBPL1, and SREBF1 with prostate cancer (Figs. 3 and 4, Supplementary Table 4). For lung cancer, one SD increase in CNOT6 expression was associated with an 18% higher tumor risk (Odds Ratio (OR): 1.18, 95% Confidence Interval (CI): 1.10–1.27, FDR = 3.46×10^{-2}). For prostate cancer, one SD increase in DNMT3B expression was associated with a 73% higher risk (OR: 1.73, 95% CI: 1.48–1.97, FDR = 1.49×10^{-3}), one SD decrease in MAP2K1 expression was associated with a 24% lower risk (OR: 0.76, 95% CI: 0.66–0.86, FDR = 5.82×10^{-6}), one SD decrease in TBPL1 expression was associated with a 16% lower risk (OR: 0.84, 95% CI: 0.76–0.92, FDR = 2.81×10^{-3}), and one SD decrease in SREBF1 expression was associated with a 11% lower risk (OR: 0.89, 95% CI: 0.85–0.94, FDR = 4.82×10^{-4}).

MR and colocalization analysis of cellular senescence genome-wide cis-mQTL with multiple cancer risks

For causal relationships between CSR genes DNA methylation and multiple cancer outcomes, after FDR corrections and HEIDI testing (Supplementary Table 5), we identified three association signals for three unique genetic loci in colorectal cancer, five association signals for five unique genetic loci in endometrial cancer, ten association signals for ten unique genetic loci in lung cancer, five association signals for three unique genetic loci in oral cavity pharyngeal, 152 association signals for 115 unique genetic loci in prostate cancer, four association signals for three unique loci in gastric cancer, and one association signal in thyroid cancer. We did not find any significant genetic associations between CSR genes DNA methylation and other cancer types.

Bayesian colocalization analysis identified 18 genes exhibiting shared causal variations between DNA methylation and cancer (Supplementary Table 6). Furthermore, these analyses indicated that distinct genetic variants regulating these genes exerted varying effects on methylation levels, thereby influencing observed outcomes.

In colorectal cancer, NTN4 and SMAD6 hypermethylation was associated with an increased risk (OR: 1.25, 95% CI: 1.15–1.35, FDR = 1.77×10^{-2} and OR: 1.31, 95% CI: 1.18–1.44, FDR = 4.52×10^{-2}, respectively). In endometrial cancer, NF1 hypermethylation was associated with an increased risk (OR: 1.34, 95% CI: 1.21–1.48, FDR = 2.62×10^{-2}), whereas TERT hypermethylation was linked to a decreased risk (OR: 0.78, 95% CI: 0.66–0.90, FDR = 4.16×10^{-2}). For lung cancer, ATG12 hypermethylation (OR: 0.92, 95% CI: 0.88–0.96, FDR = 7.43×10^{-3}), DDAH2 hypermethylation (OR ranged from 0.65 to 0.75, FDR < 0.05), ENO1 hypermethylation (OR: 0.89, 95% CI: 0.83–0.95, FDR = 1.62×10^{-2}), and TP53BP1 hypermethylation (OR: 0.80, 95% CI: 0.70–0.90, FDR = 7.07×10^{-3}) all decreased tumor risk. For oral cavity pharyngeal, KCNJ12 hypermethylation was associated with an increased risk (OR ranged from 1.17 to 1.23, FDR = 4.38×10^{-2}), whereas NOTCH1 hypermethylation decreased tumor risk (OR: 0.51, 95% CI: 0.21–0.82, FDR = 4.38×10^{-2}). In prostate cancer, different methylation sites within the same gene potentially exerted distinct effects. Methylation sites cg26553763 and cg13636640 in six CpG sites associated with rs2424905, rs6087989, rs1883730, rs1003521, and rs4911106 in DNMT3B exhibited a negative link with tumor risk (OR: 0.83, 95% CI: 0.75–0.92, FDR = 3.47×10^{-3} and OR: 0.96, 95% CI: 0.94–0.97, FDR = 1.41×10^{-4}, respectively), while cg22052056, cg21235334, cg13788819, and cg09149842 showed positive associations with tumor risk (OR: 1.15, 95% CI: 1.09–1.22, FDR = 1.07×10^{-3}; OR: 1.13, 95% CI: 1.07–1.18, FDR = 1.03×10^{-3}; OR: 1.12, 95% CI: 1.07–1.17, FDR = 5.72×10^{-4}; and OR: 1.14, 95% CI: 1.08–1.19, FDR = 7.78×10^{-4}, respectively). Similarly, the effects of different methylation sites in FERMT1 and MAD1L1 on prostate cancer risk were either positive or negative; MVK, NEK6, NUDT5, and SMAD2 hypermethylation reduced tumor risk (OR: 0.91, 95% CI: 0.86–0.95, FDR = 3.47×10^{-3}; OR: 0.96, 95% CI: 0.95–0.98,

Fig. 1 | Study design and the workflow showing genetic variation selection and analysis methods. IVs instrumental variables, SNPs single nucleotide polymorphisms, SMR summary-data-based Mendelian Randomization, TSMR two-sample Mendelian Randomization, GO Gene Ontology, KEGG Kyoto Encyclopedia of Genes and Genomes.
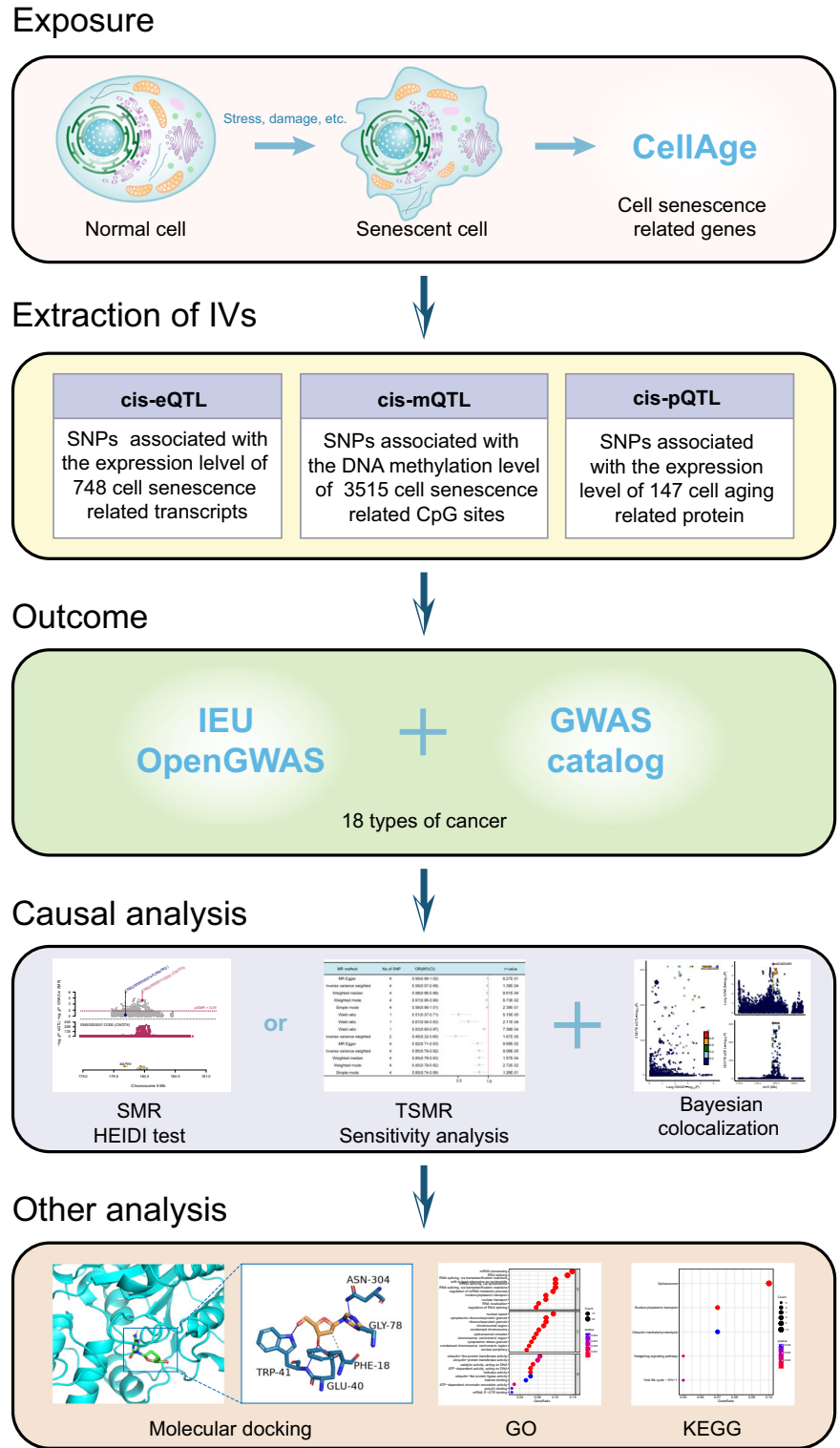
FDR = 6.28×10^{-4}; OR: 0.96, 95% CI: 0.94–0.98, FDR = 1.40×10^{-3}; and OR: 0.93, 95% CI: 0.90–0.96, FDR = 3.36×10^{-3}, respectively), and *SPI1* hypermethylation increased tumor risk (OR: 1.15, 95% CI: 1.08–1.21, FDR = 3.30×10^{-3}). Results are shown (Supplementary Table 5). For thyroid cancer, *SMAD3* hypermethylation was associated with an increased tumor risk (OR: 1.69, 95% CI: 1.47–1.92, FDR = 1.73×10^{-2}).

Moreover, altered DNA methylation patterns in specific CSR genes have a direct impact on their mRNA expression[22]. Similarly, we conducted SMR analyses to investigate causal relationships between methylation and CSR genes expression, aligning gene methylation to expression through shared genetic variations. Following FDR corrections and HEIDI testing, we compiled a list of CSR genes whose expression levels were influenced by DNA methylation at CpG sites (Supplementary Table 7). These findings indicated that the DNA methylation of *TERT, DDAH2, DNMT3B, MAD1L1, NOTCH1, TP53BP1, MAP2K1, SMAD2, SMAD3, SMAD6*, and *SREBF1* significantly impacted their gene expression levels.
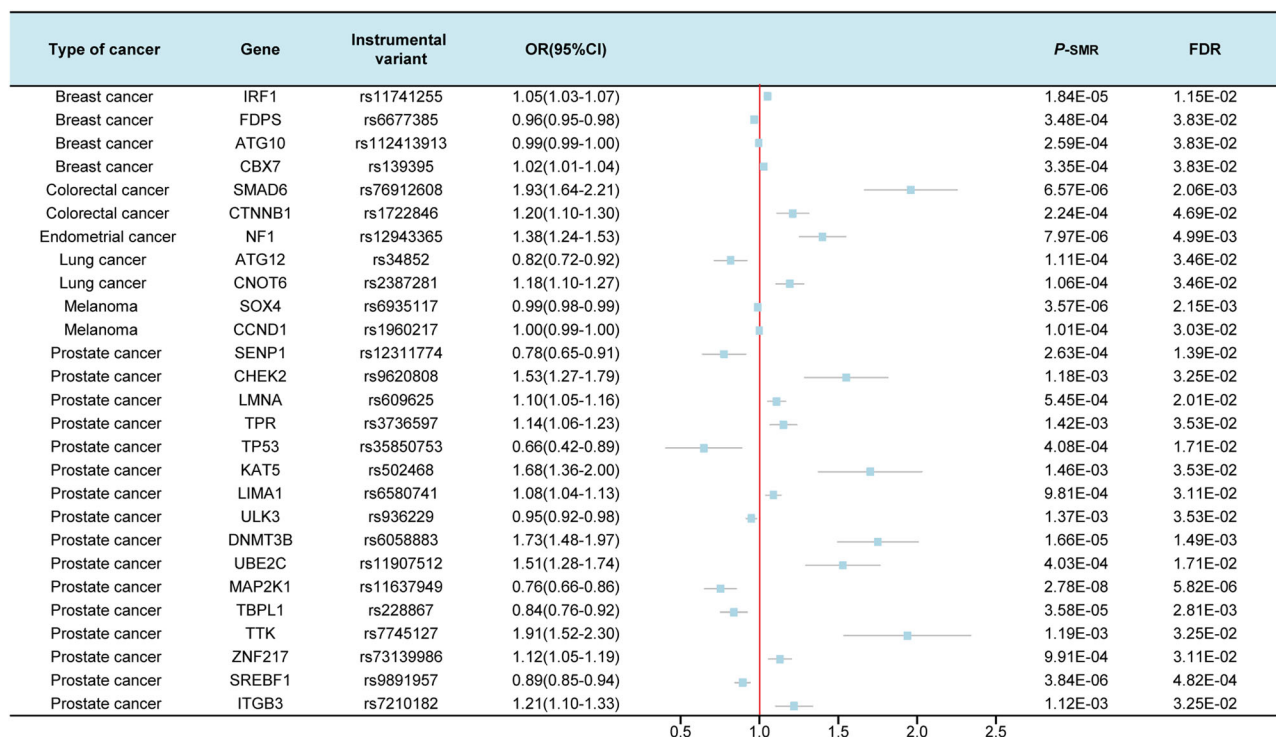
Fig. 2 | MR results showing associations between CSR genes expression and cancer risks. Instrumental variant: top SNP that best represents a gene. OR: Odds Ratio. 95% CI: 95% Confidence Interval. FDR: the false discovery rate.
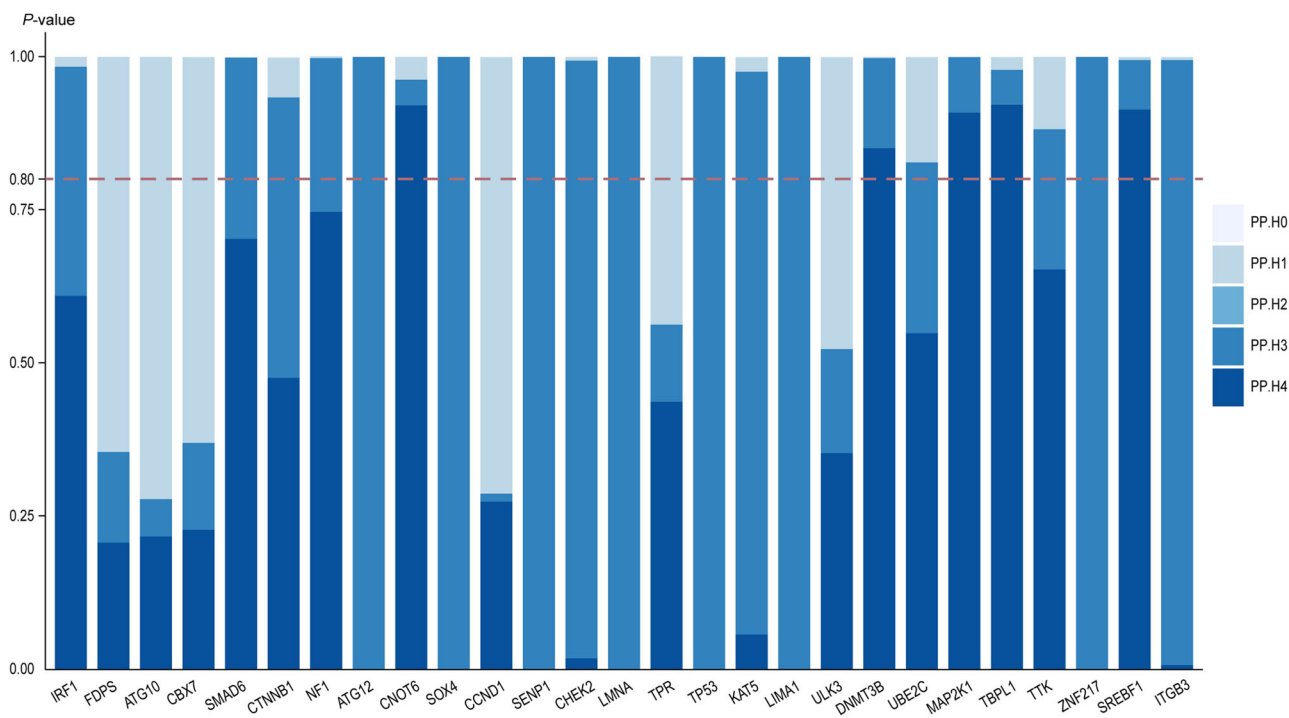

Fig. 3 | Colocalization results of eQTLs and GWAS. PP.H0: posterior probability of H1. PP.H1: posterior probability of H1. PP.H2: posterior probability of H2. PP.H3: posterior probability of H3. PP.H4: posterior probability of H4; PP.H4 > 0.8 as the cut-off for the evidence of colocalization of cancer GWAS and eQTL association.

MR and colocalization analysis of cellular senescence genome-wide cis-pQTL with multiple cancer risks

To investigate causal relationships between protein expression related to CSR genes and various cancer outcomes, we used five MR analysis methods (primarily the IVW method). Using heterogeneity, pleiotropy tests, and Bayesian colocalization, we concluded that elevated LAYN protein expression levels were linked to a reduced risk of prostate cancer (OR: 0.51, 95% CI: 0.37–0.71, FDR = 1.36×10^{-2}) and shared a causal variant with this disease (Supplementary Fig. 2, Supplementary Tables 8, 9).

A phenome-wide scan of identified genetic variants

To rule out possible pleiotropy in selected cancers and to further ruled out confounding factors related to SNPs, we performed a phenome-wide scan of variants using PhenoScanner (Supplementary Table 10). Some genes were associated with known secondary traits: rs7497064 (*SMAD6*-methylation associated) and rs731758 (*NF1*-methylation associated) were linked to secondary traits such as height, weight, and body fat metabolic rate; rs2736100 (*TERT*-methylation related) was associated with lung cancer, glioma, and testicular germ cell tumors; rs1270942 and others (*DDAH2*-methylation related) were associated with multiple traits such as asthma, lung cancer, hyperthyroidism, systemic lupus erythematosus, diabetes, and schizophrenia; rs12945438 and rs8078675 (*KCNJ12*-methylation related) were associated with body mass index; rs6087989 (*DNMT3B*-methylation related) was associated with immune cell counts; rs7783715 and rs11761270 (*MAD1L1*-methylation related) were associated with schizophrenia; rs10744826 (*MVK*-methylation related) was associated with cholesterol, low density lipoprotein (LDL), high density lipoprotein (HDL), etc.; rs17293632 (*SMAD3*-methylation related) was associated with coronary artery disease, asthma, and other allergic diseases; and rs4938784 and rs4938792 (LAYN protein expression related) were associated with allergic rhinitis, asthma, and other diseases. These observations suggested that genetic variants associated with secondary traits may have potentially introduced horizontal pleiotropy, necessitating further study to rule out pleiotropy.

Evaluating druggability and molecular docking analysis

We used DrugBank and ChEMBL platforms to explore CSR genes with causal relationships and successful colocalization identified to assess their potential as drug targets. In total, 17 drugs targeting four genes or proteins were identified, for example, hyaluronic acid (HA) for LAYN (Supplementary Table 11). We also conducted molecular docking analyses to illustrate receptor–ligand interactions between CSR genes and drugs (Fig. 5). It is noteworthy that due to secondary structure unavailability for some proteins and ligands, and the complex nature of fish oil, molecular docking was not feasible for some predictive drugs. However, the remaining compounds showed successful binding, with corresponding binding energies shown (Supplementary Table 11). The magnitude of binding energy reflected the likelihood of binding between a receptor and ligand. As the binding energy decreased, affinity and stability between a ligand and receptor increased.

Functional analysis

To explore the mechanisms of action of potentially causal genes associated with tumors, we performed functional and pathway enrichment analyses using GO (Fig. 6a–e) and KEGG (Supplementary Fig. 3a–e) methods. The genes involved in enrichment analysis are listed (Supplementary Table 12). In lung cancer, GO analysis revealed that the molecular function of *CNOT6* primarily revolved around ubiquitin-like protein transferase activity. Concurrently, KEGG analysis indicated significant enrichment in the spliceosome and the nucleoplasmic transfer pathway. CancerSEA also demonstrated a positive association between *CNOT6* and DNA damage, DNA repair, and cell proliferation and apoptosis at the single-cell level in lung cancer (Supplementary Fig. 4a, b). In prostate cancer, GO analysis showed that the principal molecular function of *DNMT3B* was in microtubule protein binding, while KEGG analysis highlighted its substantial enrichment in cell cycle pathways, among others. For *MAP2K1*, its primary molecular function was ubiquitin-like protein transferase activity, while KEGG analysis showed predominant enrichment in endocytosis and endoplasmic reticulum protein processing pathways. CancerSEA showed a positive relationship for *MAP2K1* with aging and inflammation at the single-cell level in prostate cancer (Supplementary Fig. 4c, d). For *TBPL1*, a major molecular function was related to redox activity, while KEGG analysis primarily emphasized the glutathione metabolic pathway. *SREBF1* was functionally characterized as being involved in oxidoreductase signaling

and demonstrated significant enrichment in AMP-activated protein kinase (AMPK) signaling and fatty acid metabolism pathways in KEGG analysis.

To explore possible protein expression roles causally related to tumors, we identified protein molecules closely related to the LAYN protein using the STRING database, including molecules such as MSN, TLN1, TLN2, etc. (Supplementary Fig. 4e). GO analyses indicated a significant molecular function enrichment for actin binding, while KEGG analyses showed significant enrichment in the tight junction pathway (Fig. 6f, Supplementary Fig. 3f).

Discussion

We demonstrated that CSR genes, characterized by genetic susceptibility, exerted causal effects in various tumors. We identified five gene expression levels, eighteen gene DNA methylation patterns, and one protein expression signature that were causally associated with cancer risk, with targeted drugs already developed against four proteins. The remaining genes or methylation sites represented potential targets for future small-molecule drug development studies against cancer. Our study linked genetic loci, gene expression, DNA methylation, and protein expression to different cancers, i.e., genetic variants may have influenced tumorigenesis by affecting gene expression, methylation levels, or protein expression. This provided robust evidence and promising applications for exploring underlying tumorigenesis mechanisms and reversing therapies.

The expression of five CSR genes was causally identified as related to cancer development, including *CNOT6* in lung cancer, and *DNMT3B*, *MAP2K1*, *TBPL1*, and *SREBF1* in prostate cancer. *CNOT6* is a deadenylating enzyme which degrades mRNA poly(A) tails and acts as a DNA mismatch repair regulator[23]. Notably, reduced *CNOT6* expression has been linked to distant metastasis in lung cancer[24], and our colocalization analyses revealed a potential regulatory role of the rs62405489 polymorphic site in *CNOT6*. Specifically, a one SD increase in *CNOT6* expression due to rs62405489 was associated with an 18% higher risk of lung cancer. Previously, the *CNOT6* rs2453176 SNP was identified as a potential functional susceptibility locus for lung cancer risk[25]. Our data implied that *CNOT6* rs62405489 may also serve as a susceptibility locus for lung cancer, providing insights on its regulatory expression mechanisms in this disease. Additionally, our comprehensive scanning analysis demonstrated that a causal relationship between *CNOT6* and lung cancer was not influenced by horizontal pleiotropy. Further functional enrichment analyses revealed that *CNOT6* exhibited ubiquitin-like protein transferase activity and was primarily enriched in the spliceosome and the nucleoplasmic transfer pathway. CNOT6 has previously been shown to target DNA-damaging proteins, regulate transcriptional activation, cell cycle control, apoptosis, senescence, and DNA repair processes[23]. Although no small molecule therapies targeting CNOT6 was found in DrugBank, our information provides valuable cues for future studies in designing targeted drugs against CNOT6, potentially blocking lung carcinogenesis.

For prostate cancer, both *MAP2K1* and *SREBF1* were reportedly associated with the disease and could potentially serve as therapeutic targets. Inhibitors targeting MAP2K1 and SREBF1 induced apoptosis in prostate cancer cells[26,27]. *TBPL1* regulates transcriptional functions, is essential in spermatogenesis, and has not been associated with prostate cancer before. We observed that high *TBPL1* expression reduced prostate cancer risks and shared causal variants. Whole-phenome scanning analysis confirmed that the causal relationship was not influenced by horizontal pleiotropy. KEGG analyses highlighted the glutathione metabolism pathway as the main pathway and suggested that *TBPL1*'s activator in this pathway could be used to design small-molecule drugs targeting prostate cancer.

We identified 18 CSR genes whose DNA methylation levels were causally associated with cancer development. *ENO1*, *DNMT3B*, and *MAD1L1* methylation levels were shown to influence the development or progression of lung and prostate cancers, respectively[28,29]. This underscored the significance of epigenetic regulatory mechanisms and highlighted epigenetically-driven therapeutic candidates,
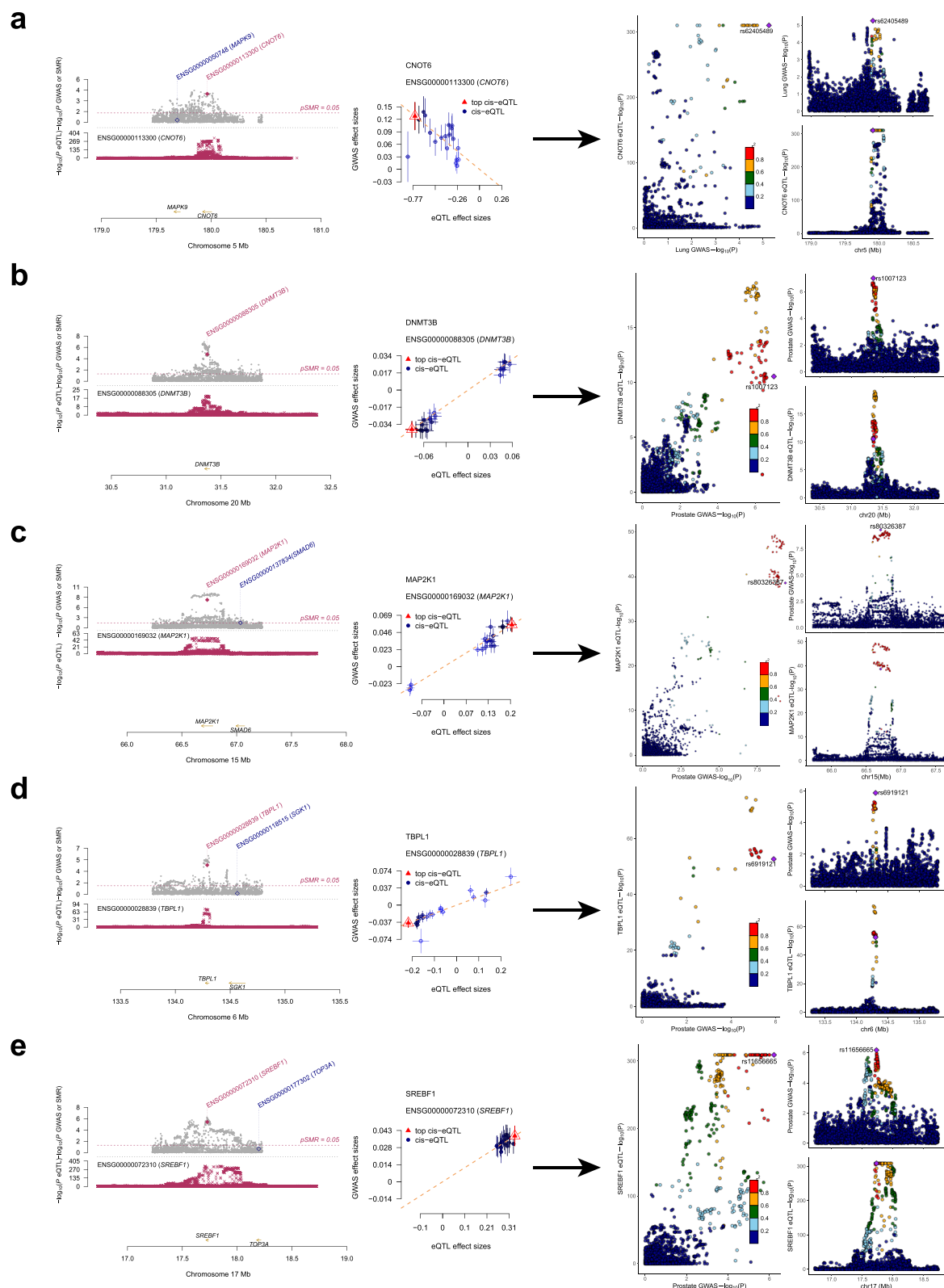
Fig. 4 | SMR and colocalization analyses prioritized CSR genes and cancer risks. Both sides of the arrow are shown SMR results between gene expression and cancer GWAS (FDR < 0.05; P_{HEIDI} > 0.01), loci comparisons between cis-eQTL and GWAS from colocalization analyses (PP.H4 > 0.8). **a–e** represents *CNOT6, DNMT3B, MAP2K1, TBPL1*, and *SREBF1* genes, respectively.

emphasizing their potential in future cancer therapeutic research. However, the relationships between methylation levels of the remaining 15 genes (*NTN4, SMAD6, TERT, NF1, ATG12, DDAH2, TP53BP1, KCNJ12, NOTCH1, FERMT1, MVK, NEK6, NUDT5,* *SMAD2* and *SMAD3*) and tumor risks remain unclear. Colocalization analyses further suggested that in aforementioned tumors, genetic polymorphisms and methylation may have acted in a coordinated manner to promote tumorigenesis.
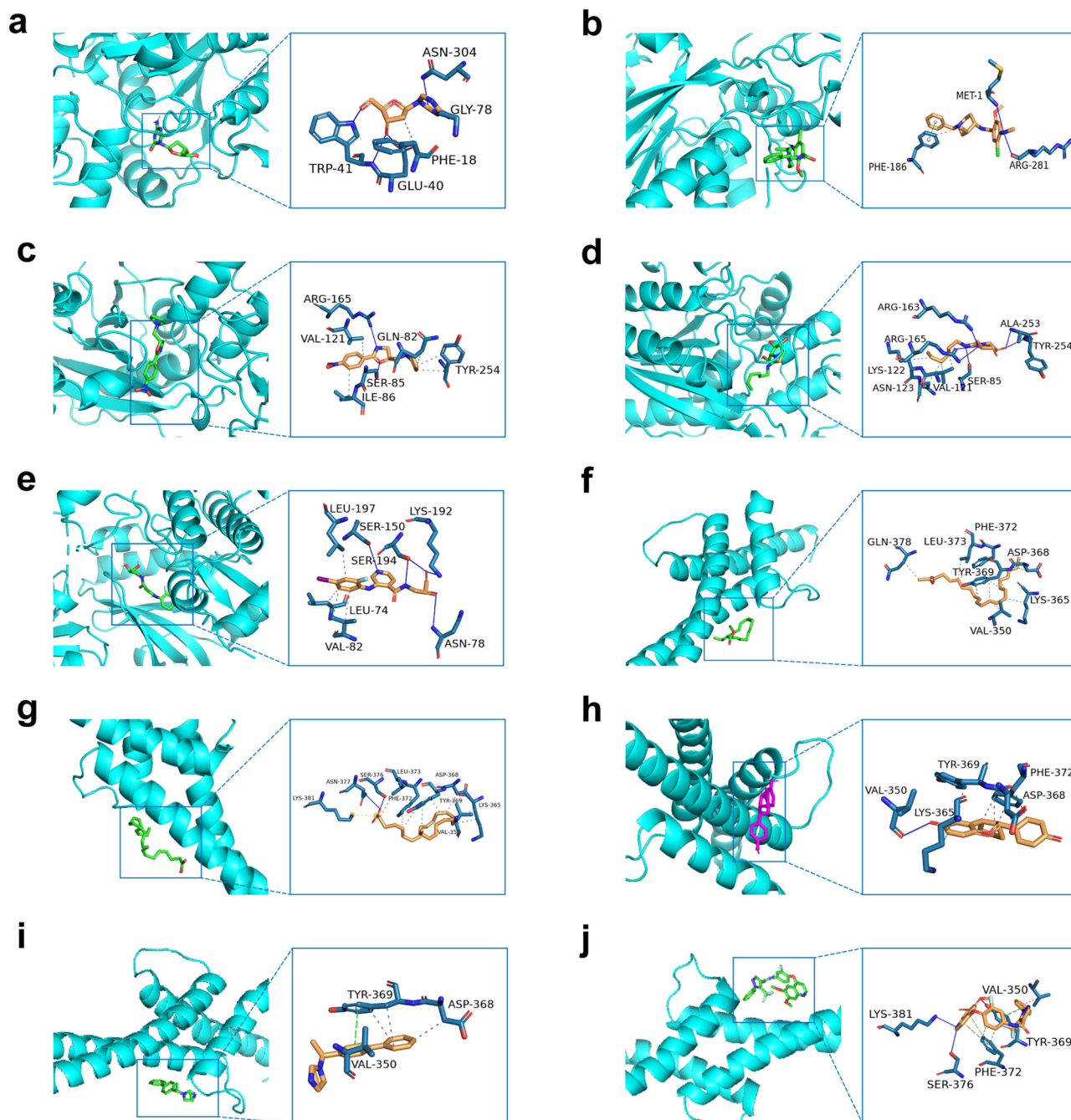
Fig. 5 | Three-dimensional structure diagram showing receptor and ligand docking. Crystal structures of acceptor and ligand complexes. The protein skeleton is a cartoon representation. Small molecular ligands are rod-like. **a–d** Three-dimensional stereogram showing DNMT3B bound to DB01262, CHEMBL109912, CHEMBL408017, and CHEMBL3916914 ligands. **e** Three-dimensional stereogram showing MAP2K1 bound to the DB14904 ligand. **f–j** Three-dimensional stereogram showing SREBF1 bound to DB11133, DB03756, CHEMBL278703, CHEMBL497499, and CHEMBL1087984 ligands.

Additionally, SMR analysis indicated that *SMAD6, TERT, DDAH2, TP53BP1, NOTCH1, DNMT3B, MAD1L1, SMAD2*, and *SMAD3* methylation levels may have influenced gene expression. Among genes, and as previously stated, *DNMT3B* expression was causally related to prostate cancer. *DNMT3B* is implicated in de novo methylation rather than methylation maintenance. Previous research indicated that methylation of the promoter region regulated its gene expression, and that *DNMT3B* overexpression was related to resistance to the androgen receptor antagonist enzalutamide, a prostate cancer therapy[30–33]. Our analyses suggested that *DNMT3B* CpG locus methylation modulated gene expression, which in turn influenced prostate carcinogenesis. Epigenetic interventions may prove beneficial for prostate cancer treatments, while targeted drug development against *DNMT3B* expression and methylation may have promising prospects.

Our study indicated that increased LAYN protein expression level was associated with a decreased risk of prostate cancer. Although not previously reported in the prostate cancer literature, elevated LAYN expression was positively correlated with tumor-associated macrophage infiltration and M2 macrophage polarization in colorectal, hepatocellular, and gastric malignancies. Moreover, its high expression signified an unfavorable prognosis[34,35]. These findings suggested that LAYN may have assumed diverse roles across different tumor types. KEGG pathway analyses revealed significant enrichment in the tight junction pathway, which requires further exploration. We also used DrugBank to predict HA as a potential target for
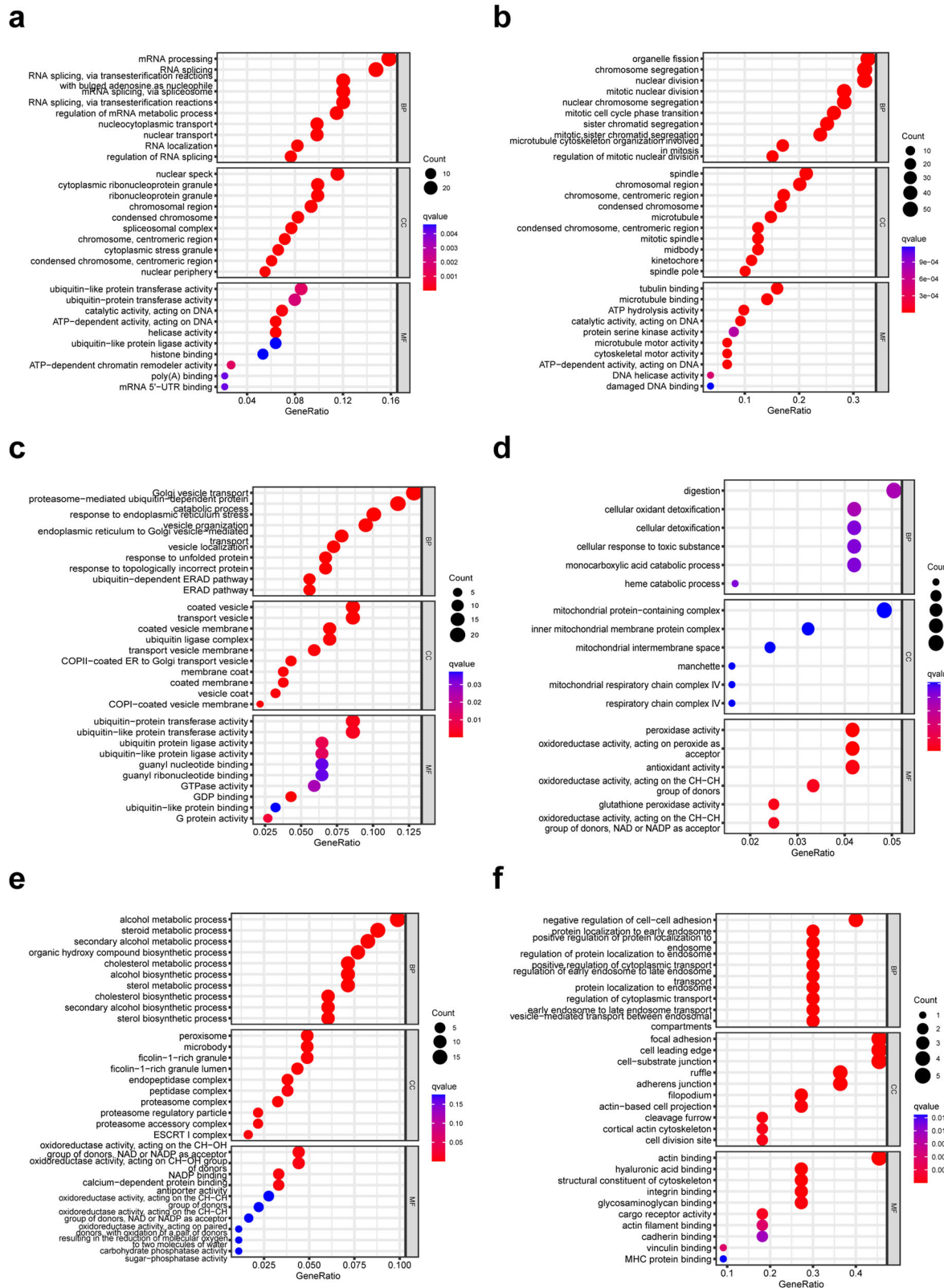
Fig. 6 | GO functional analysis of CSR genes. a–f represents *CNOT6, DNMT3B, MAP2K1, TBPL1, SREBF1*, and *LAYN* genes, respectively.

LAYN. HA is a glycosaminoglycan known to regulate cell adhesion and migration, and is currently used for joint pain relief, cancer treatment, wound healing, ophthalmic treatment, and cosmetic applications, among others. Notably, HA synthesis inhibitors (4-methylumbelliferone and sulfated HA) showed anti-tumor activity in prostate cancer[36,37]. Our study indicated that HA anti-tumor mechanisms may promote LAYN protein expression.

Additionally, in our study, we identified causal relationships between omics related to cellular senescence measured in blood and the risks of cancers in different organs. This raises the question: how can omics in blood

reflect the pathological processes occurring in the organ where the tumor is developing? We hypothesize that this may be related to the entry of circulating tumor cells, circulating tumor DNA, and the secretion of tumor markers into the bloodstream[38-41]. Furthermore, compared to tissue markers, blood markers offer several advantages, including convenient sample collection, reduced trauma, and the ability to conduct dynamic follow-up assessments. These factors render blood markers particularly suitable for early tumor detection and therapeutic applications. In summary, our findings provide valuable insights for the further exploration of potential and noninvasive biomarkers that are closely associated with primary tumors.

Our study had several strengths. Initially, we conducted a comprehensive and detailed MR analysis, incorporating genetic susceptibility data for CSR genes and their causal relationships with cancer. Given that genetic variations are randomly inherited, conditional on parental genotype, and genotypes are typically fixed from conception, we effectively eliminated selection bias and environmental confounders prevalent in previous research[13]. Furthermore, our substantial sample size and the inclusion of data from 18 different cancer outcomes, sourced from GWAS summary statistics, provided the statistical power to elucidate causality and draw conclusive estimates for various cancer types. We also used multiple MR analysis methods, integrating SMR with TSMR analyses, alongside sensitivity analyses, heterogeneity tests, and colocalization analyses. We complemented these approaches by examining genes in conjunction with multi-omics data, spanning transcriptomics, epigenomics, and proteomics, to reinforce the credibility of our findings.

Our study presented several limitations. Firstly, the sources of the summary statistics data for eQTL, mQTL, and pQTL are heterogeneous, and thus the statistical models used to generate these summary statistics were not uniformly adjusted for the same confounding factors. While, our study satisfied three core assumptions of MR as well as pleiotropy analyses were performed to furtherly remove the confounding factors, this ensured that the causal relationships between molecular traits and tumor risk were not affected by the confounders. Secondly, despite utilizing a substantial number of available data sources, the sample sizes for mQTL data, individual tumor GWAS data, and the number of genetic variants associated with protein expression in this study remain limited. Consequently, some genes potentially causally linked to cancer may have been overlooked. Thirdly, our data lacked variants associated with genes and associated expression on X and Y chromosomes. Moreover, the absence of GWAS datasets, specifically reflective of cellular senescence, impeded the assessment of causality directionality using bidirectional MR based on existing software resources. Whether cellular senescence is causally related to cancer warrants further investigations, particularly as more advanced GWAS methods become available.

In conclusion, we identified potential causal effects from CSR genes toward several cancers using MR analysis, which showed that cellular senescence mechanisms have important roles in tumorigenesis. Further exploration of the mechanisms underpinning these CSR genes in tumors may help prevent cancer and accelerate therapies. Additionally, our study has provided a robust theoretical foundation for the further exploration of CSR genes-based clinical drugs against cancer.

Methods
Data sources
To characterize the genetic susceptibility to cellular senescence, we retrieved 866 well-established CSR genes from the Human CellAge database[42]. The database is a manually curated database of human genes that drive cellular senescence that has undergone various analyses including molecular experiments, bioinformatics, and network biology.

Exposure data: Data for this MR study was sourced from several GWAS databases. The eQTLGen consortium (https://www.eqtlgen.org/cis-eqtls.html) provides information on thousands of SNPs associated with 10,317 traits from 31,684 individuals[43]. MR cis-mQTL instruments, reflecting genetic variants closely associated with CSR genes methylation,

were derived from pooled data (https://yanglab.westlake.edu.cn/software/smr/#mQTLsummarydata) in a meta-analysis of the two cohorts BSGS ($n = 614$) and LBC ($n = 1366$)[44]. MR cis-pQTL instruments, representing genetic variants associated with cellular senescence protein expression, were selected from the Icelandic proteome dataset (https://download.decode.is/form/folder/proteomics), which encompasses genome-wide association data for 4907 proteins[45]. To construct eQTL, mQTL and pQTL instruments for CSR genes, we extracted genetic variants located within 1000 kb on either side of the coding sequence (in cis), which were closely associated with gene expression, DNA methylation or protein expression.

Outcome data: GWAS cancer outcome summary statistics were retrieved from publicly available databases, covering 18 cancer types. Supplementary Table 2 shows all QTL and GWAS datasets used in this study. Most of the exposure and outcome data were derived from European populations. However, four outcome datasets, including those for colorectal cancer, gastric cancer, pancreatic cancer, and thyroid cancer, included a small proportion of individuals from East Asian populations (<30%). The specific population composition is detailed (Supplementary Table 2). Additionally, due to the different sources of data used for exposure and outcomes in this study, sample overlap was avoided from the beginning.

SMR analysis
Linux version 1.3.1 SMR software was used to execute SMR using default parameters from the command line (https://yanglab.westlake.edu.cn/software/smr/#Overview). MR has to satisfy three core assumptions: (1) the association assumption: there is a strong association between SNPs and exposure; (2) the independence assumption: SNPs do not associate with the outcome through the confounding pathway; and (3) the exclusivity assumption: SNPs do not directly affect the outcome, only indirectly through the exposure. To satisfy core assumptions, specific criteria for SNP inclusion are listed: (1) cis-QTL; (2) at least a suggestive $P < 5 \times 10^{-8}$ value; and (3) a minor allele frequency >0.01. And HEIDI tests were used to distinguish causality linkages and exclude pleiotropy. A P_{HEIDI} value ≤ 0.01 indicated that an observed association might be attributed to two distinct genetic variants in high LD with each other. Further, SNPs satisfying the above criteria were used as IVs, which ensured that the causal relationships between gene expression or DNA methylation and tumor risk were not affected by the confounders.

Due to the exploratory nature of our study, the false discovery rate (FDR) was used to adjust for multiple testing, and the FDR-corrected P threshold (FDR < 0.05) was applied to select significant probes.

TSMR analysis
In order to follow the core assumptions, SNPs need to fulfill the criteria: (1) at least a suggestive $P < 5 \times 10^{-8}$ value; (2) an F-statistic ≥ 10; and (3) a minor allele frequency >0.01. F-statistic was calculated as $(\beta_{\text{SNP-cellage}}/\text{SE}_{\text{SNP-cellage}})^2$ to represent the strength of each SNP. The $\beta_{\text{SNP-cellage}}$ denoted the estimated effect size of the SNP on cellular senescence, representing the genetic variant expression trait association. The $\text{SE}_{\text{SNP-cellage}}$ signified standard error of the sample. To eliminate the LD bias as much as possible, SNPs also fulfill a distance threshold of 10,000 kb and $r^2 < 0.001$. Further, SNPs satisfying the above criteria were used as IVs, which ensured that causal relationships between protein expression and tumor risk were not affected by the confounders. pQTL data were analyzed against GWAS data using five MR methods from the TwoSampleMR R package (https://mrcieu.github.io/TwoSampleMR/, version 0.5.6), including MR Egger, weighted median, inverse variance weighting (IVW), simple mode, and weighted mode approaches[46]. The IVW method is applicable when all SNPs meet the core assumption of being valid instrumental variables, and under these conditions, IVW results are the most reliable[47]. When more than 50% of the SNPs are valid instrumental variables, the weighted median method can be used to estimate the OR. However, if there is a multiplicity of validity or if more than 50% of the SNPs violate the core assumptions, the MR Egger method is preferred[48]. The other two methods were used as supplementary techniques. For single SNPs, the Wald ratio was used. Pleiotropy and heterogeneity tests

were conducted using MR Egger intercept tests and Cochran's Q statistic ($P > 0.05$ for all) to ensure robustness in our results. All analyses were executed using R software (http://www.R-project.org/, version 4.3.0).

Bayesian colocalization analysis

Bayesian colocalization analyses were used to assess the probability that two traits shared the same causal variance. Analyses were performed in the coloc R package (https://chr1swallace.github.io/coloc/, version 5.2.2) with default parameters[49]. This method scrutinized individual genetic regions associated with both phenotypes, exploring whether the same genetic variation propelled an association with both traits. By meticulously mapping causal variations across two traits, the approach minimized the identification of spurious pleiotropy and suggested the existence of a biological mechanism linking both traits through shared genetic predictors. Colocalization analysis encompassed five hypotheses (Supplementary Fig. 1): (1) H0: neither trait has a causal genetic variant; (2) H1: only trait 1 has a causal genetic variant; (3) H2: only trait 2 has a causal genetic variant; (4) H3: both traits have a causal genetic variant, but not the same variant; and (5) H4: both traits share the same causal variant. For the examined cancers in GWAS databases, for each major SNP, all SNPs within 100 kb upstream and downstream of the major SNP were retrieved for colocalization analysis to assess the posterior probability of a common causal variant for H4 (PP.H4). If this probability (PP.H4) between gene expression, DNA methylation, or protein expression and cancer exceeded 0.8, this suggested that robust evidence existed for colocalization between cancer GWAS and QTL.

Phenotype scanning

Using phenotype scanning, we conducted a comprehensive search in previous GWAS to uncover associations between identified QTL and other traits to the control of confounding factors. SNPs are deemed pleiotropic under the following criteria: (1) the association has genome-wide significance ($P < 5 \times 10^{-8}$); (2) the GWAS is mainly conducted in a population of European origin; and (3) the SNP exhibits an association with any known risk factor for the disease.

Druggability analysis

DrugBank (https://go.drugbank.com/) is a publicly available database of drug formulations containing 8865 compounds, including 1806 approved drugs and 7059 investigational or off-market drugs[50]. The database facilitates effective drug discovery and development by providing comprehensive molecular information on current drugs and their mechanisms, as well as easy access to interactions with targets. ChEMBL (https://www.ebi.ac.uk/chembl/) is an openly accessible chemical database, furnishing comprehensive data on molecules, providing information on chemical structures, identifiers, physicochemical properties, and biological activities[51]. It contains data exceeding 20.3 million bioactivity assessments and a compilation of 2.4 million distinct compounds. To explore interactions between CSR genes and marketed drug targets, we identified all drugs in DrugBank and ChEMBL with matching protein targets to understand their interactions.

Molecular docking analysis

For drugs with identified protein targets, the secondary structures of small molecules were retrieved from the PubChem[52] database (https://pubchem.ncbi.nlm.nih.gov/), while the Protein Data Bank (PDB) (https://www.rcsb.org/) was used to obtain three-dimensional protein structures[53]. Then, the retrieved small molecules were processed using Chem3D 19.0. AutoDockTools 1.5.6 was employed to dehydrate and add hydrogen atoms to the proteins, after which molecular docking simulations were conducted[54]. Finally, the Protein-Ligand Interaction Profiler[55] (PLIP) web tool (https://plip-tool.biotec.tu-dresden.de/plip-web) and PyMOL 2.5.2 software were used to visualize the receptor-ligand interactions.

Functional enrichment analysis

To further explore the function and possible molecular mechanism of genes or proteins linking CSR causally related to tumorigenesis, we investigated the biological functions of key genes using Gene Ontology (GO) functional annotations and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. The GEPIA database (http://gepia.cancer-pku.cn/) facilitated the identification of genes across different cancer types which exhibited similar mRNA expression patterns[56]. Selection of the top 200 genes for each key gene was based on a descending order of Pearson correlation coefficient values. Additionally, enrichment analysis process was executed using the cluster profiler package in R, with a Q-value threshold of <0.05 used as a significance criterion[57]. CancerSEA is a pioneering database used to analyze distinct functional states in diverse cancer cells at the single-cell level. It scrutinizes gene functions across 14 different functional states (including stemness, invasion, metastasis, etc.) by leveraging single-cell data from different cancers[58].

Protein-protein interaction (PPI) analysis

To explore protein expression, the STRING database (https://www.string-db.org) was used to analyze and construct PPI networks involving key proteins[59].

Statistics and reproducibility

The software used in this study includes Linux version 1.3.1 SMR software and R 4.3.0 software. In addition to this, the application of R packages and online websites is described in detail in the Methods section. In order to correct for multiple testing in performing SMR and TSMR analyses, the FDR method was used.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data used in this study are publicly available. Supplementary Table 2 provides detailed information on the source, sample size and ethnic composition of all data used.

Code availability

The main R packages used for analysis in this study include the TwoSampleMR R package (version 0.5.6) and the coloc R package (version 5.2.2), and some software and online websites were also used, such as: the Linux version 1.3.1 SMR software, AutoDockTools 1.5.6 software, the GEPIA database, the STRING database, etc., as described in the Methods section. No custom algorithms or code were used in this study. For more information about the analysis scripts or any specific code used in our study, please contact the corresponding author upon request.

References

1. Lee, S. & Schmitt, C. A. The dynamic nature of senescence in cancer. *Nat. Cell Biol.* **21**, 94–101 (2019).
2. Gorgoulis, V. et al. Cellular Senescence: Defining a Path Forward. *Cell* **179**, 813–827 (2019).
3. Shmulevich, R. & Krizhanovsky, V. Cell Senescence, DNA Damage, and Metabolism. *Antioxid. Redox Signal.* **34**, 324–334 (2021).
4. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12**, 31–46 (2022).
5. He, S. & Sharpless, N. E. Senescence in Health and Disease. *Cell* **169**, 1000–1011 (2017).
6. Campisi, J. Aging, cellular senescence, and cancer. *Annu Rev. Physiol.* **75**, 685–705 (2013).
7. Wang, K. et al. EZH2-H3K27me3-mediated silencing of mir-139-5p inhibits cellular senescence in hepatocellular carcinoma by activating TOP2A. *J. Exp. Clin. Cancer Res.* **42**, 320 (2023).

8. Zheng, H. et al. MAP4K4 and WT1 mediate SOX6-induced cellular senescence by synergistically activating the ATF2-TGFβ2-Smad2/3 signaling pathway in cervical cancer. *Mol. Oncol.* https://doi.org/10.1002/1878-0261.13613 (2024).

9. Bi, Y. et al. FBXW7 inhibits the progression of ESCC by directly inhibiting the stemness of tumor cells. *Neoplasma* **70**, 733–746 (2023).

10. Thanassoulis, G. & O'Donnell, C. J. Mendelian randomization: nature's randomized trial in the post-genome era. *JAMA* **301**, 2386–2388 (2009).

11. Burgess, S., Timpson, N. J., Ebrahim, S. & Davey Smith, G. Mendelian randomization: where are we now and where are we going. *Int. J. Epidemiol.* **44**, 379–388 (2015).

12. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–R98 (2014).

13. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**, k601 (2018).

14. Jones, H. J. et al. Associations between plasma fatty acid concentrations and schizophrenia: a two-sample Mendelian randomisation study. *lancet Psychiatry* **8**, 1062–1070 (2021).

15. Higbee, D. H., Granell, R., Sanderson, E., Davey Smith, G. & Dodd, J. W. Lung function and cardiovascular disease: a two-sample Mendelian randomisation study. *Eur. Respir. J.* **58**, https://doi.org/10.1183/13993003.03196-2020 (2021).

16. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet* **48**, 481–487 (2016).

17. Liu, Y. et al. Mendelian Randomization Integrating GWAS, eQTL, and mQTL Data Identified Genes Pleiotropically Associated With Atrial Fibrillation. *Front. Cardiovasc. Med.* **8**, 745757 (2021).

18. Zhou, D. Y. et al. Decreased CNNM2 expression in prefrontal cortex affects sensorimotor gating function, cognition, dendritic spine morphogenesis and risk of schizophrenia. *Neuropsychopharmacology*, https://doi.org/10.1038/s41386-023-01732-y (2023).

19. Lin, J., Zhou, J. & Xu, Y. Potential drug targets for multiple sclerosis identified through Mendelian randomization analysis. *Brain* **146**, 3364–3372 (2023).

20. Skrivankova, V. W. et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *JAMA* **326**, 1614–1621 (2021).

21. Skrivankova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* **375**, n2233 (2021).

22. Galamb, O. et al. Aging related methylation influences the gene expression of key control genes in colorectal cancer and adenoma. *World J. Gastroenterol.* **22**, 10325–10340 (2016).

23. Song, P. et al. CNOT6: A Novel Regulator of DNA Mismatch Repair. *Cells* **11**, https://doi.org/10.3390/cells11030521 (2022).

24. Maragozidis, P. et al. Poly(A)-specific ribonuclease and Nocturnin in squamous cell lung cancer: prognostic value and impact on gene expression. *Mol. Cancer* **14**, 187 (2015).

25. Zhou, F. et al. Susceptibility loci of CNOT6 in the general mRNA degradation pathway and lung cancer risk-A re-analysis of eight GWASs. *Mol. Carcinogenesis* **56**, 1227–1238 (2017).

26. McKinstry, R. et al. Inhibitors of MEK1/2 interact with UCN-01 to induce apoptosis and reduce colony formation in mammary and prostate carcinoma cells. *Cancer Biol. Ther.* **1**, 243–253 (2002).

27. Sun, Y. et al. Apoptosis induction in human prostate cancer cells related to the fatty acid metabolism by wogonin-mediated regulation of the AKT-SREBP1-FASN signaling network. *Food Chem. Toxicol.* **169**, 113450 (2022).

28. Karsli-Ceppioglu, S. et al. Genome-wide DNA methylation modified by soy phytoestrogens: role for epigenetic therapeutics in prostate cancer. *OMICS* **19**, 209–219 (2015).

29. Ma, L. et al. The essential roles of m(6)A RNA modification to stimulate ENO1-dependent glycolysis and tumorigenesis in lung adenocarcinoma. *J. Exp. Clin. Cancer Res.* **41**, 36 (2022).

30. Liu, W., Xie, A., Tu, C. & Liu, W. REX-1 Represses RASSF1a and Activates the MEK/ERK Pathway to Promote Tumorigenesis in Prostate Cancer. *Mol. Cancer Res.* **19**, 1666–1675 (2021).

31. Zhu, A. et al. DNMT1 and DNMT3B regulate tumorigenicity of human prostate cancer cells by controlling RAD9 expression through targeted methylation. *Carcinogenesis* **42**, 220–231 (2021).

32. Farah, E. et al. Targeting DNMTs to Overcome Enzalutamide Resistance in Prostate Cancer. *Mol. Cancer Ther.* **21**, 193–205 (2022).

33. Tzelepi, V. et al. Epigenetics and prostate cancer: defining the timing of DNA methyltransferase deregulation during prostate cancer progression. *Pathology* **52**, 218–227 (2020).

34. Yang, Y. et al. Targeting LAYN inhibits colorectal cancer metastasis and tumor-associated macrophage infiltration induced by hyaluronan oligosaccharides. *Matrix Biol. J. Int. Soc. Matrix Biol.* **117**, 15–30 (2023).

35. Cao, L., Zhu, L. & Cheng, L. ncRNA-Regulated LAYN Serves as a Prognostic Biomarker and Correlates with Immune Cell Infiltration in Hepatocellular Carcinoma: A Bioinformatics Analysis. *Biomed. Res. Int.* **2022**, 5357114 (2022).

36. Lokeshwar, V. B. et al. Antitumor activity of hyaluronic acid synthesis inhibitor 4-methylumbelliferone in prostate cancer cells. *Cancer Res.* **70**, 2613–2623 (2010).

37. Benitez, A. et al. Targeting hyaluronidase for cancer therapy: antitumor activity of sulfated hyaluronic acid in prostate cancer cells. *Cancer Res.* **71**, 4085–4095 (2011).

38. Lee, J. S. et al. Clinical Practice Guideline for Blood-based Circulating Tumor DNA Assays. *Ann. Lab. Med.* **44**, 195–209 (2024).

39. Liu, Y. et al. Significance of circulating tumor cells detection in tumor diagnosis and monitoring. *BMC Cancer* **23**, 1195 (2023).

40. Church, T. R. et al. Prospective evaluation of methylatedSEPT9in plasma for detection of asymptomatic colorectal cancer. *Gut* **63**, 317–325 (2014).

41. Hammarström, S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin. Cancer Biol.* **9**, 67–81 (1999).

42. Avelar, R. A. et al. A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol.* **21**, 91 (2020).

43. Võsa, U. et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet* **53**, 1300–1310 (2021).

44. McRae, A. F. et al. Identification of 55,000 Replicated DNA Methylation QTL. *Sci. Rep.* **8**, 17605 (2018).

45. Ferkingstad, E. et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet* **53**, 1712–1721 (2021).

46. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, https://doi.org/10.7554/eLife.34408 (2018).

47. Nie, D. et al. Association between green tea intake and digestive system cancer risk in European and East Asian populations: a Mendelian randomization study. *Eur. J. Nutr.* **63**, 1103–1111 (2024).

48. Bowden, J. et al. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat. Med.* **36**, 1783–1802 (2017).

49. Rasooly, D., Peloso, G. M. & Giambartolomei, C. Bayesian Genetic Colocalization Test of Two Traits Using coloc. *Curr. Protoc.* **2**, e627 (2022).

50. Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

51. Zdrazil, B. et al. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.*, https://doi.org/10.1093/nar/gkad1004 (2023).

52. Kim, S. et al. PubChem 2023 update. *Nucleic Acids Res.* **51**, D1373–D1380 (2023).

53. Burley, S. K. et al. RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res.* **51**, D488–D508 (2023).

54. Forli, S. et al. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **11**, 905–919 (2016).

55. Adasme, M. F. et al. PLIP 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA. *Nucleic Acids Res.* **49**, W530–W534 (2021).

56. Tang, Z. et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**, W98–W102 (2017).

57. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

58. Yuan, H. et al. CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res.* **47**, D900–D908 (2019).

59. Szklarczyk, D. et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).

Competing interests
The authors declare no competing interests.

Ethics approval
This study did not contain clinical studies or patient data and was based on large-scale summary GWAS datasets rather than individual-level data. In all corresponding original studies, all participants gave informed consent and no additional ethical approvals applied.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-06755-9.

Correspondence and requests for materials should be addressed to Bengang Wang or Yuehua Gong.

Peer review information *Communications biology* thanks Diptavo Dutta, Evangelina López de Maturana and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Tobias Goris. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.