Am J Transl Res 2024;16(11):7071-7086
www.ajtr.org /ISSN:1943-8141/AJTR0157931

Original Article

MMP9 in pan-cancer and computational study to screen for MMP9 inhibitors

Xianjie Ai[1], Xinyu Wang[2], Taotao Ren[1], Zhong Li[1], Bo Wu[1], Ming Li[1]

[1]Lower Extremity Division, Orthopedic Trauma Department, Honghui Hospital, Xi'an Jiaotong University, Youyi East Road No. 555, Beilin District, Xi'an, Shaanxi, China; [2]Department of Orthopaedic Trauma, Center of Orthopaedics and Traumatology, The First Hospital of Jilin University, Street Xinmin 71, Changchun, Jilin, China

Received May 21, 2024; Accepted September 26, 2024; Epub November 15, 2024; Published November 30, 2024

Abstract: Purpose: The stromal cell protein metalloproteinase 9 (MMP9), associated with extracellular matrix degradation and remodeling, promotes tumor invasion and metastasis and regulates cell adhesion molecule and cytokine activity. This study evaluated MMP9 in pan-cancer and screened for compounds and drug candidates that can inhibit it. Methods: MMP9 expression in pan-cancer tissues was evaluated in a pan-cancer dataset from the University of California Santa Cruz database, along with the correlation between MMP9 and the tumor microenvironment (TME), RNA modification genes, and tumor mutation burden. MMP9 crystal structures were downloaded, and a ligand-based pharmacophore model was constructed. A machine learning model was constructed for further screening. The identified compounds were pooled into Discovery Studio 4.5 for absorption, distribution, metabolism, and excretion (ADME) and toxicity prediction. Molecular docking was used to demonstrate the binding affinity and mechanism between the compounds and MMP9, and the stability of the ligand-receptor complex was assessed. Results: The expression levels of MMP9 differed between tumor tissues. Prognostic analysis showed that high MMP9 expression indicates poor survival and tumor progression in glioma (GMBLGG), pan-kidney (KIPAN; KICH+KIRC+KIRP), uveal melanoma (UVM), low-grade glioma (LGG), adrenocortical carcinoma (ACC), and liver hepatocellular carcinoma (LIHC). MMP9 expression in GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC was positively correlated with the TME. The ligand-based pharmacophore model and the machine learning model identified 49 small molecules. ADME and toxicity prediction identified CEMBL82047 and CEMBL381163 as potential MMP9 inhibitors, showing robust binding affinity with MMP9. The resulting complexes are stable in the natural environment. Conclusion: CHEMBL82047 and CHEMBL381163 are ideal compounds for inhibiting MMP9. The findings of this study will contribute to the design and improvement of MMP9-targeting drugs.

Keywords: MMP9, pan-cancer, ligand-based pharmacophore model, machine learning model, virtual screening, molecular dynamics simulation

Introduction

Tumors are formed when normal cells proliferate and differentiate abnormally under the action of various initiating and promoting factors. Tumors, especially malignant ones, destroy normal tissues and organs and can cause gradual organ dysfunction until failure or death due to compression, consumption, or destruction [1]. Malignant cancer is one of the leading causes of death worldwide [2], with an extremely low cure rate in developed and developing countries [3, 4]. Although great progress has been made in cancer therapy, many patients still have poor prognoses and low survival rates. Thus, novel therapeutic methods and drugs are urgently needed.

Matrix metalloproteinase-9 (MMP9), a member of the zinc-dependent endopeptidase family, is a gelatinase involved in a variety of biological processes (e.g., proteolytic extracellular matrix (ECM) degradation, cell-ECM and cell-cell interactions, and cell surface cleavage activities). In addition, it degrades and regulates ECM proteins and releases bioactive proteins, including cytokines, chemokines, and growth factors [5, 6]. MMP9 degrades type IV collagen and dis-

https://doi.org/10.62347/NXMR6806

rupts basement membranes associated with tumor invasion and metastasis. The expression level of MMP9 mRNA is significantly higher in nasopharyngeal carcinoma tissues than in nasopharyngeal tissues, and MMP9 overexpression accelerates tumor growth by inducing angiogenesis and enhanced local cell invasion and metastasis by degrading the ECM [7]. In esophageal cancer, MMP9 overexpression is significantly correlated with the depth of tumor infiltration, lymphatic infiltration, lymph node metastasis, and the degree of pathological differentiation [7]. The ECM is a key component of the local tumor microenvironment (TME) and undergoes extensive remodeling during breast cancer evolution. MMP9 is reported as a key player in ECM remodeling during cancer initiation and progression through a variety of mechanisms [8].

Currently, several chemotherapeutic agents target MMP9. MMP9-IN-1, a highly selective MMP9 inhibitor with oral efficacy [9, 10], selectively inhibits MMP9 to control the development, progression, invasion, and metastasis of nasopharyngeal carcinoma, but it also affects the function of the human respiratory system and reduces the activity of other proteases and cytokines because of its strong and effective inhibitory effect [7, 11]. JNJ0966 is another highly selective MMP9 inhibitor that blocks the conversion of MMP9 zymogen to a catalytically active enzyme [12]. However, it is currently only used in scientific research. Other MMP9 inhibitors exist but with extensive effect targets, which means they have more side effects. Therefore, novel MMP9-targeting drugs are needed.

This study combined a pharmacophore model and a machine learning model to screen for novel MMP9 inhibitors. Pharmacophores are combinations of characterized three-dimensional structural elements [13, 14] and have been used to design and screen new drugs on the basis of specific ligand structures [15, 16]. Machine learning is used to predict or classify drugs using data analysis [17] and is helpful in many fields, such as clinical data processing [18, 19]. We explored the role of MMP9 in pan-cancer and assessed the relevance of MMP9 in the tumor immune microenvironment and mRNA modifications. We then constructed a pharmacophore model and a machine learning model to screen for inhibitors of MMP9, followed by absorption, distribution, metabolism, excretion (ADME) and toxicity analysis, protein-ligand docking, and molecular dynamics (MD) simulation. This research provides a novel investigation strategy and a group of therapeutic candidates for MMP9, which might serve as a strong foundation for further agonist research.

Methods

Analysis of the expression level of MMP9 in pan-cancer datasets

The unified and standardized pan-cancer dataset TCGA (The Cancer Genome Atlas) Pan-Cancer (PANCAN, N = 10535, G = 60499) was downloaded from the University of California Santa Cruz (UCSC) database (https://xenabrowser.net/). The expression data of the ENSG00000100985 (MMP9) gene was extracted from each sample, and the samples from normal solid tissue, primary blood-derived cancer - peripheral blood, and primary tumors were further screened, followed by log2 (x+0.001) transformation of each expression value. Cancers with fewer than three samples were excluded. The difference in expression between normal and tumor samples in each tumor was calculated using R software (version 3.6.4), and significance analysis was performed using unpaired Wilcoxon rank sum and signed rank tests. Finally, a plot showing the differences in MMP9 expression between cancers was created.

Identification of the correlation between MMP9 expression levels and survival in pan-cancer

Several metrics (overall survival [OS] and progression-free survival [PFS]) were selected from TCGA samples to investigate the association between MMP9 expression and patient outcomes. A high-quality prognostic dataset (TCGA) was obtained from a previously published TCGA prognosis study published in *Cell*; cancers with fewer than 10 samples and samples with a follow-up time of less than 30 days were excluded. The R software package “survival” was used to obtain a forest map for Cox to analyze the relationship between MMP9 gene expression and survival in each tumor. The patients with each tumor type in the TCGA dataset were divided into two groups according

to the best cut-off value of MMP9 to compare the prognostic differences. The prognostic differences between the two groups were further analyzed using the "survfit" function of the R software package "survival", and the log-rank test was used to evaluate significant prognostic differences between the samples of different groups.

Association between MMP9 expression and the TME in pan-cancer

The gene expression profiles of each tumor were extracted separately, and the expression profiles were mapped to "Gene Symbol". The R software package "ESTIMATE" was used to calculate the stromal, immune, and ESTIMATE scores of each patient with each tumor type according to gene expression. The corr.test function of the R software package "psych" was used to calculate the Pearson's correlation coefficient between genes and immune invasion and immune cell invasion scores in each tumor to determine whether immune invasion scores were significantly correlated.

Correlation between MMP9 expression and mRNA-modifying genes in pan-cancer

The expression data of the marker genes of the MMP9 gene and three types of RNA modification genes (m1A, m5C, and m6A) in each sample were extracted. Primary blood-derived cancer - peripheral blood samples and primary tumor samples were screened, and the Pearson correlation coefficients between MMP9 and the marker genes of the five types of immune pathways were calculated by filtering all normal samples and transforming each expression value. These data were used to estimate the role of RNA modifications in cancer using the gene expression dataset and further summarize their therapeutic potential for abnormal deposition in cancer.

Association between MMP9 expression and tumor mutation burden in pan-cancer

Simple nucleotide variation data were downloaded from the database and processed. A simple nucleotide variation dataset was used to plot the mutational landscape of MMP9 in four tumor types. Tumor mutation burden (TMB) scores were calculated using mutation data of four tumor samples from TCGA, and patients were divided into low-TMB and high-TMB groups according to the TMB score quartile. Differentially expressed genes (DEGs) were identified in the low- and high-TMB groups.

Construction and verification of pharmacodynamic mass models

Pharmacophore models are useful for screening ideal compounds, and two types of pharmacophore models are known: structure-based pharmacological models derived directly from the X-ray structure of protein - ligand complexes and ligand-based pharmacological models derived from the structure of known active compounds. The crystal structures of human MMP9 receptors with different ligands (protein data bank [PDB] IDs: 2OW0, 2OW1, 4H3X, and 4WZV) were analyzed using LigandScout v4.3, which provides automated construction of three-dimensional pharmacophores. LigandScout identifies 3D chemical features; ligand options containing hydrogen bond donors (HBDs) and acceptors (HBAs) are shown as concentrated vectors, along with negative and positive ignitable spheres. Moreover, lipophilic regions are indicated by spheres. In addition, to expand selectivity, the LigandScout indicator incorporates spatial data about regions into each promising inhibitor. Pharmacophore signatures were entered into the web server Pharmit (http://pharmit.csb.pitt.edu/) to search for and identify small molecules that bind to the target molecule (MMP9 receptor) on the basis of structural and chemical similarities between small molecules. By combining the code from the PDB, 1,752,844 possible small molecules are obtained. Then, the deep learning model was built by DeepScreening (http://deepscreening.xielab.net/) for further screening, and the performance of the model was evaluated using test loss, accuracy, recall, precision, the F1 (F1-score), and Matthew's correlation coefficient (MCC).

ADME and toxicity prediction

The ADME module of Discovery Studio 4.5 was used to calculate the ADME of selected compounds, along with their water solubility, blood-brain barrier permeability, cytochrome P-450 2D6 (CYP2D6) inhibition, hepatotoxicity, human enteric absorption, and plasma protein binding levels. The topcat module of Discovery Studio

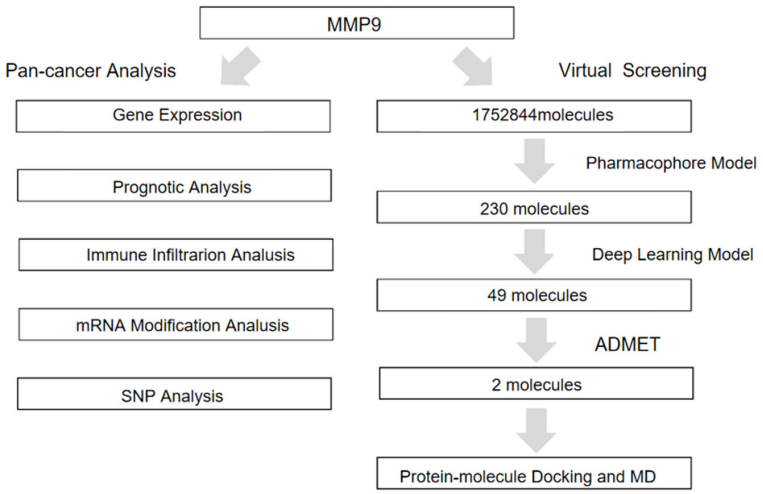


Figure 1. Mind map of this study.

4.5 was used to calculate the potential compounds' toxicity and other properties, such as the National Toxicology Program rodent carcinogenicity, the Ames mutagenicity, the developmental toxicity potential, the median oral lethal dose (LD50), and the chronic oral minimum observed adverse reaction level (LOAEL) in rats. These pharmacological properties were considered when selecting appropriate drug candidates for MMP9.

Protein molecule docking

Molecular docking was assessed using the Glide module of the Schrödinger kit to collect the active conformation of small molecules interacting with the MMP9 receptor. Top-level compounds from the pharmacophore screening were prepared in Maestro using the LigPrep module to obtain the starting structure for docking. Ligand-acceptor interactions included hydrogen bond interactions, van der Waals interactions, π-π stacked interactions, and ionic interactions. The molecular docking results were analyzed according to the binding energy (kcal/mol) between small molecules and amino residues and the number of binding interactions.

Molecular dynamics simulation

The best binding conformations of the ligand-MMP9 complexes among the potential compounds predicted by the molecule docking program were submitted to the MD simulation using Discovery Studio 4.5. The ligand-acceptor complex was placed into an orthogonal box and solvated with an explicit periodic boundary-solvated water model. To simulate the physiological environment, sodium chloride was added to a system with an ionic strength of 0.145. The system was then subjected to a CHARMM force field for analogy-based ligand parameterization. For this system, the following simulation protocols were applied: 1000 minimization steps for the fastest descent and conjugate gradient; 5ps equilibrium simulation at 300 K (slow drive 2ps from 50 K initial temperature) and atmospheric pressure; 25ps-MD simulation (production mode) at NPT (atmospheric pressure and temperature). The Particle Grid Ewald (PME) algorithm was used to calculate remote electrostatic, and the Linear Constraint Solver (LINCS) algorithm was used to fix all bonds involving hydrogen. With the initial complexity setting as a reference, the trajectories of the root mean square deviation (RMSD), potential energy, and structural features were determined by the Discovery Studio 4.5 analysis trajectory protocol.

Results

MMP9 expression in pan-cancer

The complete data analysis process is depicted in **Figure 1**. We analyzed the expression data of 26 cancer types and found that MMP9 was highly expressed in the vast majority of tumor samples. The expression differed significantly between most tumors, including glioblastoma multiforme (GBM), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), lung adenocarcinoma (LUAD), colon adenocarcinoma (COAD), colon adenocarcinoma/rectum adenocarcinoma esophageal carcinoma (COADREAD), breast invasive carcinoma (BRCA), esophageal carcinoma (ESCA), stomach and esophageal carcinoma (STES), kidney renal papillary cell carcinoma (KIRP), kidney pancreas carcinoma (KIPAN), stomach adenocarcinoma (STAD), prostate adenocarcinoma (PRAD), uterine corpus endometrial carcinoma (UCEC), head and

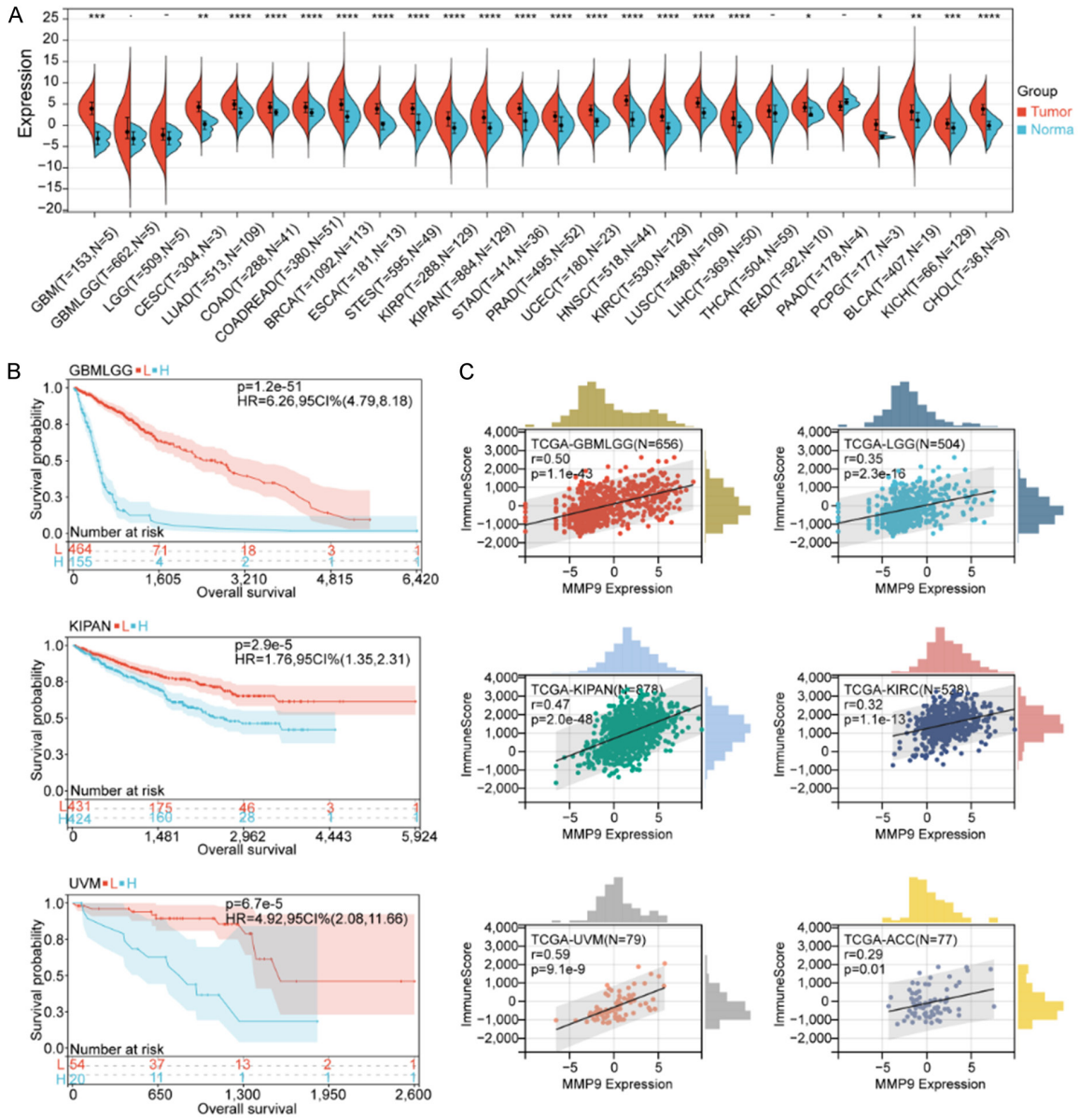


Figure 2. Pan-cancer analysis of MMP9 expression. A. Differential expression of MMP9 between tumor and normal tissues in pan-cancer analysis. MMP9 expression correlates with overall survival time (OS). B. Survival curves of MMP9 expression in GBMLGG, KIPAN and UVM. L represents low expression of MMP9 group, and H represents high expression of MMP9 group. C. Pan-cancer cohort (GBMLGG, KICH, KIRC, KIRP, KIPAN and UVM). Correlation between MMP9 expression and immune scores.

neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), lung squamous cell carcinoma (LUSC), liver hepatocellular carcinoma (LIHC), rectum adenocarcinoma (READ), pheochromocytoma and paraganglioma (PCPG), bladder urothelial carcinoma (BLCA), kidney chromophobe carcinoma (KICH), and cholangiocarcinoma (CHOL) ($P < 0.05$). MMP9 was highly expressed in brain low-grade glioma (LGG), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), and pancreatic adenocarcinoma (PAAD); however, because of the small sample size of the control group (normal), no significant differences were detected. Furthermore, the expression of MMP9 in thyroid carcinoma (THCA) did not differ significantly from that of normal samples (**Figure 2A**).

Pan-cancer prognostic analysis of MMP9

To further explore the association between MMP9 and the prognosis of pan-cancer, we performed prognostic analysis on 39 cancer types. The OS results (Supplementary Figure 1A) showed that for glioma (GMBLGG), KIPAN, uveal melanoma (UVM), LGG, adrenocortical carcinoma (ACC), liver hepatocellular carcinoma (LIHC), BLCA, and testicular germ cell tumors (TGCTs), higher MMP9 expression was associated with a lower survival rate ($P < 0.05$). For skin cutaneous melanoma (SKCM) and SKCM-M, higher MMP9 expression was associated with a higher survival rate, suggesting that MMP9 is a beneficial factor for these two tumor types ($P < 0.05$). For the other 28 tumors, expression was not significantly associated with survival ($P > 0.05$). We also plotted the survival curves of GMBLGG, KIPAN, UVM, LGG, ACC, KIRC, LIHC, BLCA, and TGCT (**Figure 2B**, Supplementary Figure 1B, 1C). In addition, we analyzed the PFS of pan-cancer (Supplementary Figure 1D) and found that for GMBLGG, KIPAN, KIRC, UVM, LGG, ACC, THCA, GBM, and KICH, higher MMP9 expression was associated with faster tumor progression ($P < 0.05$); for lymphoid neoplasm diffuse large B-cell lymphoma (DLBC) and ovarian serous cystadenocarcinoma (OV), higher MMP9 expression was associated with slower tumor progression, suggesting that MMP9 is a suppressor of tumor development in these cancers ($P < 0.05$). For the other 28 tumors, MMP9 expression was not significantly associated with tumor progression ($P > 0.05$). In summary, higher MMP9 expression was associated with a lower survival rate and tumor progression in GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC.

Correlation between MMP9 expression, the TME, and immune infiltration

The TME is composed of various components, such as immune cells, non-immune stromal cells, and ECM proteins, including innate immune cells, adaptive immune cells, extracellular immune factors, and cell surface molecules. TME, also known as the tumor immune microenvironment (TIME), has unique internal interactions and plays an important role in tumor biology [20, 21]. To further explore the correlation between MMP9 and tumor immune infiltration, we performed immune analysis on six tumors with MMP9 expression. We found that MMP9 expression in GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC was positively correlated with the immune score, ESTIMATE score, and stromal score (**Figure 2C**, Supplementary Figure 2).

In addition, we analyzed the correlation of MMP9 expression with immune cells in each tumor (**Figure 3A**). We found that macrophages were significantly associated with MMP9 expression. Specifically, M0 macrophages were significantly positively correlated with MMP9 expression in all six tumors; classically activated M1 macrophages were positively correlated with MMP9 expression in GMBLGG, KIPAN, UVM, LGG, and ACC; alternative activated M2 macrophages were positively correlated with MMP9 expression in GMBLGG and LGG. High macrophage expression leads to the release of more cytokines (such as epidermal growth factor (EGF)), which promotes the metastasis and invasion of cancer cells [22, 23]; this may explain the high correlation between MMP9 expression and metastasis. Monocytes were negatively correlated with MMP9 expression in five tumors but not in ACC, suggesting that the ability to recognize and kill tumor cells was inhibited [24]. Activated natural killer cells were negatively correlated with MMP9 expression in GMBLGG, KIPAN, KIRC, and ACC, indicating that their ability to kill tumor cells decreases when tumors express more MMP9. Furthermore, MMP9 expression was positively correlated with regulatory T cells (Tregs) in GMBLGG, KIPAN, UVM, LGG, and KIRC, which could suppress the immune system [25].

Correlation of MMP9 expression with RNA modification genes

Chemical RNA modifications play an important role in fundamental cellular processes, such as cell differentiation, protein production, cell signaling, and the maintenance of circadian rhythms [26, 27], and these modifications can be critical in tumor suppression or tumor-promoting effects. We found that GBMLGG was positively correlated with most of the genes in m1A modification, with significant differences between tumor types; the gene *ALKBH3* was positively associated with MMP9 expression in four tumors - GBMLGG, KIPAN, ACC, and LGG - with statistically significant differences between tumors (**Figure 3B**). ALKBH3 can pro-

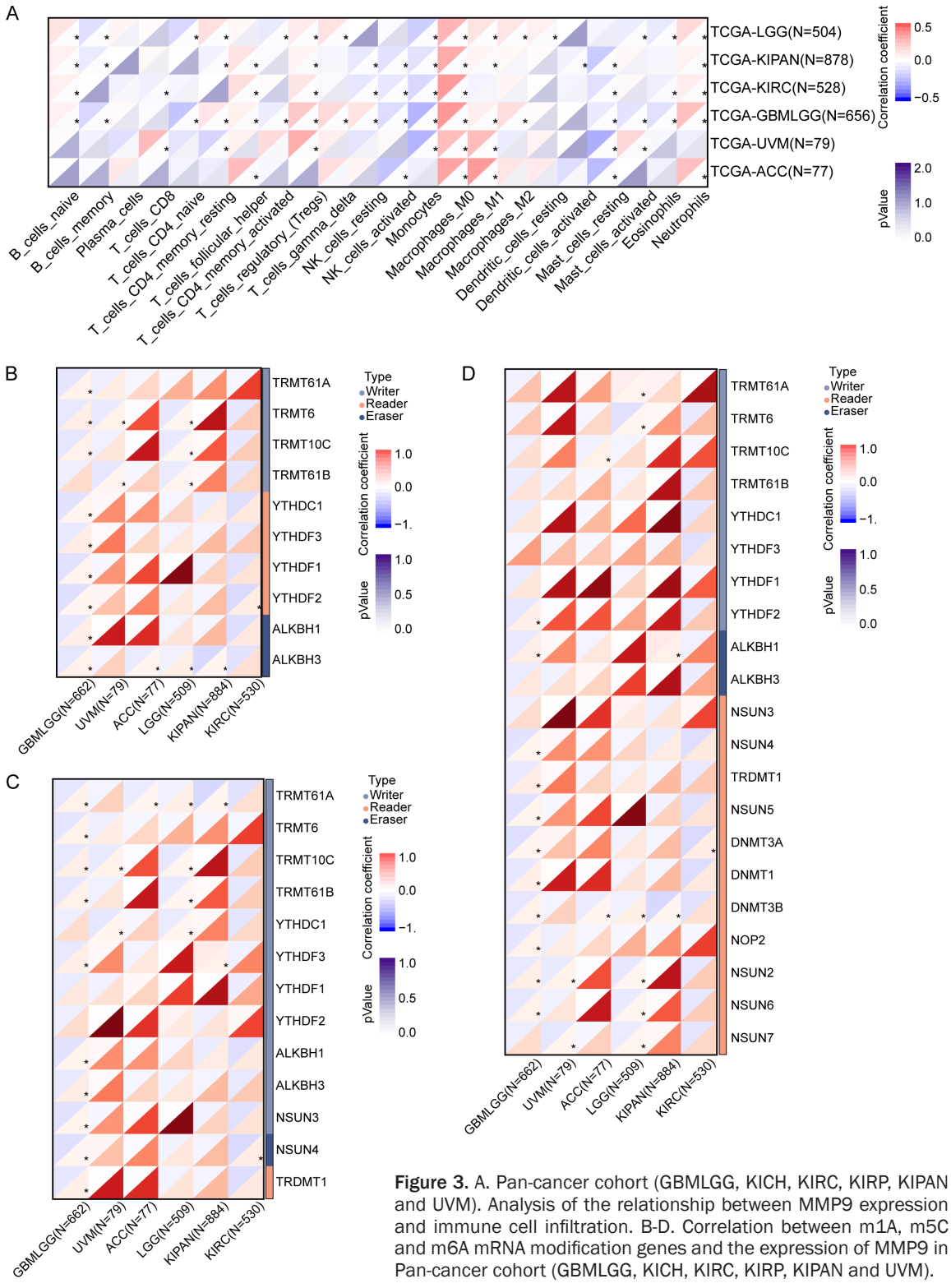


Figure 3. A. Pan-cancer cohort (GBMLGG, KICH, KIRC, KIRP, KIPAN and UVM). Analysis of the relationship between MMP9 expression and immune cell infiltration. B-D. Correlation between m1A, m5C and m6A mRNA modification genes and the expression of MMP9 in Pan-cancer cohort (GBMLGG, KICH, KIRC, KIRP, KIPAN and UVM).

mote the proliferation, migration, and invasion of cancer cells [28]. In m6A modification (**Figure 3C**), MMP9 expression was positively correlated with most genes in GBMLGG, with significant differences between tumor types. TRMT61A was positively correlated with MMP9 expression in four tumors - GBMLGG, KIPAN, ACC, and LGG - with statistically significant dif-

ferences between tumors (**Figure 3C**). A58 in m1tRNA is composed of the RNA-binding component TRMT6 and the catalytic component TRMT61A, which is crucial for maintaining m1tRNA stability, affects translation initiation, and has profound effects on various biological processes [29]. In m5C modification (**Figure 3D**), MMP9 expression was positively correlated with most genes in GBMLGG, with significant differences between tumor types. DNMT3B was positively correlated with MMP9 in four tumors - GBMLGG, KIPAN, ACC, and LGG - with significant differences between tumors (**Figure 3D**). DNMT3B is involved in de novo DNA methylation in embryonic stem cells and early embryos. It is overexpressed in several human tumors and is an indicator of early tumor recurrence and poor prognosis in hepatocellular carcinoma [30].

Correlation of MMP9 expression with TMB

We further performed single nucleotide polymorphism (SNP) analysis by dividing patients into two groups: a high MMP9 expression group and a low MMP9 expression group. In LGG (Supplementary Figure 3A), the genes *IDH1*, *TP53*, and *ATRX* had high mutation frequencies (> 20%), and *EGFR*, *MYH13*, *EPPK1*, *MYO15A*, *SI*, *KIAA1109*, *CDH17*, *SLCO1B1*, *SYNE2*, *CFAP47*, *SSPO*, and *ZFHX4* also had higher mutation rates and more mutation types in the high MMP9 expression group. In KIRC (Supplementary Figure 3B), the genes *VHL* and *PBRM1* had high mutation frequencies (> 20%), and the mutation types were mostly missense mutations, frameshift deletion mutations, nonsense mutations, splice site mutations, and in-frame insertions. *THSD7B*, *ADGRV1*, *XPO7*, *LAMC2*, and *UBR4* also had higher mutation rates and mutation types in the MMP9 high expression group. *TP53*, *CTNNB1*, and *MUC16* showed high mutation frequencies (> 20%) in ACC (Supplementary Figure 3C) as well as higher mutation rates in the high MMP9 expression group. *DST*, *FAT4*, *ASXL3*, *CNTNAP5*, and *NF1* also had higher mutation rates and more mutation types in high MMP9 expression group. In UVM (Supplementary Figure 3D), the genes *GNAQ*, *GNA11*, *BAP1*, and *SF3B1* had high mutation frequencies (> 20%), whereas *BAP1* had a higher mutation frequency in the high MMP9 expression group. Finally, *SF3B1* and *EIF1AX* showed higher mutation rates in patients with high MMP9 expression.

Construction and validation of the pharmacophore model

To further screen for novel inhibitors of MMP9, we constructed a ligand-based pharmacophore model. We first considered evaluating the major residues obtained by analyzing the crystal structures (PDB IDs: 2OW0, 2OW1, 4H3X, and 4WZV) to obtain the major residues of the MMP9 receptor (**Figure 4A-D**), identifying small active molecules and target proteins and the physicochemical interaction patterns between them and then mapping them to 3D array features (e.g., hydrogen bonds, lipophilic contacts, and ionic or aromatic interactions).

As shown in **Figure 4A**, the crystal structure 2OW0 exhibited two hydrophobic interactions, binding with the residues TYR423, LEU397, LEU418, VAL398, and ZN444. Two hydrogen bond acceptors were found with ALA189, GLN402, HOH503, HOH608, and LEU188. In addition, a positively ionized region was also detected. The crystal structure 2OW1 (**Figure 4B**) exhibited two hydrophobic interactions, binding with the residues VAL398, LEU418, TYR423, LEU397, and ZN444. Five hydrogen bond acceptors were found with LEU188, HOH593, HOH557, ALA189, and GLN402, and three hydrogen bond donors were also observed, along with a positively ionized region. The crystal structure 4H3X (**Figure 4C**) exhibited two hydrophobic interactions, binding with the residues LEU243, TYR248, VAL223, and ZN301. Two hydrogen bond acceptors were found with LEU188 and ALA189, and three hydrogen bond donors with ALA189, HIS226, and HOH415 were also observed, along with a positively ionized region. The crystal structure 4WZV (**Figure 4D**) exhibited two hydrophobic interactions, binding with the residues TYR245, MET247, ZN302, VAL223, and TYR248. Four hydrogen bond acceptors were found with ALA191, HOH401, LEU188, and ALA189, and ALA189, HIS230, and GLU227 hydrogen bond donors were also observed, along with a positively ionized region. As shown in Supplementary Figure 4A-D, these compounds exerted the largest effect with the amino acid residue H401.

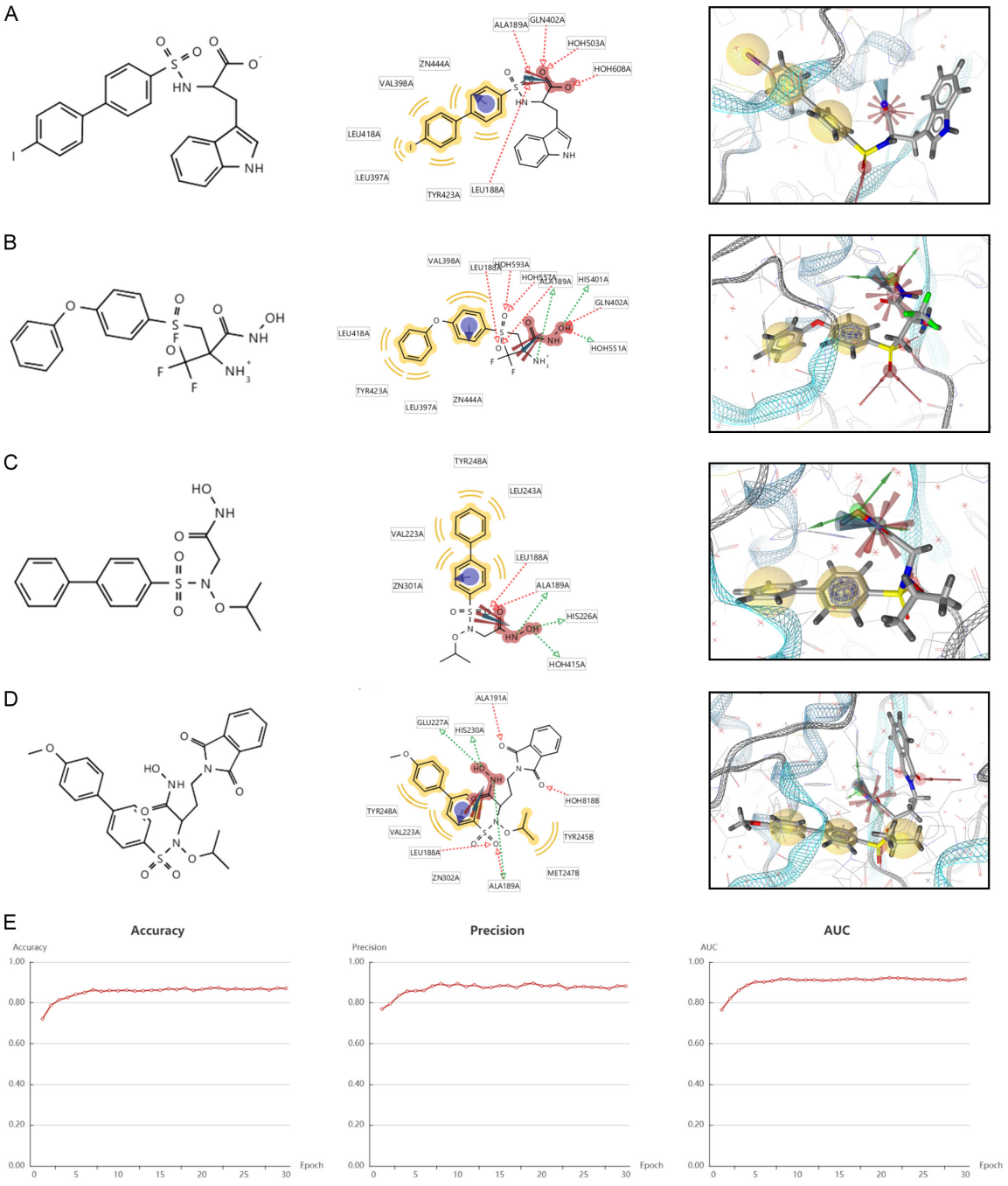


Figure 4. Chemical structure formula and pharmacophore analysis of (A) 20W0, (B) 20W1, (C) 4H3X and (D) 4WZV. Chemical features of the co-crystal structures were analyzed for summarizing common features. Red arrows indicate hydrogen bond acceptors, green arrows indicate hydrogen bond donors and yellow spheres indicate hydrophobes. (E) Evaluation index of deep learning model.

Virtual screening

We performed a prospective virtual screening (VS) of a database of compounds of natural origin and synthetic drugs, in which we used fitted values as pharmacology-based screening criteria. After removing duplicates, we screened 230 small molecules with the same pharmacophore from 1,752,844 small molecules. Then, we built a deep learning model with MMP9 and 3479 small molecules and validated it. The accuracy, precision, and area under the curve

(AUC) of the model gradually stabilized with increases in the Epoch, and finally stabilized at around 0.9 (**Figure 4E**). Recall and F1 also gradually stabilized around 0.9, and loss and MCC gradually stabilized around 0.45 and 0.7, respectively (Supplementary Figure 4E-H). After screened by the machine learning model, 49 small molecules (score = 1) from the 230 small molecules were identified.

ADME and toxicity prediction

Pharmacokinetics is an important analytical method for detecting effective compounds in the process of drug discovery, and the analysis of its properties plays a key role in drug design (Supplementary Table 1). Water solubility predictions (defined in water at 25°C) indicated that 33 compounds were soluble in water. In addition, 21 compounds showed good human intestinal absorption levels. Furthermore, 40 compounds were highly bound to plasma proteins, whereas the rest were not. CYP2D6 is an important enzyme involved in drug metabolism, and all 49 compounds were predicted to be non-inhibitors of cytochrome P450 2D6 (CYP2D6). Regarding hepatotoxicity, seven compounds were predicted to be nontoxic. CHEMBL82047 and CHEMBL381163 have good water solubility, intestinal absorption, and protein binding and can act as non-inhibitors of CYP2D6 without hepatotoxicity (Supplementary Table 2). We conducted a comprehensive investigation of the safety of these small molecules; the results showed that two small molecules, CEMBL82047 and CEMBL381163, are non-mutagenic and predicted to have less Ames mutagenic, rodent carcinogenic, and developmental toxicity potential than other compounds.

Protein molecular docking

To further study the binding properties of small molecules to proteins, we carried out molecular docking experiments (**Figures 5**, **6A**, **6B**, Supplementary Figure 5A-D). As shown in **Table 1**, CEMBL82047 and CEMBL381163 have higher binding affinity to the protein compared with the drugs JNJ0966 and MMP9-IN-1. Supplementary Figure 5E, 5F shows the π-dependent interactions and hydrogen bonds determined by the structural calculations. The results of the structural calculation studies showed that CEMBL82047 forms four pairs of hydrogen bonds with the MMP9 residue acceptor, and the complex itself forms four pairs of π-related interactions with the MMP9 residue acceptor. CHEMBL381163 forms four pairs of hydrogen bonds and seven pairs of π-related interactions with the MMP9 residue acceptor (**Tables 2** and **3**).

Molecular dynamics simulation

Molecular dynamics simulation is a method for simulating the physical motion trajectories and states of atoms and molecules based on Newtonian mechanics. We build a molecular dynamics simulation module to evaluate the stability of small molecule-protein complexes under natural environment conditions. **Figure 6C**, **6D** shows the potential energy and RMSD plots for each complex. The trajectories of each complex reached equilibrium, and the potential energy and RMSD of complexes CEMBL82047-MMP9 and CEMBL381163-MMP9 reached a steady state over time. This indicates that the complexes can exist stably in the natural environment.

Discussion

Tumors are among the leading causes of death worldwide [2], and MMP9 is a reported cancer biomarker [6] that promotes tumor invasion and metastasis, greatly contributing to the occurrence and development of tumors [5, 6]. Although great progress has been made in the design and development of drugs targeting MMP9, these drugs have many shortcomings. This study systematically assessed the expression pattern and prognostic value of MMP9 in pan-cancer and screened for specific MMP9-targeting drugs.

We found that the expression level of MMP9 differed significantly between tumor samples and normal samples in most of the 26 cancers investigated. Higher MMP9 expression was associated with poorer survival and tumor progression in GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC. These findings are consistent with those of previous reports; for example, elevated MMP9 expression in breast cancer has been identified as a predictor of shortened patient survival [31-33]; it also acts as a prognostic biomarker for thyroid cancer [34]. To further confirm the correlation between MMP9 expression and tumors, we performed immune

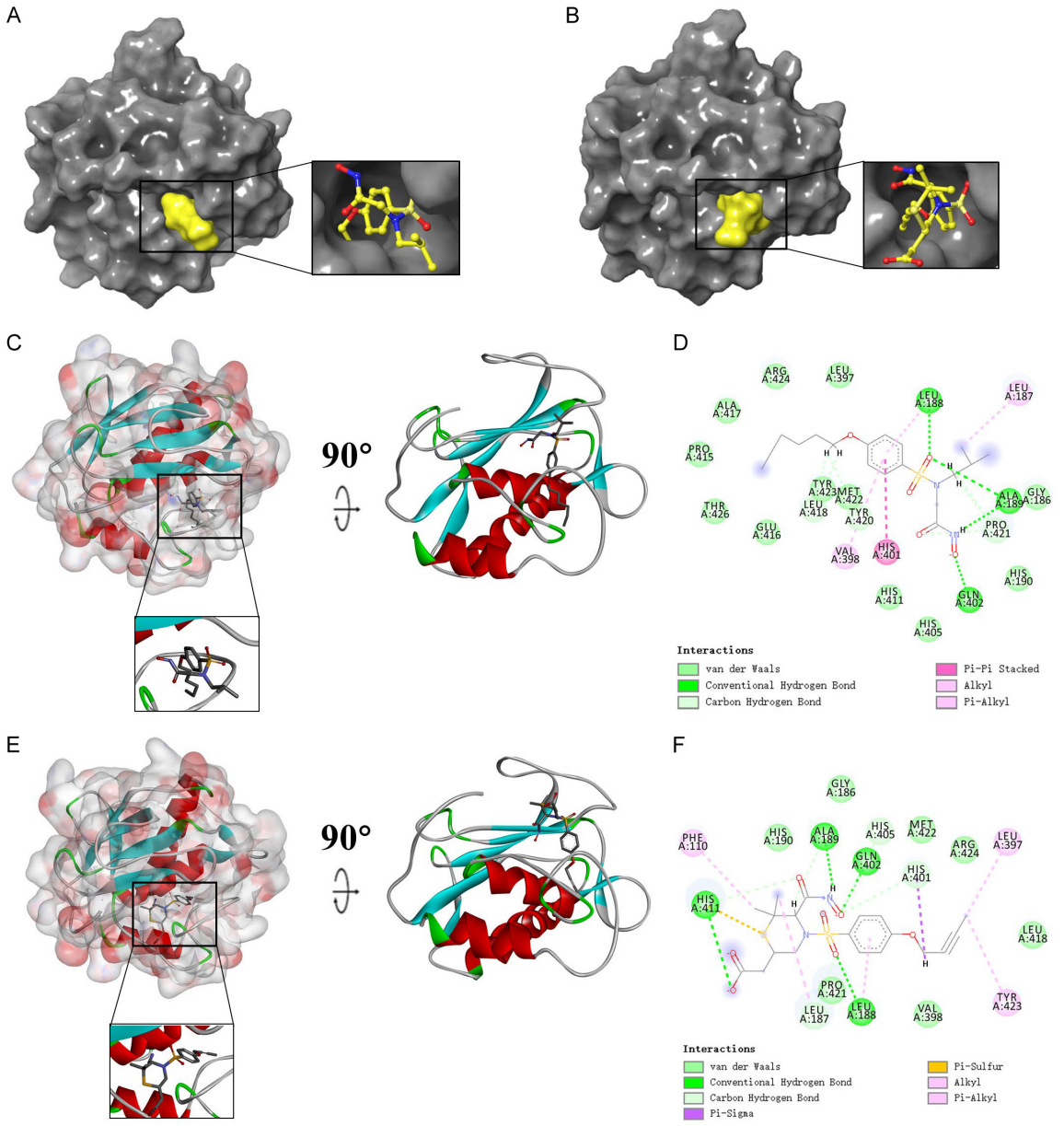


Figure 5. Schematic drawing of interactions between ligands and MMP9. (A) Ligand interaction diagram of CHEMBL82047-MMP9 complex. (B) Ligand interaction diagram of CHEMBL381163-MMP9 complex. (C) CHEMBL82047-MMP9 complex. (D) Schematic of intermolecular interaction of the predicted binding modes of CHEMBL82047 with MMP9. (E) CHEMBL381163-MMP9 complex. (F) Schematic of intermolecular interaction of the predicted binding modes of CHEMBL381163 with MMP9.

infiltration analysis on six abovementioned tumors with high MMP9 expression in TCGA. We found that MMP9 expression in patients with tumors was significantly correlated with the stromal score, immune score, and ESTIMATE score. We also examined the relationship between MMP9 expression and the infiltration of 22 immune cell subtypes, and our findings showed that the level of immune cell infiltration was significantly correlated with MMP9 expression in most cancer types. This also demonstrates that immune escape occurs in patients with tumors with high MMP9 expression; moreover, it illustrates the mechanism of MMP9 in tumors. For example, macrophages were significantly positively associated with all six tumors, and high macrophage expression promotes cancer initiation and malignant progres-

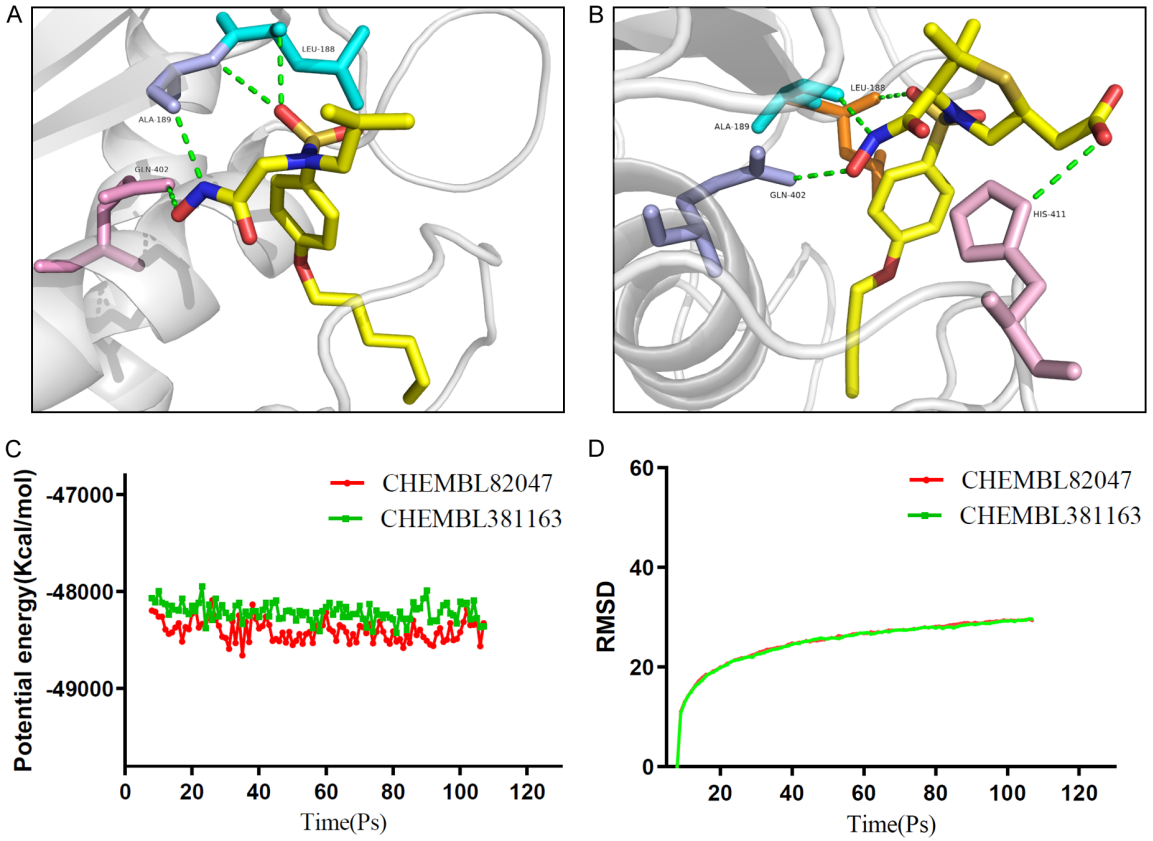


Figure 6. Schematic of intermolecular interaction of the predicted binding modes of (A) CHEMBL82047 with MMP9, and (B) CHEMBL381163 with MMP9. Results of molecular dynamics simulation of the compounds CHEMBL82047-MMP9 complex and CHEMBL381163-MMP9 complex. (C) Potential energy, average backbone root-mean-square deviation. (D) RMSD, root-mean-square deviation.

Table 1. COCKER potential energy of compounds

	COCKER potential energy
CEMBL82047	-12.164
CEMBL381163	-11.623
JNJ0966	-6.629
MMP9IN1	-8.618

sion. During tumorigenesis, macrophages create a mutagenic and growth-promoting inflammatory environment; as tumors progress to malignant tumors, macrophages stimulate angiogenesis, enhance tumor cell migration and invasion, and suppress antitumor immunity [34]. Monocytes were negatively associated with six tumors, and their ability to generate antitumor effectors and activate antigen-presenting cells was suppressed [24]. NK cells were also significantly inhibited in these six tumors, and their ability to directly kill tumor cells and release soluble factors affecting innate and adaptive immune responses was significantly inhibited. In the TME, Tregs can be induced and differentiated by traditional T cells; they have strong immunosuppressive functions, inhibit anti-tumor immunity, and promote the occurrence and development of tumors, which also explains why Treg levels are positively correlated with these tumor types [35]. Activated CD4 memory T cells can suppress anticancer immunity, thereby hindering protective immune surveillance of tumors and hindering effective antitumor immune responses of tumor hosts, promoting tumor development and progression. This finding is consistent with the results of a previous study, in which activated CD4 memory T cell expression was positively correlated with tumors [36].

Exploring the mutational landscape of MMP9 in different cancers further, we found that UVM, KIRC, ACC, and LGG - four types of tumors with high MMP9 expression - had much higher mutation numbers and more mutation types

Table 2. Hydrogen bond interaction parameters for each compound with MMP9 residues

Receptor	Compound	Donor Atom	Receptor Atom	Distances (Å)
2OW1	CEMBL82047	LEU188:H	UNK900:O2	1.84939
		ALA189:H	UNK900:O2	2.5637
		GLN402:HE22	UNK900:O5	2.04371
		UNK900:H29	ALA189:O	2.21685
	CEMBL381163	LEU188:H	UNK900:O3	2.73564
		GLN402:HE22	UNK900:O7	1.84541
		HIS411:HD1	UNK900:O4	2.63353
		UNK900:H22	ALA189:O	2.04562
	JNJ0966	LEU188:H	UNK900:N3	2.7175
		UNK900:H1	MET422:O	2.01943
	MMP9IN1	GLN227:HE21	UNK900:N2	2.6066
		UNK900:H11	TYR245:O	1.90625

Table 3. π-related interaction parameters for each compound with MMP9

Receptor	Compound	Donor Atom	Receptor Atom	Distances (Å)
2OW1	CEMBL82047	HIS401	UNK900	4.2886
		UNK900:C15	LEU187	4.82839
		UNK900	LEU188	5.32252
		UNK900	VAL398	4.8719
	CEMBL381163	UNK900:H4	HIS401	2.77631
		UNK900:C1	LEU397	4.16409
		UNK900:C14	LEU187	4.4138
		PHE110	UNK900:C15	4.14676
		HIS411	UNK900	5.04727
		TYR423	UNK900:C1	4.81758
		UNK900	LEU188	4.80899
	JNJ0966	HIS401	UNK900	4.00463
		UNK900	TYR423	5.5835
		ALA189	UNK900:C9	3.89591
		UNK900:C9	LEU188	4.59029
		UNK900:C9	VAL398	4.42465
		UNK900	LEU188	4.21123
		UNK900	VAL398	4.71999
		UNK900	LEU397	5.13481
		UNK900	LEU418	5.36355
	MMP9IN1	UNK900	LEU243	5.32463

than normal tissues. This also verified that MMP9 promotes tumorigenesis and development. TMB reflects the number of cancer mutations, and a higher TMB generally indicates better outcomes. Mutations are processed as neoantigens and presented to T cells by major histocompatibility complex (MHC) proteins, and a higher TMB results in more neoantigens, increasing the chances of T-cell recognition and improving immunotherapy efficacy [37].

Although MMP9 is highly expressed in most tumors and closely related to tumor metastasis, only a few drugs specifically target MMP9, and they have many limitations. JNJ0966 is a specific inhibitor of MMP9; it is reportedly involved in the progression and development of various diseases, and it can regulate a series of physiological response processes in the body by regulating the expression of MMP9. However, as mentioned above, JNJ0966 is currently only

used in scientific research [12]. Similarly, MMP9-IN-1, as a specific MMP9-targetingdrug, has not been put into clinical use on a large scale because of several defects, such as respiratory system inhibition [7, 11]. Although the mechanism of action of MMP9 in tumor progression is relatively clear, the application of existing drugs is not satisfactory. Therefore, it is necessary to use various cell biology experiments and other methods to screen for and develop new drugs targeting MMP9.

We virtually screened 1,752,844 small-molecule compounds in a natural source compound and synthetic drug database. By constructing a pharmacophore model, we screened 230 small-molecule compounds with the same pharmacophore and then constructed a pharmacophore model. We used a machine learning model to further screen 49 small molecule compounds with high binding affinity to MMP9 and pooled them for further study.

The ADME and toxicity prediction results indicated that CEMBL82047 and CEMBL381163 had good water solubility, absorption levels, and plasma protein binding properties, with no hepatotoxicity or toxicity and low Ames mutagenicity, rodent carcinogenicity, and developmental toxicity, indicating their potential as ideal compounds. Then, we further performed docking analysis and the results showed that CEMBL82047 and CEMBL381163 had higher binding affinity to MMP9 than JNJ0966 and MMP9-IN-1. Because these two compounds form more chemical bonds with MMP9 than JNJ0966 and MMP9-IN-1, they have a higher interaction force and more stable binding with MMP9, which may enhance their inhibition of MMP9, thereby improving the tumor-killing effect. Finally, we conducted a molecular dynamics simulation, and the results showed that the potential energy and RMSD of these complexes reached a steady state over time, indicating that the two complexes remain stable in natural environments.

In conclusion, MMP9 is highly expressed in most cancers. Higher MMP9 expression in GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC is associated with poorer survival and tumor progression. In GMBLGG, KIPAN, UVM, LGG, ACC, and LIHC, higher MMP9 expression is associated with increased infiltration of immune cells, such as macrophages and regulatory T cells, and more RNA modifications. In UVM, LGG, ACC, and LIHC, higher MMP9 expression indicates that the tumor has a higher TMB. A total of 49 candidate inhibitors against MMP9 were screened with a ligand-based pharmacophore model and a machine learning model. CHEMBL82047 and CHEMBL381163 have good water solubility, absorption levels, and plasma protein binding properties. They also have low Ames mutagenicity, rodent carcinogenicity, and developmental toxicity potential, with no hepatotoxicity or toxicity. These molecules have a high binding affinity to proteins and are stable in the natural environment. Therefore, CEMBL82047 and CEMBL381163 show potential as MMP9-inhibiting drugs.

Acknowledgements

This work was supported by Natural Science Foundation of Shaanxi province (2022JQ-299), and Research Program of Xi'an Heath Commission (2021yb26).

Disclosure of conflict of interest

None.

Address correspondence to: Drs. Ming Li and Bo Wu, Lower Extremity Division, Orthopedic Trauma Department, Honghui Hospital, Xi'an Jiaotong University, Youyi East Road No. 555, Beilin District, Xi'an, Shaanxi, China. E-mail: limingguke123@163.com (ML); bocai527@163.com (BW)

References

[1] Hossain SMM, Khatun L, Ray S and Mukhopadhyay A. Pan-cancer classification by regularized multi-task learning. Sci Rep 2021; 11: 24252.

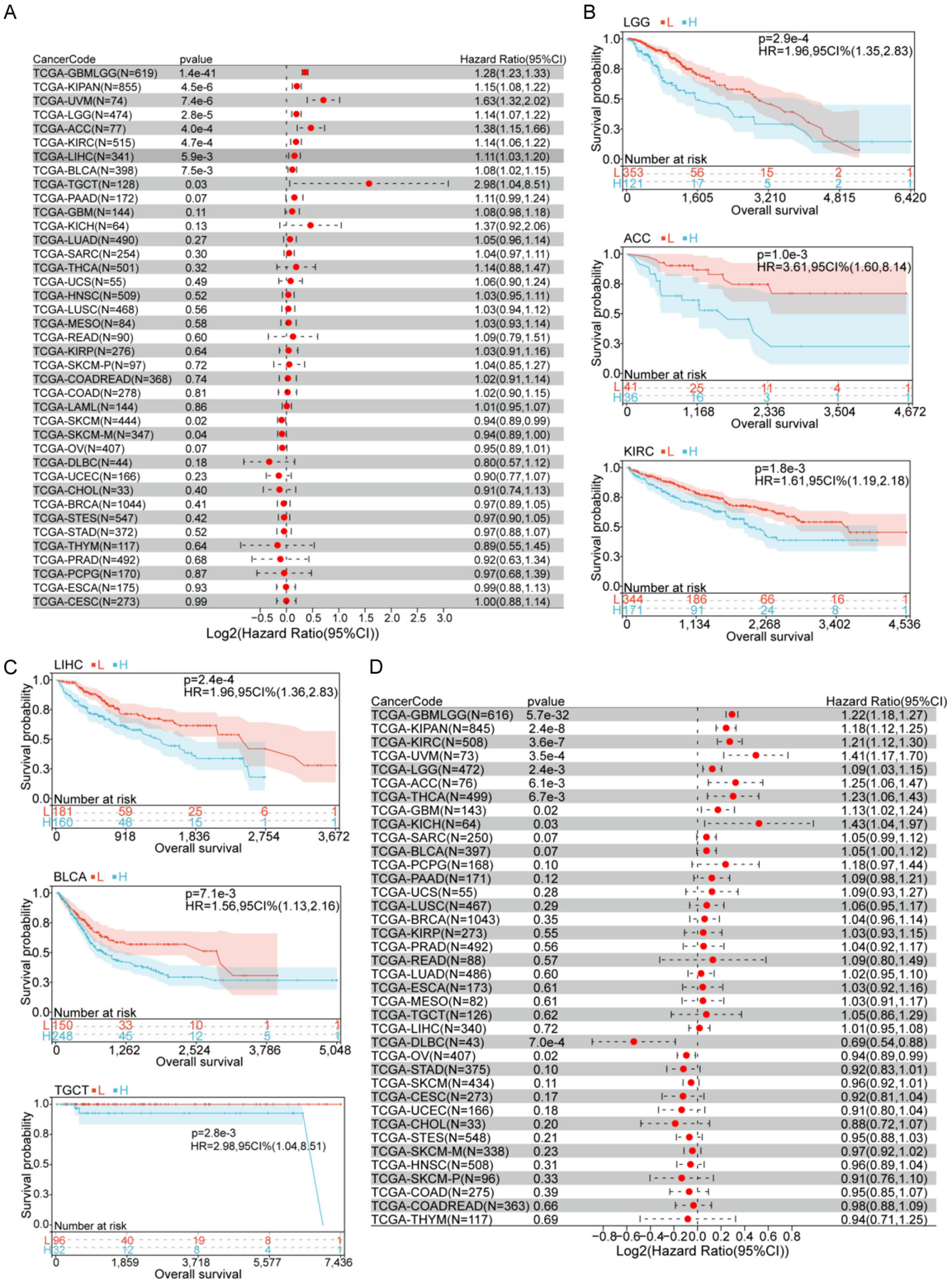
[2] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015; 65: 87-108.

[3] Hu J, Xu J, Feng X, Li Y, Hua F and Xu G. Differential expression of the TLR4 gene in pan-cancer and its related mechanism. Front Cell Dev Biol 2021; 9: 700661.

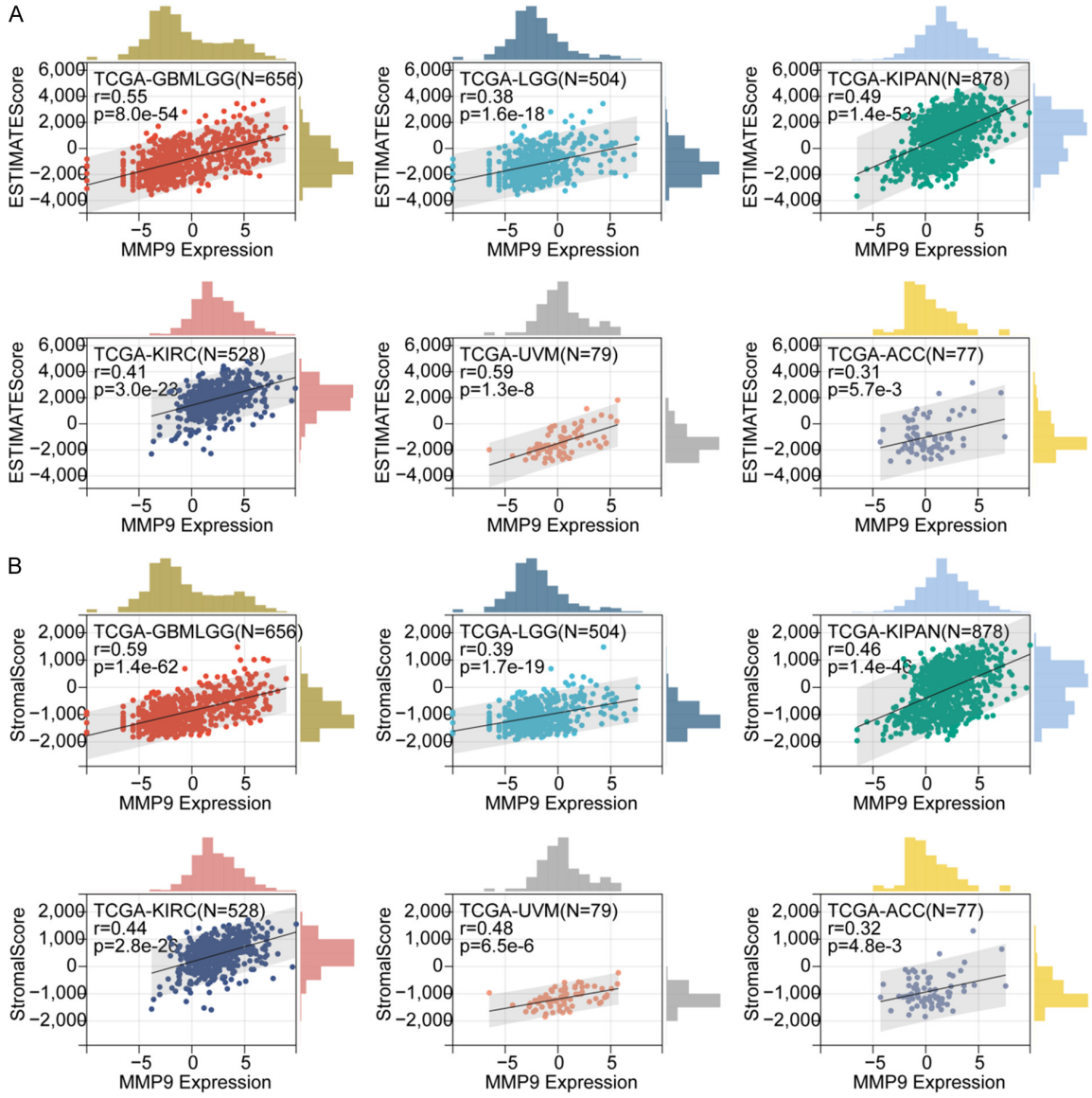
[4] Roche L, Danieli C, Belot A, Grosclaude P, Bouvier AM, Velten M, Iwaz J, Remontet L and Bossard N. Cancer net survival on registry data: use of the new unbiased pohar-perme estimator and magnitude of the bias with the classical methods. Int J Cancer 2013; 132: 2359-2369.

[5] Liu N, Wang X, Wu H, Lv X, Xie H, Guo Z, Wang J, Dou G, Zhang C and Sun M. Computational study of effective matrix metalloproteinase 9 (MMP9) targeting natural inhibitors. Aging (Albany NY) 2021; 13: 22867-22882.
[6] Huang H. Matrix metalloproteinase-9 (MMP-9) as a cancer biomarker and MMP-9 biosensors: recent advances. Sensors (Basel) 2018; 18: 3249.
[7] Liu Z, Li L, Yang Z, Luo W, Li X, Yang H, Yao K, Wu B and Fang W. Increased expression of MMP9 is correlated with poor prognosis of nasopharyngeal carcinoma. BMC Cancer 2010; 10: 270.
[8] Owyong M, Chou J, van den Bijgaart RJ, Kong N, Efe G, Maynard C, Talmi-Frank D, Solomonov I, Koopman C, Hadler-Olsen E, Headley M, Lin C, Wang CY, Sagi I, Werb Z and Plaks V. MMP9 modulates the metastatic cascade and immune landscape for breast cancer anti-metastatic therapy. Life Sci Alliance 2019; 2: e201800226.
[9] Tamura Y, Watanabe F, Nakatani T, Yasui K, Fuji M, Komurasaki T, Tsuzuki H, Maekawa R, Yoshioka T, Kawada K, Sugita K and Ohtani M. Highly selective and orally active inhibitors of type IV collagenase (MMP-9 and MMP-2): N-sulfonylamino acid derivatives. J Med Chem 1998; 41: 640-649.
[10] Dufour A, Sampson NS, Li J, Kuscu C, Rizzo RC, Deleon JL, Zhi J, Jaber N, Liu E, Zucker S and Cao J. Small-molecule anticancer compounds selectively target the hemopexin domain of matrix metalloproteinase-9. Cancer Res 2011; 71: 4977-4988.
[11] Song Z, Wang J, Su Q, Luan M, Chen X and Xu X. The role of MMP-2 and MMP-9 in the metastasis and development of hypopharyngeal carcinoma. Braz J Otorhinolaryngol 2021; 87: 521-528.
[12] Scannevin RH, Alexander R, Haarlander TM, Burke SL, Singer M, Huo C, Zhang YM, Maguire D, Spurlino J, Deckman I, Carroll KI, Lewandowski F, Devine E, Dzordzorme K, Tounge B, Milligan C, Bayoumy S, Williams R, Schalk-Hihi C, Leonard K, Jackson P, Todd M, Kuo LC and Rhodes KJ. Discovery of a highly selective chemical inhibitor of matrix metalloproteinase-9 (MMP-9) that allosterically inhibits zymogen activation. J Biol Chem 2017; 292: 17963-17974.
[13] Ravichandran S, Singh N, Donnelly D, Migliore M, Johnson P, Fishwick C, Luke BT, Martin B, Maudsley S, Fugmann SD and Moaddel R. Pharmacophore model of the quercetin binding site of the SIRT6 protein. J Mol Graph Model 2014; 49: 38-46.
[14] Vaidyanathan J, Vaidyanathan TK and Ravichandran S. Computer simulated screening of dentin bonding primer monomers through analysis of their chemical functions and their spatial 3D alignment. J Biomed Mater Res B Appl Biomater 2009; 88: 447-457.
[15] Pascual R, Almansa C, Plata-Salamán C and Vela JM. A new pharmacophore model for the design of sigma-1 ligands validated on a large experimental dataset. Front Pharmacol 2019; 10: 519.
[16] Lindvall M, McBride C, McKenna M, Gesner TG, Yabannavar A, Wong K, Lin S, Walter A and Shafer CM. 3D pharmacophore model-assisted discovery of novel CDC7 inhibitors. ACS Med Chem Lett 2011; 2: 720-723.
[17] Auslander N, Gussow AB and Koonin EV. Incorporating machine learning into established bioinformatics frameworks. Int J Mol Sci 2021; 22: 2903.
[18] Taniguchi H, Sato H and Shirakawa T. A machine learning model with human cognitive biases capable of learning from small and biased datasets. Sci Rep 2018; 8: 7397.
[19] Deo RC. Machine learning in medicine. Circulation 2015; 132: 1920-1930.
[20] Fu T, Dai LJ, Wu SY, Xiao Y, Ma D, Jiang YZ and Shao ZM. Spatial architecture of the immune microenvironment orchestrates tumor immunity and therapeutic response. J Hematol Oncol 2021; 14: 98.
[21] Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg S, Hedrick CC, Vonderheide RH, Pittet MJ, Jain RK, Zou W, Howcroft TK, Woodhouse EC, Weinberg RA and Krummel MF. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med 2018; 24: 541-550.
[22] Siveen KS and Kuttan G. Role of macrophages in tumour progression. Immunol Lett 2009; 123: 97-102.
[23] Lewis CE and Pollard JW. Distinct role of macrophages in different tumor microenvironments. Cancer Res 2006; 66: 605-612.
[24] Ugel S, Canè S, De Sanctis F and Bronte V. Monocytes in the tumor microenvironment. Annu Rev Pathol 2021; 16: 93-122.
[25] Bazewicz CG, Dinavahi SS, Schell TD and Robertson GP. Aldehyde dehydrogenase in regulatory T-cell development, immunity and cancer. Immunology 2019; 156: 47-55.
[26] Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS and Novoa EM. The RNA modification landscape in human disease. RNA 2017; 23: 1754-1769.
[27] Gao L, Chen R, Sugimoto M, Mizuta M, Kishimoto Y and Omori K. The impact of m1A methylation modification patterns on tumor immune microenvironment and prognosis in oral squamous cell carcinoma. Int J Mol Sci 2021; 22: 10302.

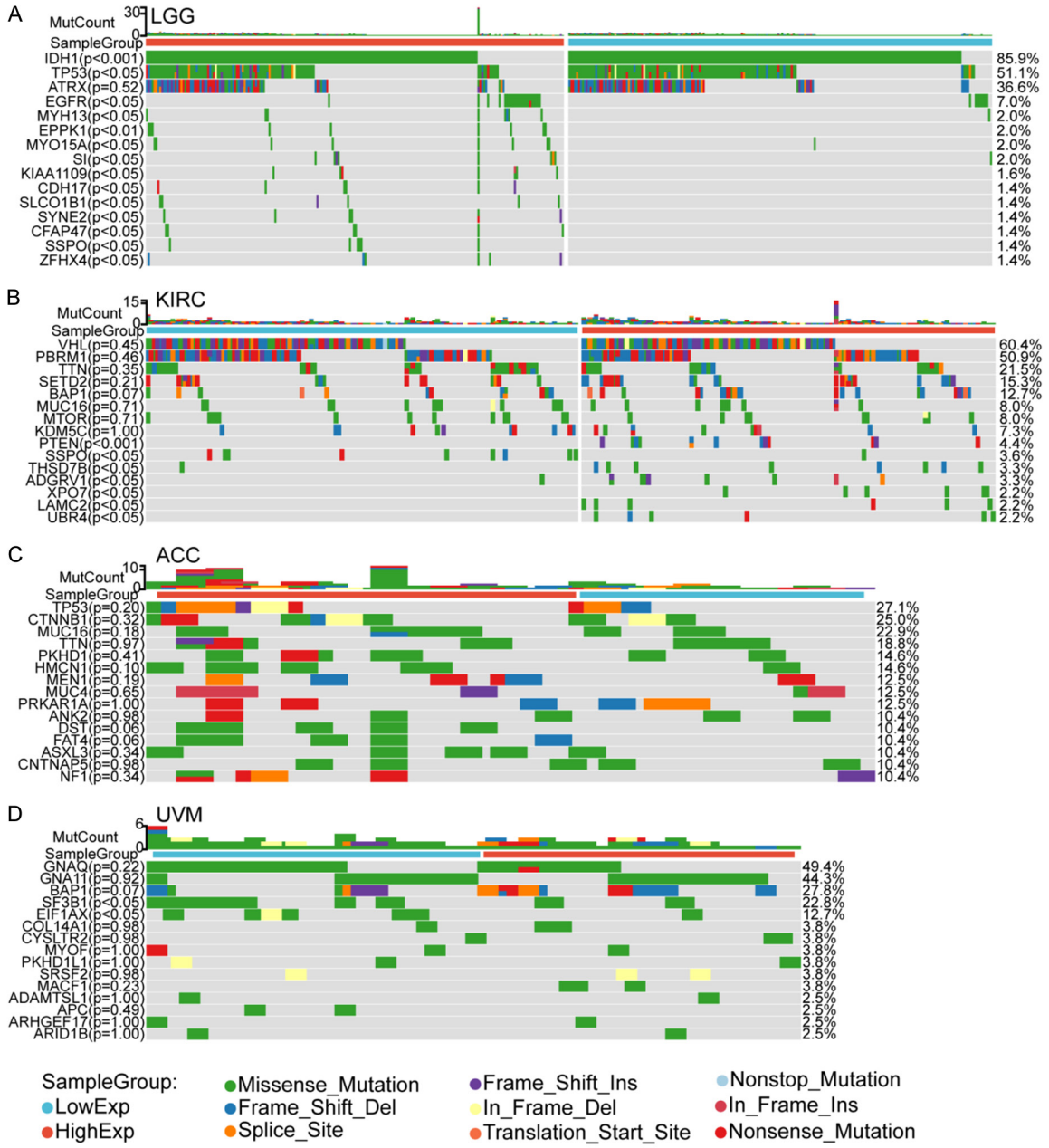
[28] Chen Z, Qi M, Shen B, Luo G, Wu Y, Li J, Lu Z, Zheng Z, Dai Q and Wang H. Transfer RNA demethylase ALKBH3 promotes cancer progression via induction of tRNA-derived small RNAs. Nucleic Acids Res 2019; 47: 2533-2545.
[29] Wang Y, Wang J, Li X, Xiong X, Wang J, Zhou Z, Zhu X, Gu Y, Dominissini D, He L, Tian Y, Yi C and Fan Z. N(1)-methyladenosine methylation in tRNA drives liver tumourigenesis by regulating cholesterol metabolism. Nat Commun 2021; 12: 6314.
[30] Lai SC, Su YT, Chi CC, Kuo YC, Lee KF, Wu YC, Lan PC, Yang MH, Chang TS and Huang YH. DNMT3b/OCT4 expression confers sorafenib resistance and poor prognosis of hepatocellular carcinoma through IL-6/STAT3 regulation. J Exp Clin Cancer Res 2019; 38: 474.
[31] Joseph C, Alsaleem M, Orah N, Narasimha PL, Miligy IM, Kurozumi S, Ellis IO, Mongan NP, Green AR and Rakha EA. Elevated MMP9 expression in breast cancer is a predictor of shorter patient survival. Breast Cancer Res Treat 2020; 182: 267-282.
[32] Xue Q, Cao L, Chen XY, Zhao J, Gao L, Li SZ and Fei Z. High expression of MMP9 in glioma affects cell proliferation and is associated with patient survival rates. Oncol Lett 2017; 13: 1325-1330.
[33] Niu H, Li F, Wang Q, Ye Z, Chen Q and Lin Y. High expression level of MMP9 is associated with poor prognosis in patients with clear cell renal carcinoma. PeerJ 2018; 6: e5050.
[34] Zarkesh M, Zadeh-Vakili A, Akbarzadeh M, Fanaei SA, Hedayati M and Azizi F. The role of matrix metalloproteinase-9 as a prognostic biomarker in papillary thyroid cancer. BMC Cancer 2018; 18: 1199.
[35] Li C, Jiang P, Wei S, Xu X and Wang J. Regulatory T cells in tumor microenvironment: new mechanisms, potential therapeutic strategies and future prospects. Mol Cancer 2020; 19: 116.
[36] Togashi Y, Shitara K and Nishikawa H. Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. Nat Rev Clin Oncol 2019; 16: 356-371.
[37] Jardim DL, Goodman A, de Melo Gagliato D and Kurzrock R. The challenges of tumor mutational burden as an immunotherapy biomarker. Cancer Cell 2021; 39: 154-173.



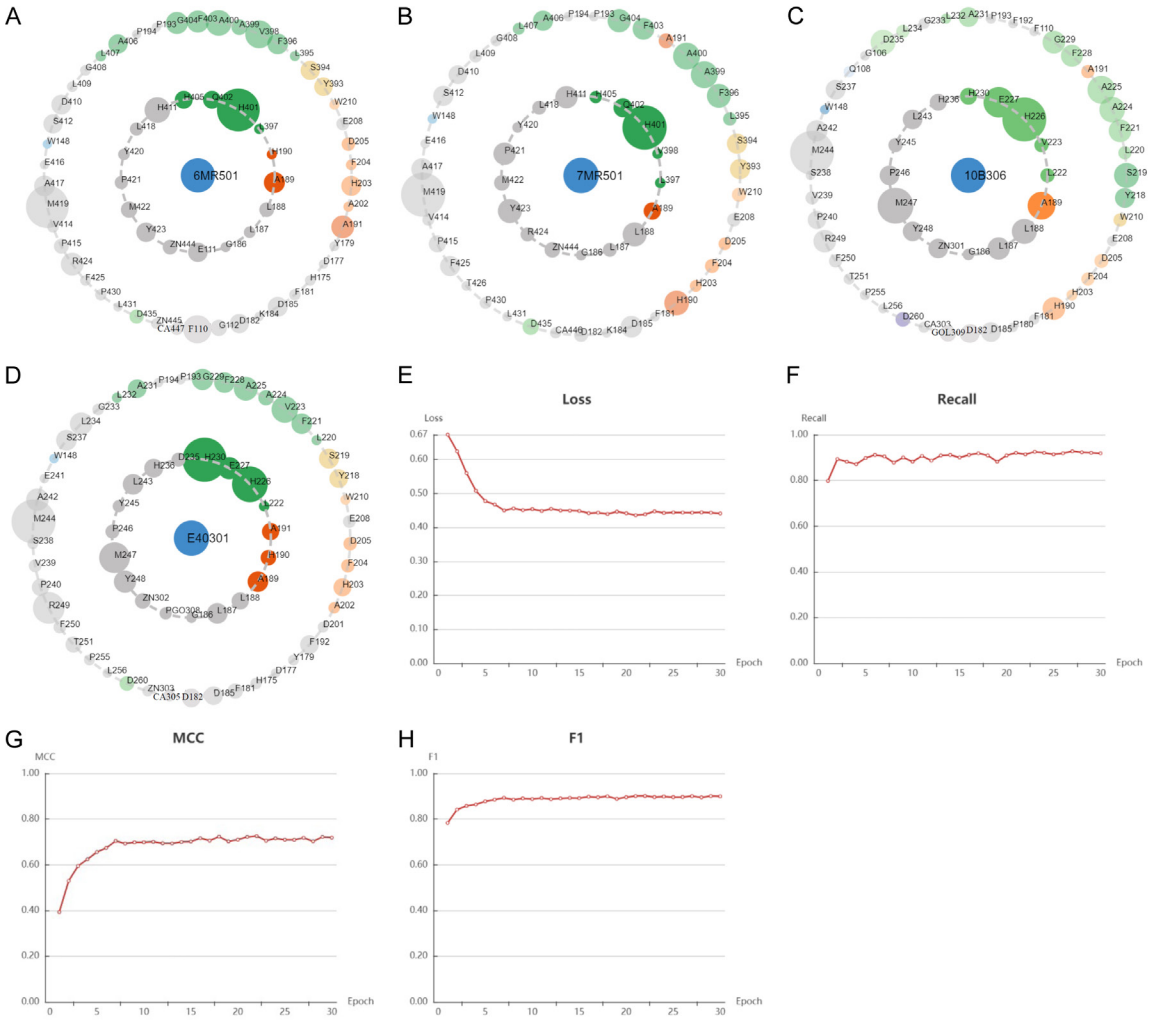
Supplementary Figure 1. (A) MMP9 expression correlates with overall survival time (OS). Forest plots showing the correlations between OS and MMP9 expression across 39 types of cancers. (B, C) Survival curves of MMP9 expression in LGG, ACC, KRIC, LIHC, BLCA and TGCT. L represents low expression of MMP9 group, H represents high expression of MMP9 group. (D) Forest plots showing the correlations between Progression-free survival time (PFS) and MMP9 expression across 39 types of cancers.



Supplementary Figure 2. Pan-cancer cohort (GBMLGG, KICH, KIRC, KIRP, KIPAN and UVM). Correlation between MMP9 expression and pan-cancer (A) estimate score and (B) stroma score.



Supplementary Figure 3. Correlation between MMP9 expression and tumor mutation burden in (A) LGG, (B) KIRC, (C) ACC and (D) UVM.



Supplementary Figure 4. The interaction between MMP9 and MMP9's inhibitors (A) 20W0 (B) 20W1 (C) 4H3X and (D) 4WZV. (E-H) Evaluation index of deep learning model. Loss, Recall, MCC and F1.

Supplementary Table 1. Adsorption, distribution, metabolism, and excretion properties of compounds

	Solubility Level	BBB level	CYP2D6	Hepatotoxicity	Absorption Level	PPB Level
CHEMBL344828 PubChem-10764489	3	3	0	1	0	0
CHEMBL2425940 PubChem-73293197	3	4	0	1	3	1
CHEMBL139884 PubChem-10502046	3	3	0	1	0	1
CHEMBL381554 PubChem-44409390	3	4	0	1	0	1
CHEMBL2425944 PubChem-73293200	4	4	0	1	3	0
CHEMBL82047 PubChem-10738924	3	3	0	0	0	1
CHEMBL196647 PubChem-44402021	4	4	0	1	3	1
CHEMBL381163 PubChem-44409365	3	4	0	0	1	1
CHEMBL206481 PubChem-44409389	3	4	0	1	2	0
CHEMBL207776 PubChem-21304710	3	3	0	1	0	1
CHEMBL138643 PubChem-23523890	2	4	0	1	3	1
CHEMBL382227 PubChem-44411830	2	4	0	1	1	1
CHEMBL419503 PubChem-44325156	4	4	0	1	3	0
CHEMBL252711 PubChem-44445823	4	4	0	1	3	0
CHEMBL433171 PubChem-21130561	2	4	0	0	0	1

CHEMBL1801052 PubChem-9847113	2	4	0	1	1	1
CHEMBL234529 PubChem-25181080	3	4	0	0	3	1
CHEMBL126004 PubChem-10389610	3	4	0	1	0	1
CHEMBL236167 PubChem-23655323	3	3	0	1	0	1
CHEMBL429800 PubChem-23656291	3	3	0	1	0	1
CHEMBL358812 PubChem-10549612	4	4	0	1	2	0
CHEMBL1801395 PubChem-22707860	2	4	0	1	1	1
CHEMBL1916211 PubChem-57403331	2	2	0	1	0	1
CHEMBL1770697 PubChem-20620715	2	4	0	1	2	1
CHEMBL47728 PubChem-44291532	3	3	0	1	0	1
CHEMBL303082 PubChem-44306344	2	4	0	1	3	1
CHEMBL71227 PubChem-44309863	2	4	0	1	1	1
CHEMBL1770712 PubChem-20620688	3	4	0	1	0	1
CHEMBL164980 PubChem-11070343	3	4	0	1	1	0
CHEMBL44045 PubChem-44289352	3	3	0	0	0	1
CHEMBL362797 PubChem-22644895	3	3	0	1	0	1
CHEMBL561625 PubChem-45269631	3	3	0	1	0	1
CHEMBL35606	3	4	0	1	0	1
CHEMBL2425935 PubChem-73292710	3	4	0	1	1	0
CHEMBL2204827 PubChem-71459505	3	3	0	1	0	1
CHEMBL369302 PubChem-22644965	3	4	0	1	0	1
CHEMBL1771223 PubChem-54587429	2	4	0	1	3	1
CHEMBL92778 PubChem-9913479	2	4	0	1	1	1
CHEMBL292671 PubChem-44299758	3	4	0	1	3	1
CHEMBL1771216 PubChem-20620240	3	4	0	0	3	1
CHEMBL1801431 PubChem-10280852 PubChem-46939559	2	4	0	1	2	1
CHEMBL381505 PubChem-44409164	3	4	0	1	2	1
CHEMBL1771222 PubChem-54580544	2	4	0	1	2	1
CHEMBL1771215 PubChem-10483139	2	4	0	1	3	1
CHEMBL1771221 PubChem-54583511	2	4	0	1	2	1
CHEMBL42771 PubChem-44289604	3	3	0	1	0	1
CHEMBL1801398 PubChem-46938727	2	4	0	1	0	1
CHEMBL44250 PubChem-10572544	3	3	0	0	0	1
CHEMBL338007 PubChem-10789711	4	4	0	1	1	0

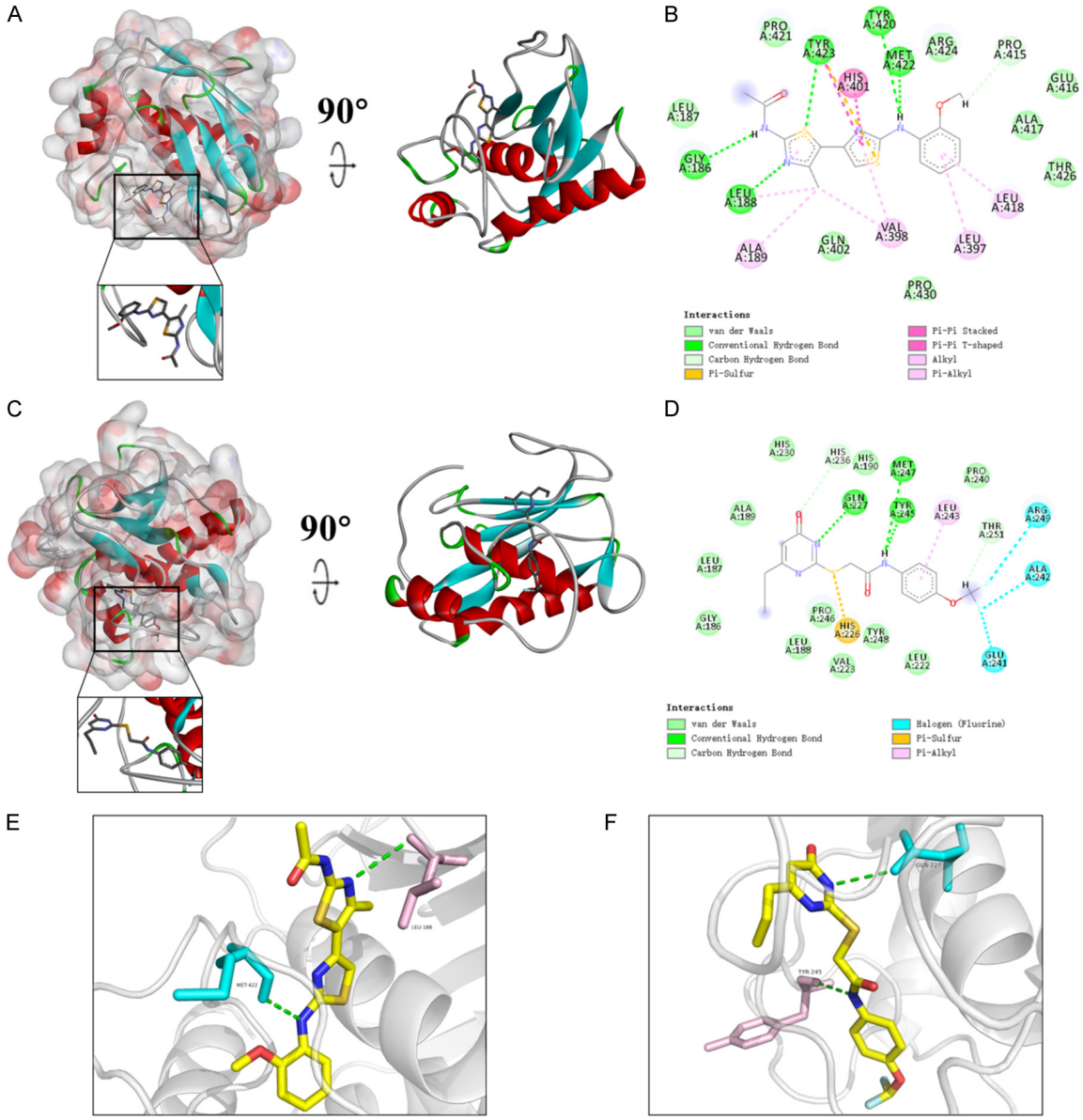
Aqueous-solubility level: 0 (extremely low); 1 (very low, but possible); 2 (low); 3 (good). Blood brain barrier level: 0 (Very high penetrant); 1 (High); 2 (Medium); 3 (Low); 4 (Undefined). Cytochrome P450 2D6 level: 0 (Non-inhibitor); 1 (Inhibitor). Hepatotoxicity: 0 (Nontoxic); 1 (Toxic). Human-intestinal absorption level: 0 (good); 1 (moderate); 2 (poor); 3 (very poor). Plasma protein binding: 0 (Absorbent weak); 1 (Absorbent strong).

Supplementary Table 2. Toxicities of compounds

	Mouse NTP		Rat NTP		AMES	DTP
	Female	Male	Female	Male		
CHEMBL344828 PubChem-10764489	0	0.68	0.809	0.383	1	0.978
CHEMBL2425940 PubChem-73293197	0	0	1	1	1	0
CHEMBL139884 PubChem-10502046	0	0.016	0	0.701	0.396	0.999
CHEMBL381554 PubChem-44409390	0	0	0	0	0.844	1
CHEMBL2425944 PubChem-73293200	0	0.109	1	1	1	0.001
CHEMBL82047 PubChem-10738924	0	0.964	0.439	1	0.99	0.11
CHEMBL196647 PubChem-44402021	0.59	0	1	1	1	1
CHEMBL381163 PubChem-44409365	0	0	0.062	0.023	0.964	1
CHEMBL206481 PubChem-44409389	0	0	0	0	0.623	1
CHEMBL207776 PubChem-21304710	0	0.003	0	0.898	0.498	0.946

CHEMBL138643 PubChem-23523890	0	1	1	1	1	1
CHEMBL382227 PubChem-44411830	0.032	0.05	1	1	0	1
CHEMBL419503 PubChem-44325156	0	0.004	1	1	0.995	1
CHEMBL252711 PubChem-44445823	0	0	1	1	1	1
CHEMBL433171 PubChem-21130561	0.898	0.005	0.953	1	1	1
CHEMBL1801052 PubChem-9847113	0.076	0.352	1	1	0.424	1
CHEMBL234529 PubChem-25181080	0.985	0.028	1	0	1	0.946
CHEMBL126004 PubChem-10389610	0	0.891	1	1	0.137	0.002
CHEMBL236167 PubChem-23655323	0.003	0.012	1	1	1	0.434
CHEMBL429800 PubChem-23656291	0.003	0.012	1	1	1	0.434
CHEMBL358812 PubChem-10549612	0	0.006	1	1	0.639	0.003
CHEMBL1801395 PubChem-22707860	0	0.002	0.999	1	0.948	1
CHEMBL1916211 PubChem-57403331	0	0	0	1	1	0.98
CHEMBL1770697 PubChem-20620715	0.006	0.821	1	1	0	1
CHEMBL47728 PubChem-44291532	0	0	0	0	0.777	0
CHEMBL303082 PubChem-44306344	1	0.636	1	1	0.999	0.521
CHEMBL71227 PubChem-44309863	0	0.001	0.993	1	0.983	1
CHEMBL1770712 PubChem-20620688	0.018	0.998	1	1	0.92	1
CHEMBL164980 PubChem-11070343	0	1	0	1	1	0.273
CHEMBL44045 PubChem-44289352	0	0.752	0.999	0.999	0	1
CHEMBL362797 PubChem-22644895	0	0.001	0	0	0.071	0.649
CHEMBL561625 PubChem-45269631	0	0.009	0.001	0.984	1	0
CHEMBL35606	0	0.002	0	1	0.968	0
CHEMBL2425935 PubChem-73292710						
CHEMBL2204827 PubChem-71459505	0.025	0.364	0.998	1	1	0.829
CHEMBL369302 PubChem-22644965	1	0.63	1	1	1	0.002
CHEMBL1771223 PubChem-54587429	0	0.227	1	1	0.374	0
CHEMBL92778 PubChem-9913479	1	0.005	1	1	1	1
CHEMBL292671 PubChem-44299758	0.987	1	0.903	1	1	0.002
CHEMBL1771216 PubChem-20620240	0	0.997	1	1	0.689	0
CHEMBL1801431 PubChem-10280852 PubChem-46939559	0	0.001	0.992	1	0.588	1
CHEMBL381505 PubChem-44409164	0	0.17	0	0.712	0.024	0.998
CHEMBL1771222 PubChem-54580544	0	0.776	1	1	1	0.813
CHEMBL1771215 PubChem-10483139	0	0.24	1	1	0.334	0
CHEMBL1771221 PubChem-54583511	0	0.187	1	1	1	0.996
CHEMBL42771 PubChem-44289604	0	0.875	1	0.899	0.001	1
CHEMBL1801398 PubChem-46938727	0	0.024	0.966	1	0	1
CHEMBL44250 PubChem-10572544	0.001	0.611	1	1	0.002	1
CHEMBL338007 PubChem-10789711	0	0.998	1	1	1	1

NTP < 0.3 (Non-Carcinogen); > 0.7 (Carcinogen). AMES < 0.3 (Non-Mutagen); > 0.7 (Mutagen). DTP < 0.3 (Nontoxic); > 0.7 (Toxic).



Supplementary Figure 5. Schematic drawing of interactions between control drugs and MMP9. (A) JNJ0966-MMP9 complex, (B) JNJ0966 with MMP9, (C) MMP-9-IN-1-MMP9 complex, (D) MMP-9-IN-1 with MMP9. Schematic of intermolecular interaction of the predicted binding modes of (E) JNJ0966 with MMP9, and (F) MMP-9-IN-1 with MMP9.