Title
Chemical Dissection of the Cell Cycle for Anticancer Drug Discovery and Target Identification

Author
Lo, Yu-Chen

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Chemical Dissection of the Cell Cycle for Anticancer Drug Discovery and Target Identification

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Biomedical Engineering

by

Yu-Chen Ben Lo

2016

ABSTRACT OF THE DISSERTATION

Chemical Dissection of the Cell Cycle for Anticancer Drug Discovery and Target Identification

by

Yu-Chen Ben Lo

Doctor of Philosophy in Biomedical Engineering

University of California, Los Angeles, 2016

Professor Jorge Torres, Co-Chair

Professor Michael Alan Teitell, Co-Chair

The cell cycle is governed by highly regulated mechanisms that mandate organism's proliferation and survival. On the other hand, malignant human diseases like cancer often arise by uncontrolled cell cycle progression. The cell cycle of mammalian cells can be easily perturbed by chemical compounds (drugs), leading to check point activation, cell cycle arrests and eventual apoptosis (cell death). Consequently, identifying critical cell cycle targets whose bioactivities can be modulated by small molecules present a promising strategy to advance our understanding and treatment of neoplasm disease. Here, we present a new approach to dissect mammalian cell cycles using small molecule cell cycle screen to identify diverse hit compounds with potent anticancer activities. In particular, we identified MI-181, a potent tubulin destabilizing agent that is active against multiple cancer cell lines. Further structure study indicates that it binds to a distinct site close to the colchicines binding pocket. As an alternative, we further applied cell cycle analysis to repurpose anticancer agents from non-cancer drugs and have identified six FDA compounds with novel cytotoxic properties. To

enable large-scale profiling of cell cycle targets, we developed a new computational approach for target identification called “CSNAP” based on network algorithm and consensus statistics. CSNAP analysis of M-phase inhibitors identified three novel mitotic targets not previous associated with mitosis and one novel mitotic inhibitor interacting with the colchicine site of beta tubulin. To improve target predictability and identify scaffold hoppers, we considered ligand 3D conformation in an improved CSNAP3D program. CSNAP3D analysis of M-phase compounds discovered low molecular weight taxol mimetics that bound to the taxane site, stabilized microtubule formation, and demonstrated promising transport properties.

The dissertation of Yu-Chen Ben Lo is approved.

Elliot M Landaw

Tseng Hsieng-Rong

Jorge Torres, Committee Co-Chair

Michael Alan Teitell, Committee Co-Chair

University of California, Los Angeles 2016

Dedicated to my parents Tony and Sophia,
my aunts Gina and Lisa,
my grandparents Peilieh and Chiyoko,
my Ph.D. advisor Jorge,
and all the friends, families and teachers.

Table of Contents

List of Figures

Abbreviations

ACE	Angiotensin-converting enzyme
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
ATR	Ataxia telangiectasia and Rad3-related protein
CCI	Cell cycle index
CDK2	Cyclin-dependent kinase 2
DMSO	Dimethyl sulfoxide
EGFR	Epidermal growth factor receptor
GSK3β	Glycogen synthase kinase 3β
HIVRT	HIV reverse-transcriptase
HMGA	HMG-CoA reductase
HSP90	heat-shock protein 90
HTS	High-throughput screening
MAPK	Mitogen-activated protein kinase
PARP	Poly [ADP-ribose] polymerase
PKC	Protein kinase C
Plk1	Polo-like kinase 1
SAC	Spindle assembly checkpoints
SAR	Structure-activity relationship
TOP2	Topoisomerase 2

Acknowledgments

More acknowledgments go here!

I would like to acknowledge my co-authors and collaborators on the work presented in this dissertation: Dr. Silvia Senese for performing small molecule cell cycle screen in Chapter 2 and her assistance on cell-based assays in Chapter 3 and 5; Dr. Robert Damoiseaux and Bryan France for their assistance on assay development and optimization in Chapter 3; Dr. Huang Yong and Dr. Chien-Ming Li for performing mass spectrometry competition assay in Chapter 4.

The research in Chapter 2 was based on the published article "Senese S., Lo Y.C., Huang D., Zangle T.A., Gholkar A.A., Robert L., Homet B., Ribas A., Summers M.K., Teitell M.A., Damoiseaux R., Torres J.Z. "*Chemical dissection of the cell cycle: probes for cell biology and anti-cancer drug development"* Cell Death & Disease 10/2014."

The research in Chapter 4 was based on the published article "Lo Y.C., Senese S., Li C.M, Hu.Q., Huang Y., Damoiseaux R., Torres J.Z., "*Large-scale chemical similarity networks for target profiling of compounds identified in cell-based chemical screens"* PLoS Computational Biology 03/2015."

Finally, I acknowledge the generous support I received to complete this dissertation including IDRE scholarship awards, graduate research fellowship and teaching assistant fellowship. This work was also made possible by National Science Foundation under Grant Number MCB1243645 awarded to Jorge Torres.

Vita

2006	B.S., Biomedical Engineering, UC Berkeley
2006	Honors in Biomedical Engineering, UC Berkeley
2006	Student Member, Bioengineering Honor Society, UC Berkeley
2007	Research Assistant, Arbor Vita Corp., Sunnyvale, CA
2008	Enzyme Chemist, Biomedix USA Inc., Fremont, CA
2010	M.S., Biomedical Engineering, UCLA
2010	Graduate Researcher, Molecular Pharmacology, UCLA
2011	Graduate Researcher, Chemistry and Biochemistry, UCLA
2011	Teaching Assistant, Chemistry and Biochemistry, UCLA
2011	IDRE Scholarship Awards, UCLA
2012	Graduate Field Mentor, Bioengineering, UCLA
2013	HHMI Awards Nominee, UCLA

Publications and Presentations

Lo Y.C., Senese S., France B., Damoiseaux R., Torres J.Z., "*Computational flow cytometry for drug repositioning and the discovery of new anticancer drugs*" (manuscript submitted).

Lo Y.C., Senese S., Damoiseaux R., Torres J.Z., "*CSNAP3D: 3D chemical similarity networks for structure-based drug target prediction and identification of scaffold hoppers*" (manuscript submitted).

Senese S., **Lo Y.C.** (co-author), Gholkar A.A., Li C.M., Huang Y., Mottahedeh J., Kornblum H.I., Damoiseaux R., Torres J.Z., "*Microtubins: a novel class of small synthetic microtubule targeting drugs that inhibit cancer cell proliferation*" (manuscript submitted).

Gholkar A.A, Senese S**., Lo Y.C.**, Capri J., Deardorff W.J., Dharmarajan H., Contreras E., Hodara E., Julian P.W., Jackson P.K., Torres J.Z., "*Tctex1d2 associates with short-rib polydactyly syndrome proteins and is required for ciliogenesis*" **Cell cycle** 10/2015;14(7):1116-25. DOI: 10.4161/15384101.2014.985066.

Lo Y.C., Senese S., Li C.M, Hu.Q., Huang Y., Damoiseaux R., Torres J.Z., "*Large-scale chemical similarity networks for target profiling of compounds identified in cell-based chemical screens"* **PLoS Computational Biology** 03/2015; 11(3):e1004153. DOI:10.1371/journal.pcbi.1004153 (**featured paper**).

Senese S., **Lo Y.C.**, Huang D., Zangle T.A., Gholkar A.A., Robert L., Homet B., Ribas A., Summers M.K., Teitell M.A., Damoiseaux R., Torres J.Z." *Chemical dissection of the cell cycle: probes for cell biology and anti-cancer drug development"* **Cell Death & Disease** 10/2014; 5:e1462. DOI:10.1038/cddis.2014.420.

Senese S., Cheung K., **Lo Y.C.**, Gholkar A.A., Xia X., Wohlschlegel J.A., Torres J.Z.," *A Unique Insertion in STARD9's Motor Domain Regulates Its Stability."* **Molecular Biology of the Cell** 12/2014; 26(3). DOI:10.1091/mbc.E14-03-0829.

Torres J.Z., Summers K.M., Peterson D., Brauer M.J., Lee J., Senese S., Gholkar A.A., **Lo Y.C.**, Lei X., Jung K., Anderson D.C., Davis D.P., Belmont L., Jackson P.K., "*The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly"* **Cell** 12/2011; 147(6):1309-23. DOI:10.1016/j.cell.2011.11.020.

CHAPTER 1

Introduction

1.1 Cell cycle and cancers

The cell cycle in a mammalian cell is a set of highly coordinated events that can be divided into four major phases including G1 (growth phase), S (DNA synthesis), G2 (growth phase) and M (mitosis) [1, 2]. G1 phase is the first growth phase in which the cell's transcription and translation activity increases to produce the required cellular proteins for DNA synthesis. DNA damage during this phase often activates the G1 checkpoint and cells arrest through an ATM/ATR dependent pathway [3]. During S-phase, double stranded DNA is unwound and duplicated. In G2 phase, cells grow in size in preparation for cell division. Damaged DNA at this stage will activate the G2/M checkpoint, which arrests cells for DNA repair using a mechanism similar to the G1/S transition [4]. M phase marks the start of a cells mitosis where DNA is condensed and duplicated DNA is segregated into two daughter cells. In metaphase, misalignment of the chromosomes also triggers spindle checkpoints by activating spindle-assembly checkpoint (SAC) response [5]. In cancer cells, the cell cycles are often deregulated, leading to uncontrolled cell division and proliferation [6, 7]. Since cell cycle progression from one cell phase to another is gated by several checkpoints, one potential strategy for cancer treatment is to use small molecules to disrupt the critical cellular machinery required for cell division (drug target), leading to checkpoint activation and apoptosis (cell death) [8]. For example, inhibitors like 5-fluorouracil that target DNA replication and taxol that binds and stabilizes microtubules, have been used successfully for treating a broad array of cancers including breast and colorectal cancers [9, 10]. Still, these drugs have low efficacy and high dose-limiting toxicities, which limit their potential for the treatment of neoplasm[11]. Consequently, identification of novel cell cycle modulators and targets will be critical to improve the current state of cancer treatments.

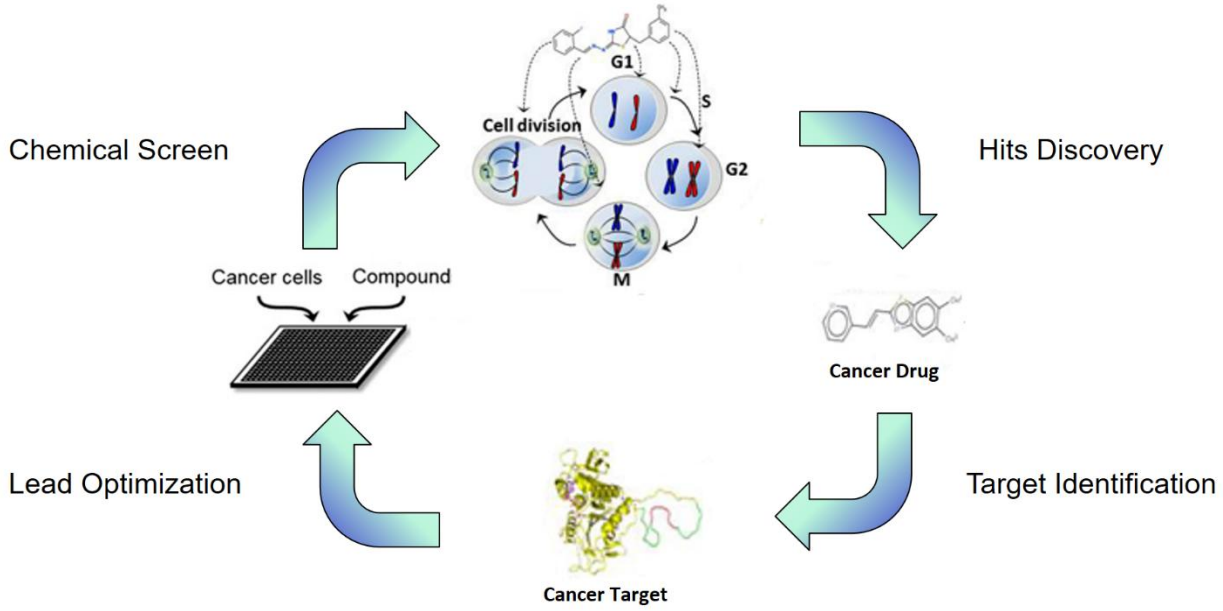

Figure 1.1 – **Overview of chemical dissection of the cell cycle.** A chemical library consisting of diverse small synthetic molecules was used to screen for cell cycle modulators that cause mitotic arrest and apoptosis. The drug targets of the discovered hits were deconvoluted to determine the underlying anticancer mechanisms and pathways. The structure information of the drug target can be further applied in medicinal chemistry efforts to improve pharmacological and safety profiles of the lead compounds in additional rounds of chemical screening.

1.2 Cell cycle drug discovery

With ever increasing understanding of human diseases at the genetic and proteomic level and the advance in information technology and robotics automation, large-scale screens of drug-like molecules are often the predominant method of modern drug discovery programs [12]. In a high throughput screen (HTS) campaign, chemical libraries consisting of diverse chemical structures are often prepared and the bioactivity of the drug targets are tested in micro plates assisted by robotic arms [12, 13]. This approach can be used to screen millions of compounds within days to identify potential hits. To further improve the pharmacological profiles of the hit compounds, medicinal chemistry efforts are often involved to convert hits into leads through chemical modification [14]. For example, an *in vitro* chemical screen targeting the Plk1 kinase identified the

small molecule BI2536 [15]. BI2536 was not only used to define novel roles for Plk1 during cell division, it was further developed into an anti-cancer drug whose efficacy is being evaluated in clinical trials [15]. Alternatively, computational virtual screens based on knowledge of receptor structure can be applied [16]. This approach involves *in-silico* docking of the small molecule to the receptor site points to identify the binding ligands with optimal binding energy [17]. The advantage of virtual screening over high-throughput screening is the possibility of screening a large chemical space including compounds not readily available. Computational screens have also been successfully applied to identify several potent inhibitors of Plk1 with distinct chemical structures [18].

However, for cell cycle drug discovery, the aforementioned approaches often necessitate a defined drug target. While genome wide studies depleting the expression of human genes have provided a wealth of potential drug targets required for proliferation, there is no guarantee that the identified targets are druggable. Furthermore, there are no inhibitors to the majority of the cell cycle machinery. As an alternative, unbiased cell-based or phenotypic chemical screens can be used to identify inhibitors that target multiple phases of the cell cycle without regarding to underlying molecular mechanisms (Figure 1.1)[19]. Although cell-based chemical screens have been performed for a single cell phase, these screens were conducted with a limited number of compounds (100–38 000) or cell extract fractions, with several screens reusing the identical library, thus limiting compound diversity, chemical coverage, and opportunities for novel discoveries. Most screens also lacked further chemical analyses to understand their cellular targets. Thus, there is a critical need to develop new screening strategies to discover novel anti-cancer drugs.

1.3 Cell cycle target identification

Drug target identification is currently the major bottleneck in modern drug discovery, particularly for hits derived from a cell-based or phenotype-based chemical screens (Figure 1.1)[20]. Classical methods for target identification like chemical proteomics rely on compound modification and immobilization to generate compound affinity matrixes that can be used to pull down associated proteins (Figure 1.2A)[21]. Without prior knowledge of compound structure-activity-relationship (SAR), the modification of key functional groups can occlude compound activity and hamper protein-ligand interactions. Additionally, these approaches are labor intensive, costly and have a low success rate [22]. Furthermore, many hits compounds identified from a chemical screen have suboptimal binding specificity and may have promiscuous binding tendency to multiple drug targets. Thus, differentiating between on and off targets is often necessary.

One viable alternative for experimental target identification is computational target fishing (Figure 1.2B) [23, 24]. Computational approaches for predicting the targets, off-targets and poly-pharmacology of hit compounds have been used widely in recent years due to their speed, flexibility and ability to be easily coupled to experimental validation techniques. *In-silico* target inference methods include ligand-based and structure-based approaches [23]. Ligand-based approaches, such as similarity ensemble approach (SEA), compare hit compounds to a database of annotated compounds, and drug targets of hit compounds are inferred from the targets of the most similar annotated compounds, based on their chemical structure similarity [25]. However, current ligand-based approaches analyze bioactive molecules in an independent sequential fashion, which has several disadvantages. For example, target inference is based on finding a single most similar annotated compound for a given query ligand, which may not provide consistent target prediction for a group of structurally similar ligands. Additionally, subtle structural changes in the functional groups of active molecules can alter their potency and specificity toward drug targets; thus, analyzing each molecule independently may not offer a coherent SAR for a congeneric series. This suggests that a more global and systematic analysis of compound bioactivity is required to improve the current state of *in-silico* drug target prediction [26]. Alternatively, if identifying complementary ligands is not required, proteomics approach can be used [27]. The proteomic

approach often involves large-scale siRNA knockdown of all critical genes and subsequent screening for the desired drug targets based on the effected phenotypes. Although the targets identified in this approach may not be immediately druggable, this approach can nevertheless discover highly specific drug targets that can be followed up by target-based chemical screens.

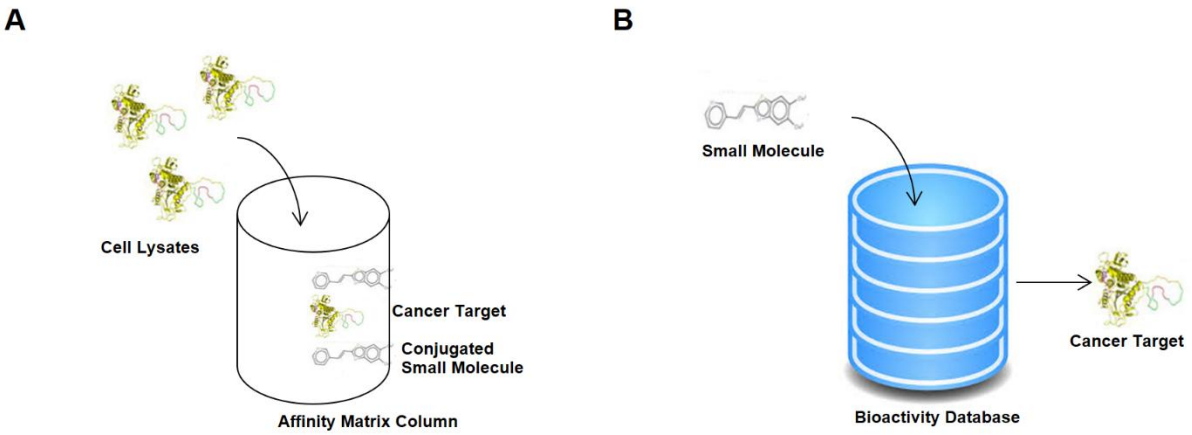

Figure 1.2 – Approaches for drug target identification. (A) Chemical proteomic approach for target identification requires conjugation of small molecule onto the affinity matrix column to pull down the target from the cell lysates. (B) Computational target prediction uses the small molecule as query to search for similar compounds from chemical databases with known targets. The drug targets can be inferred from the annotated ligands sharing the highest chemical similarity.

1.4 Overview of the dissertation

This dissertation will be focusing on the discovery of drugs and targets critical for cell cycle regulation and will have important ramification for their future application in cancer treatment. In chapter 2, we first present a new cell-based chemical screen to identify several novel cell cycle modulators that target different phases of the cell cycle. In particular, we identified MI-181, a potent M-phase inhibitors that showed broad activities in wide array of cancer cell types including melanomas. In chapter 3, we present a new drug target identification strategy by cell cycle profiling. We used this approach to profile 1200 FDA approved drugs and identified 36 approved drugs with novel anticancer indications. Further cell cycle and structural similarity clustering predicted that many of these compounds are tubulin or DNA binding agents. In chapter 4, we proposed a new computational approach called CSNAP for large-scale drug target profiling based on

chemical similarity networks. We showed that CSNAP improved target prediction performance over traditional ligand-based approaches and can further be applied to predict on and off targets of promiscuous ligands. CSNAP analysis of M-phase inhibitors identified three novel mitotic targets and multiple tubulin targeting agents that are stabilizing or destabilizing agents. In chapter 5, we incorporated 3D structure similarity metrics into the CSNAP algorithm for structure-based drug target prediction. We showed that the CSNAP3D program improved the target predictability of HIVRT inhibitors by structural clustering. We subsequently applied CSNAP3D to identify low-weight taxol mimetics that bind to the taxane site and stabilize microtubule formation similar to Paclitaxel.

CHAPTER 2

Cell-based Chemical Screens for Cell Cycle Drug Discovery

2.1 Small molecule cell cycle screen

Cell-based small molecule screens represent a powerful strategy to identify novel cell cycle compounds when potential drug targets remain unknown (Figure 2.1A). The identified modulators can be further deconvoluted to determine their binding targets. The knowledge of drug target in particular regarding how the drug interacts with the binding pockets can be used to initiate additional rounds of chemical synthesis and experimental testing with the goal to optimize drug binding affinity and compound potency while minimizing toxicity.

A cell cycle can be divided into four phases: G1, S, G2/M and subG1 (Figure 2.1B) [2]. As the cell cycle progresses, the DNA content can be effectively quantified by flow cytometry (Figure 2.1C) [28]. Using a fluorescent dye that intercalates into DNA, the major cell population migrates from a low fluorescent intensity region at G1 phase to a high fluorescent intensity region at the end of G2/M phase of the cell cycle [28]. Briefly, human HeLa cancer cells were plated into 384-well plates and a diverse compound library (79 827 small molecules) encompassing broad chemical space was used to place one compound per well in triplicate at 10 μM final concentration. These compounds were pre-selected based on their drug-like properties: predominantly conform to Lipinski's rule of five for acceptable molecular properties for orally active drugs in humans. Twenty hours later, the cells were fixed and stained with the DNA-selective stain Vybrant DyeCycle Green, which is cell membrane permeate and after binding to DNA emits a fluorescent signal that is proportional to DNA mass when excited at

488 nm. Plates were scanned with a fluorescence micro-plate cytometer and a cell cycle histogram profile was generated for each well, which had been treated with one compound. Cell cycle profiles were grouped and ranked according to the extent of G1, S, and G2/M arrest. The top hits from each phase were cherry picked and retested in triplicate and only those testing positive were considered further. In total we uncovered 69 G1-phase inhibitors (>4 S.D. from the mean), 148 S-phase inhibitors (> 5 S.D. from the mean), and 273 G2/M-phase inhibitors (% G2/M ≥ 67%).

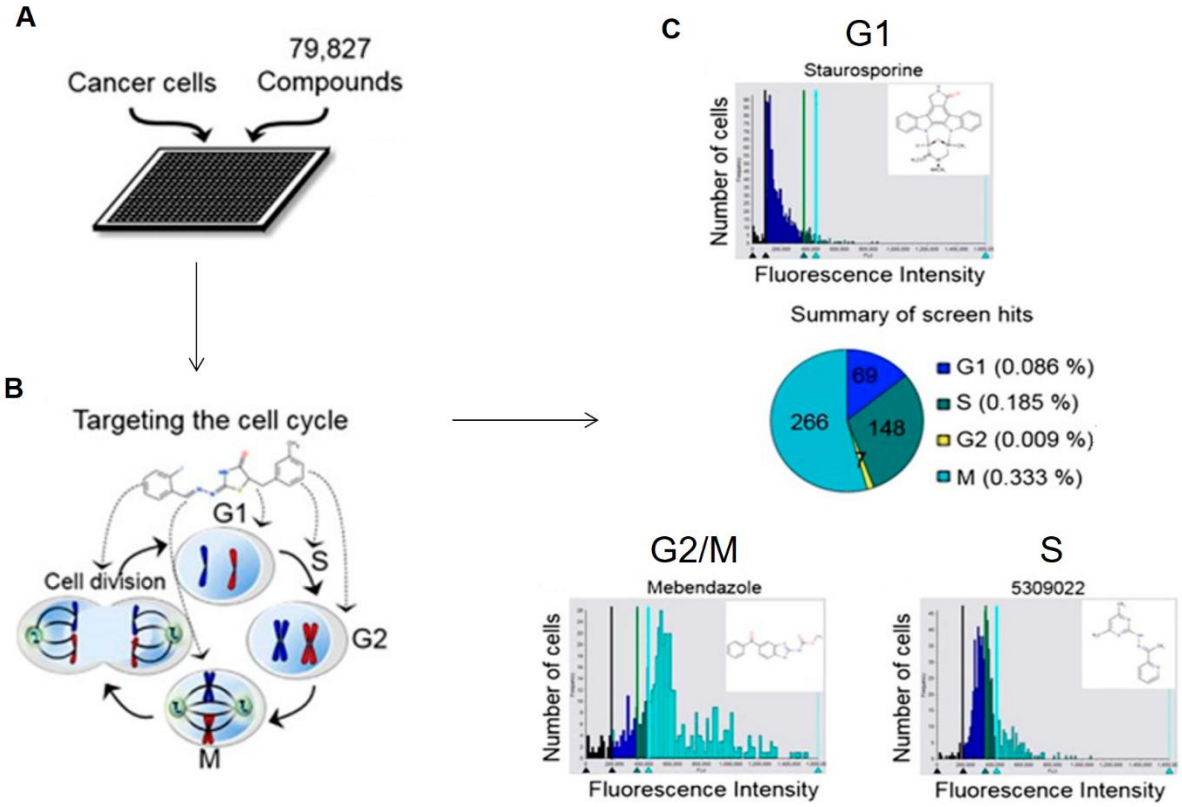

Figure 2.1 – Cell cycle small molecule screen. (A) HeLa cancer cells were plated into 384-well plates and a diverse compound library (79 827 small molecules) encompassing broad chemical space was used to place one compound per well at 10 μM final concentration. (B) The tested compounds target different phases of the cell cycle, leading to cell cycle arrest in G1, S, or G2/M phases. (C) The cells were fixed and stained with the DNA-selective stain Vybrant DyeCycle Green, and after binding the DNA emits a fluorescent signal that is proportional to DNA mass when excited at 488 nm. Plates were scanned with a fluorescence micro-plate cytometer and a cell cycle histogram profile was generated for each well, which had been treated with one compound.

2.2 Discovery of cell cycle modulators

The cell cycle small molecule screen identified diverse compounds capable of inducing cell cycle arrests and many of which are anticancer agents (Figure 2.2). The identified G1-arrest compounds include kinase inhibitors like Staurosporine, Tyrphostin, and their analogs that mimicked the ATP substrate of PKC and EGFR, which were known to block the MAPK signaling pathway critical for tumor proliferation (Figure 2.2A) [29, 30]. In addition, compounds capable of modulating the intracellular calcium concentration including the ion channel inhibitors Thapsigargin (sarco-endoplasmic reticulum Ca^{2+} ATPase inhibitor), Ouabain (Na^{+}/K^{+} ATPase inhibitor), and the ionophore antibiotic A-23187 were also identified (Figure 2.2A) [31]. This was consistent with reports indicating that calcium is an important secondary messenger and that oscillatory calcium signaling is required for MAPK activity and cyclin D1/E synthesis at the G1/S transition[32].

The discovered S-phase compounds included a group of compounds 5309022 and 5113916 that are likely inhibiting ribonucleotide reductase (catalyzes the reduction of ribonucleotides to deoxyribonucleotides; the building blocks for DNA replication and repair) activities by iron chelating through a hydrazone motif similar to that of Triapine and its analog 311 (Figure 2.2B) [33]. In addition, two GSK3β inhibitors (5100772 and 5583777) were identified among the S-phase inhibitors; consistent with its role in regulating cyclin D1 expression required for S-phase entry and progression (Figure 2.2B) [34].

On the other hand, seven G2-phase compounds were identified including DNA topoisomerase II (TOP2) inhibitors such as Etoposide and Amsacrine-like analogs (Figure 2.2C). These DNA intercalating agents trap TOP2 : DNA covalent complexes, which induce DNA damage and G2 checkpoint arrest [35].

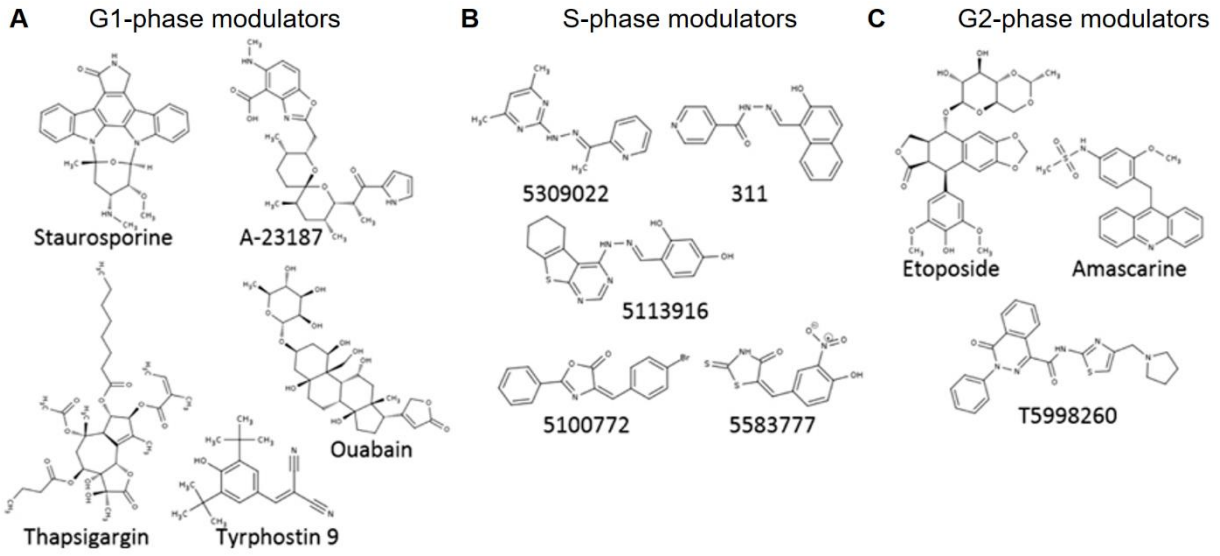

Figure 2.2 – **Discovery of cell cycle modulators.** (A) The identified G1-arrest compounds include kinase inhibitors like Staurosporine, Tyrphostin, and their analogs that mimicked the ATP substrate of PKC and EGFR, which were known to block the MAPK signaling pathway critical for tumor proliferation. In addition, compounds capable of modulating the intracellular calcium concentration including the ion channel inhibitors Thapsigargin (sarco-endoplasmic reticulum Ca^{2+} ATPase inhibitor), Ouabain (Na^{+}/K^{+} ATPase inhibitor), and the ionophore antibiotic A-23187 were also identified. (B) The discovered S-phase compounds include a group of compounds 5309022 and 5113916 that are likely inhibiting ribonucleotide reductase (catalyzes the reduction of ribonucleotides to deoxyribonucleotides; the building blocks for DNA replication and repair) activities by iron chelating through a hydrazone motif similar to that of Triapine and its analog 311. In addition, two GSK3β inhibitors (5100772 and 5583777) were identified among the S-phase inhibitors. (C) The identified G2-phase compounds include DNA topoisomerase II (TOP2) inhibitors such as Etoposide and Amsacrine-like analogs. These DNA intercalating agents trap TOP2: DNA covalent complexes, which induce DNA damage and G2 checkpoint arrest.

2.3 Characterization of M-phase inhibitors

We further performed detailed characterization of M-phase inhibitors to identify novel antimitotics with potent anticancer activities. To assess the potential of these M-phase inhibitors as anticancer agents, we performed mitotic arrest assays and cell viability assay to determine their cytotoxicity in HeLa cell cultures. For mitotic arrest assays, HeLa cells were treated with a 20 2-fold-titration (190pM to 10 μM) of each compound for 20 h and the half maximal inhibitory concentration (IC_{50}) was derived for each compound using the cell cycle profile assay used in the initial screen. Within the titration series, all compounds displayed a percent mitotic arrest between 48.3 and 83.6% at their IC_{50}. For viability assays, cells were treated with a 14 2-fold-titration (12.2 nM to 100 μM) of each compound for 72 h and cell viability was measured using the CellTiter-Glo luminescent cell viability assay. Interestingly, all 211 compounds arrested cells in mitosis and decreased cell viability. Most of these compounds were potent; 16 had a mitotic arrest IC_{50} of $\leq$100 nM, 56 had $\leq$500 nM, and 98 had $\leq$1 μM. Similarly, 13 compounds had a cell viability IC_{50} of $\leq$100 nM, 56 had $\leq$500 nM, and 95 had $\leq$1 μM.

To acquire further understanding of the compound mechanisms, we followed up with multiparametric phenotypic analyses using high-resolution immunofluorescence (IF) microscopy to determine the mitotic defects induced by these mitotic compounds (Figure 2.3A). Briefly, HeLa cells were treated with each of 211 available compounds at a concentration corresponding to their mitotic arrest IC_{90} for 20 h. Cells were then fixed, permeabilized, and co-stained for DNA and α-tubulin. Six major classes of phenotypes were observed: multipolar spindle, normal bipolar spindle with unaligned chromosomes, defective bipolar spindle with unaligned chromosomes, mixed phenotype (containing more than one of the six phenotypes), depolymerized microtubules, and stabilized microtubules (Figure 2.3B). The compound class that induced microtubule depolymerization was the most abundant with 84 compounds falling into this category, followed by the mixed phenotype (54 compounds) and the multipolar phenotype (37 compounds).

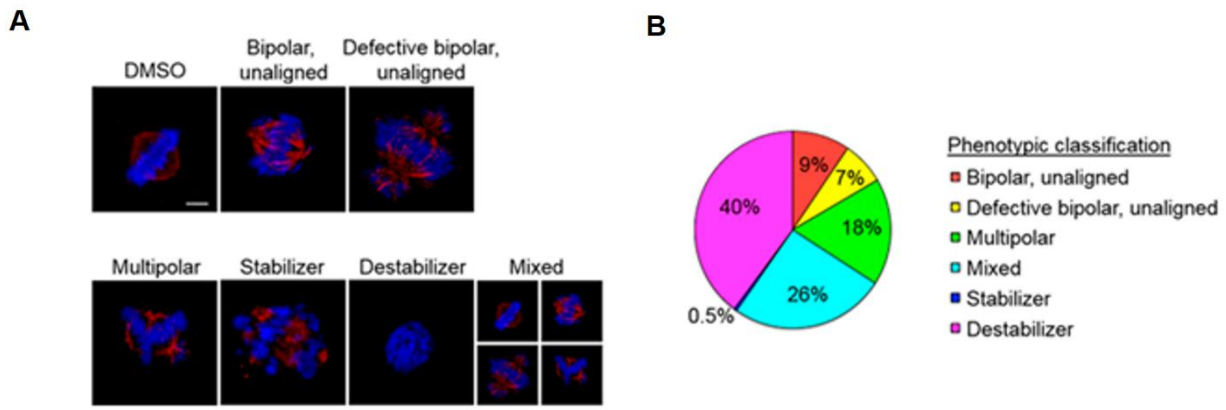

Figure 2.3 – **Multi-parametric analysis of M-phase modulators** (A) Multiparametric phenotypic analysis using high-resolution immunofluorescence (IF) microscopy to determine the mitotic defects induced by mitotic compounds. Briefly, HeLa cells were treated with each of the 211 compounds at a concentration corresponding to their mitotic arrest IC_{90} for 20 h. Cells were then fixed, permeabilized, and co-stained for DNA and α-tubulin. (B) Six major classes of phenotypes were observed: multipolar spindle, normal bipolar spindle with unaligned chromosomes, defective bipolar spindle with unaligned chromosomes, mixed phenotype (containing more than one of the six phenotypes), depolymerized microtubules, and stabilized microtubules. The compound class that induced microtubule depolymerization was the most abundant with 84 compounds falling into this category, followed by the mixed phenotype (54 compounds) and the multipolar phenotype (37 compounds).

2.4 Characterization of a novel antimitotics: MI-181

Based on the mitotic arrests assay, cell viability assay and phenotypic analysis, we identified MI-181 as a potent antimitotic (mitotic arrest IC_{50}=23 nM and cell death IC_{50}= 17 nM), similar to the taxol (mitotic arrest IC_{50}=37 nM and cell death IC_{50}=27 nM) and colchicine (mitotic arrest IC_{50}=24 nM and cell death IC_{50}=12 nM) controls (Figure 2.4A). In addition, MI-181 arrested more cells in mitosis compared with colchicine and taxol, 77% *versus* 76%, and 74%, respectively at their IC_{50}. MI-181 is a small (266Da) synthetic benzothiazole-based compound, and *in silico* analysis of its physiochemical properties indicated that it conformed to parameters needed for oral bioavailability in humans [36]. To evaluate MI-181's mechanism of action, we treated HeLa cells

with MI-181 for 20 h, fixed them, co-stained them for DNA and α-tubulin, and imaged them by IF microscopy. These analyses revealed that MI-181-treated cells failed to form a mitotic spindle and only small tubulin polymers were observed, indicative of microtubule depolymerization. This was further confirmed *in vitro* where MI-181 inhibited tubulin polymerization in an *in vitro* tubulin polymerization assay, similar to colchicines (Figure 2.4B).

Since MI-181 arrests cell in mitosis, we performed further experiment to confirm SAC activation (Figure 2.4C). The status of mitotic arrest can be confirmed by p-H3 phosphorylation. Briefly, HeLa cells were treated with DMSO, MI-181, or colchicine for 18 h, cells were fixed, co-stained for DNA, microtubules, kinetochores, and the mitotic marker p-H3, and imaged by fluorescence microscopy. Similar to colchicine treatment, MI-181 treatment arrested cells in mitosis with p-H3-positive staining and unaligned tightly condensed chromosomes. The SAC activation was confirmed by co-staining for the SAC component Bub1. Indeed, Bub1 remained localized to the kinetochore region in colchicine-and MI-181-treated cells.

To determine if MI-181 had broad anti-cancer activity, we treated a diverse panel of cancer cell lines including cervical adenocarcinoma (HeLa), breast adenocarcinoma (MCF7), melanoma (M233), osteosarcoma (U2OS), acute lymphoblastic leukemia (CCRF-CEM), non-small cell lung carcinoma (NCI-H460), and breast adenocarcinoma (MCF7) with MI-181 and determined its cell viability IC_{50}. Interestingly, MI-181 showed great efficacy across most cancer cell lines with a cell viability IC_{50} ranging from 0.03 to 0.36 μM, with the exception of MCF7 cells (IC_{50}=11 μM). These results indicated that MI-181 was potent across a broad array of cancers and was most effective against cervical adenocarcinoma and melanoma cell lines.

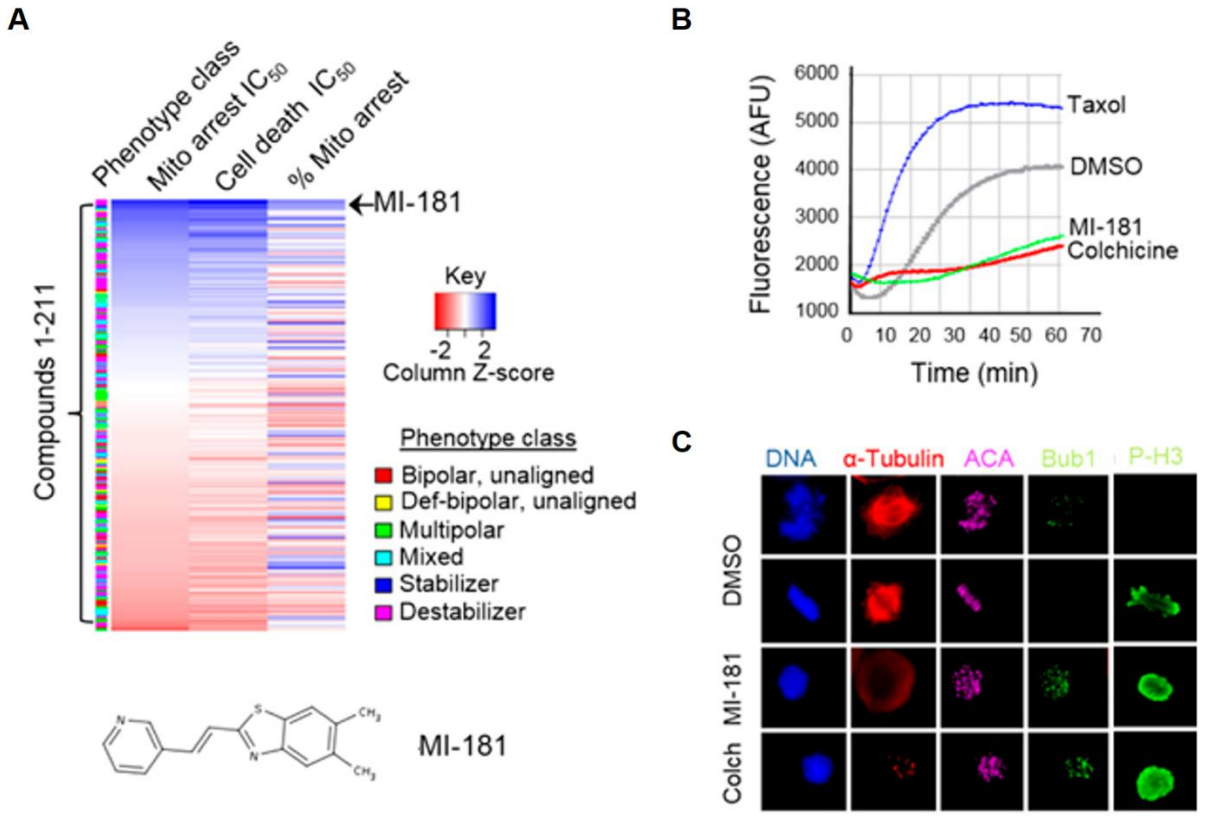

Figure 2.4 – **Evaluation of a novel antimitotic: MI-181** (A) Based on the mitotic arrests assay, cell viability assay and phenotypic analysis, we identified MI-181 as a potent antimitotic (mitotic arrest IC_{50}=23 nM and cell death IC_{50}= 17 nM), similar to the taxol (mitotic arrest IC_{50}=37 nM and cell death IC_{50}=27 nM) and colchicine (mitotic arrest IC_{50}=24 nM and cell death IC_{50}=12 nM) controls. (B) Tubulin polymerization assay showed that MI-181 inhibited tubulin polymerization in an *in vitro* tubulin polymerization assay, similar to colchicine. (C) To confirm SAC activation by MI-181, the status of mitotic arrest were confirmed by p-H3 phosphorylation. Briefly, HeLa cells were treated with DMSO, MI-181, or colchicine for 18 h, cells were fixed, co-stained for DNA, microtubules, kinetochores, and the mitotic marker p-H3, and imaged by fluorescence microscopy. Similar to colchicine treatment, MI-181 treatment arrested cells in mitosis with p-H3-positive staining and unaligned tightly condensed chromosomes. The SAC activation was confirmed by co-staining for the SAC component Bub1. Indeed, Bub1 remained localized to the kinetochore region in colchicine-and MI-181-treated cells.

2.6 Discussion

Cell-based chemical screens represent a fruitful approach to dissect the cell cycle to discover novel compounds and targets for cell cycle modulation. In particular, the cell cycle screen devised in this study avoided the cell sorting step and can be adapted for high-throughput screening. We have successfully applied this approach to identify diverse compounds capable of inhibiting cell proliferation by inducing arrests in different phases of the cell cycle. Many of these hits are novel and their interacting targets warrant further study and characterization.

We focused on M-phase inhibitors and identified MI-181 as novel antimitotic with a benzothiazole-based scaffold. The compound arrests cells in mitosis due to SAC activation by spindle assembly checkpoints and leads to eventual cell death. Subsequent crystal structure results showed that MI-181 binds to same site as colchicine but in a distinct location[37]. We further showed that MI-181 possessed potent anticancer effects through mitotic arrest assays, cell viability assays and multiparametric phenotypic analyses and the cytotoxic effects are competitive against known anticancer agents including taxol and colchicine. Most importantly, MI-181 had broad anti-cancer activity against multiple cancer cell lines particularly for cervical adenocarcinoma and melanoma. Consequently, our study concluded that MI-181 could be potentially used as an alternative to current anticancer agents in clinical application.

CHAPTER 3

Repurposing Anticancer Drugs by Cell Cycle Profiling

3.1 Introduction to drug repurposing

Cancer remains a severe disease that contributes to a high mortality rate in modern society. Despite tremendous investment in cancer drug discovery including high-throughput screening and structure-based drug design, these efforts have not resulted in a significant increase in the number new cancer drugs introduced into clinics [38]. Additionally, the length of time required for developing a new drug has increased from an average 7.9 years to 13.9 years with an average expenditure of US$1.8 billion to introduce a drug to the market [38, 39]. The decline in the productivity from pharmaceutical industry may have severely impacted cancer patients' quality of life and survival. Thus, an alternative approach that allows rapid anticancer drug discovery and deployment in clinical use is urgently needed. High attrition rate in modern anticancer drug discovery can often be attributed to low efficacy or high toxicity in the late stages of clinical trials [40]. On the other hand, FDA approved drugs have superior safety profile and pharmacokinetic properties relating to adsorption, metabolism, toxicity. Consequently, repositioning known drugs for new anticancer indication, known as "drug repurposing" or alternatively, "drug repositioning", "therapeutic switching" represents a promising strategy to accelerate the approval and clinical application of these drugs for cancer treatment. It has been estimated that drug repurposing can effective reduce drug development time down to 3 to 12 years due to significant shortening of lead optimization phase [41]. The basic idea behind drug repurposing is polypharmacology, which suggests that a drug not only can interact with a primary target, but also multiple secondary off-targets. Thus, it is possible to reverse the drug action for the treatment of the original indication to attend other secondary indications. Furthermore, repurposing old drug for new indications only

require minimal or no structural modifications that enable rapid drug approval for the clinics.

Several approaches for drug repurposing have been proposed [39, 42]. The early repurposed drugs were discovered serendipitously by their unexpected side-effects. One notable example is sildenafil (Viagra), a well-known drug for the treatment of erectile dysfunction that was originally intended for heart conditions [43]. Recently, drug repositioning for anticancer agents using high-throughput activity-based screen of disease phenotype as well as *in-silico* prediction are commonly used [39, 44-46]. For example, screening for 2,200 approved drugs for antiproliferative activity in endothethial and prostate cancer cell lines successfully identified the antifungal drug, intraconazole and cardiac glycoside, digoxin as potential anticancer drugs [47, 48]. Similarly, chemoinformatics and bioinformatics approaches through large-scale mining of drug-target information and structure of target and drugs also identified simvastatin, a lipid lower agent as a potential anticancer therapy [49]. Furthermore, many of these repurposed anticancer agents are currently being evaluated in phase I and II clinical trials [39]. In this study, we reported a new systematic approach to repurpose FDA-approved drugs for oncology indications using cell cycle profiling. In contrast to target-based drug repositioning, cell cycle profiling doesn't necessitate predefined drug target or mechanism and can be used to repurpose anticancer drugs targeting diverse pathways. Most importantly, the drug mechanism can be inferred by cell cycle profile similarity, which can be validated experimentally.

A mammalian cell cycle consists of four phases: G1, S, G2 and M phases [2]. To ensure proper cell division, each cell phase is tightly regulated by cell cycle check points. For instance, when cellular DNA undergoes ultraviolent radiation, the DNA damage checkpoint is activated through an ATM/ATR dependent pathway, which leads to a slow progression from G1 to S phase and the accumulation of a G1 cell population for DNA repairs [4, 50]. Similarly checkpoint mechanisms are also presented in G2 phase of the cell cycle, leading to G2/M arrest [51]. On the other hand, chemical damage to spindle apparatus often induced the spindle assembly checkpoint (SAC) response, which arrests cells in M-phase of mitosis and subsequently undergoes apoptosis (cell death) [52, 53]. Based on the accumulated cell phase populations, the mechanisms that induce cycle arrests can be determined. The cell cycle can be profiled using flow cytometry by monitoring

the DNA content with a fluorescent dye [28]. As the cell progress from G1, S to G2/M phase, DNA is gradually synthesized and the cell populations shifts from low fluorescent intensity to high fluorescent intensity region [28].

Still, cell cycle profile changes under chemical perturbation and the correlation between drug mechanism and cell cycle variations are less well understood. Traditionally, anticancer drugs were known to cause arrest in a specific cell phase of the cell cycle. For example, anticancer compounds like staurosporine and paclitaxel that induce strong arrests in G2 or M phases of the cell cycle are categorized as G2 and M-phase inhibitors respectively [28]. Additional cell phase specific inhibitors include camptothecin that arrest cells in S-phase and M-phase inhibitors that target aurora kinases have also been discovered [54-56]. However, a close inspection of the cell cycle profile indicates that these compounds that arrest one cell phase also affect other cell phases [28]. Similarly, for compounds arresting in same cell phase, a slight variation in cell cycle profile was also observed [28]. This indicated the limitations of traditional approach to categorize anticancer compounds by a single cell phase and the whole cell cycle must be fully evaluated to elucidate the anticancer mechanism of the drugs.

Here, we reported a new approach to characterize anticancer drug mechanism based on whole cell cycle profiling and applied computational flow cytometry to repurpose anticancer compounds from FDA approved drugs (Figure 3.1). We found that anticancer compounds often induced cell profile deviations from that of DMSO treatment, thus the drug repositioning can be conducted in a systematic and unbiased fashion disregarding underlying targets or mechanism of actions. To compare the cell cycle profile of different compounds, the cell populations in each phase were encoded as a cell cycle fingerprint and the cell cycle deviation from the control condition were computed as a cell cycle index (CCI) by determining the Euclidean distance from the drug-treated and control profile using the expression CCI= $\sqrt{\mathrm{RG1}^2 + RS^2 + RG2M^2 + RSG1^2}$, where RG1, RS, RG2M and RSG1 are the difference between two fingerprints of four cell phase components.

. A larger CCI value correlate with a stronger cell cycle arrest and can be used as a predictor for compound potency of cytotoxic drugs. Similarly, the cell cycle profile similarity can be used to infer anticancer mechanism of repurposed drugs from the cell cycle profile of known compounds. As a proof-of-principle, we applied this approach to profile 1200 FDA drug library and repurposed 36 approved drugs with potent anticancer activities including 6 compounds with novel anticancer association. Cell cycle profile and structural similarity clustering analysis indicated that many of these drugs are DNA binders and were validated using genotoxicity assays.

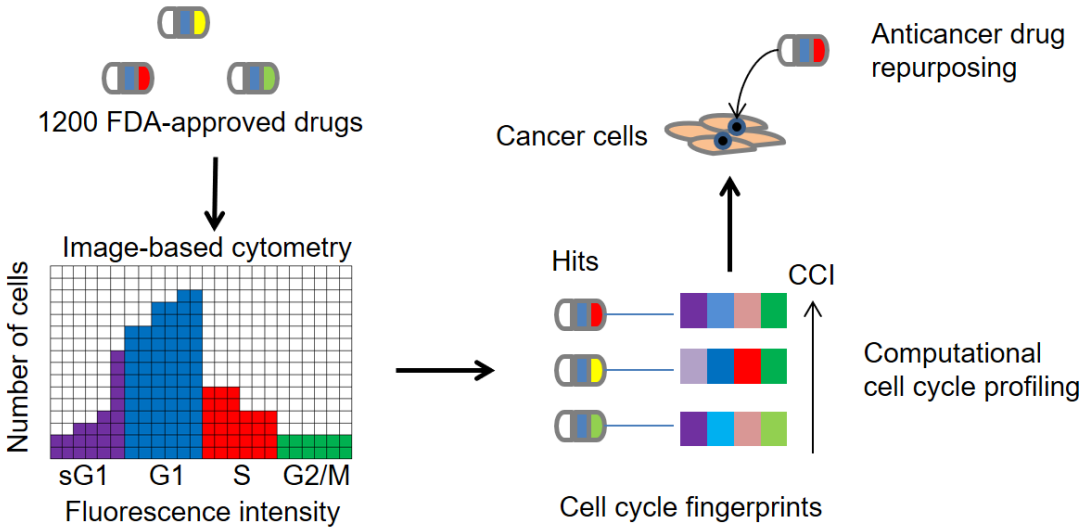

Figure 3.1 – Repurposing anticancer drugs by cell cycle profiling. Drug-induced cell cycle profile of 1200 FDA-approved drugs were constructed using image-based flow cytometry. Computational cell cycle profiles were used to assess potential compound cytotoxicity and anticancer effects. The predicted anticancer FDA-approved compounds were evaluated in cell-based viability assays to determine compound potency. The most potent non-cancer drugs were repurposed as anticancer agents and their mechanism of actions were reevaluated using cell cycle profile similarity and validated by *in-vitro* assays.

3.2 Cell Cycle Profiling of FDA-approved Drugs

We determined the drug-induced cell cycle profile of 1200 FDA-approved drug library in HeLa cells using flow cytometry. The DNA contents of the cells were monitored using fluorescent dye, which allowed the cell population distribution within different cell phases to be measured. The

fluorescent intensity gating for each cell phase was first determined in control cell cycle profile treated with DMSO and taxol normalized to 100%. The control DMSO cell cycle profile indicated that more than 50% of cell populations were in G1 phase followed by 30% in G2/M phase and 10% in S phase. Less than 5% cell subpopulation in subG1 phase due to apoptotic cell death. As further validation, taxol, a known tubulin binder, induced a strong accumulation of 80% of M-phase cell population. Similarly, the DNA binding agent mitoxantrone induced an S-phase arrest response. From this, we hypothesized that antiproliferative properties of the anticancer agents can be approximated by cell cycle profile deviation from that of DMSO control. To quantify this change, we converted cell cycle profiles into fingerprint consisting of four phases (G1, S, G2/M and subG1) and computed the Euclidean distance between drug-treated and DMSO-treated profile as cell cycle index (CCI) (Figure 3.2). Using a CCI cutoff of 10 around one standard deviation from the mean of the CCI distribution (mean=5.532, standard deviation=7.2), we identified 91 approved drugs with diverse cell cycle profile that may exhibit high cytotoxic effects. Several known anticancer drugs were identified in this set including doxorubicin (CCI=77.03), paclitaxol (CCI=57.84), etoposide (CCI=61.92) and others. Among the 91 compounds, 30 compounds are characterized anticancer drugs.

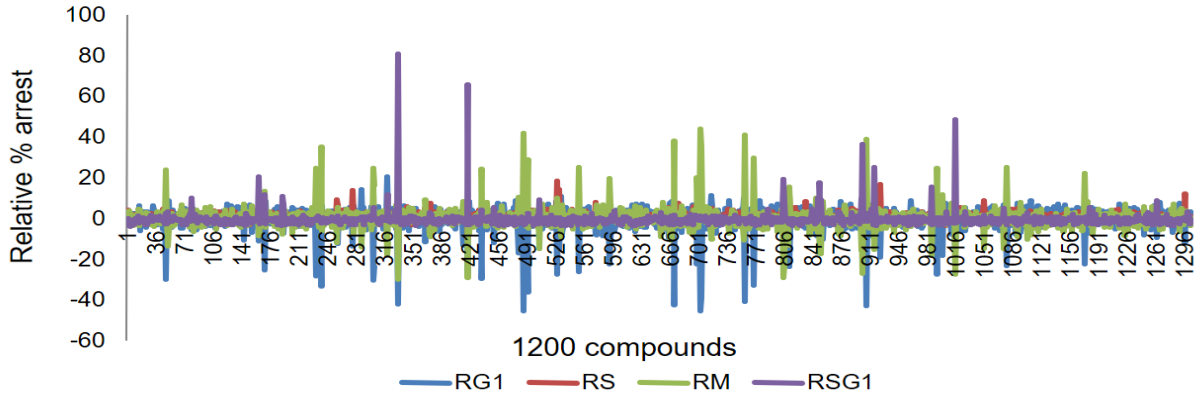

Figure 3.2 – Computational Cell Cycle Profiling of FDA-approved drugs. Drug-induced cell cycle profile of 1200 selected FDA-approved drugs were expressed as cell cycle fingerprint consisting of G1, S, G2/M and subG1 phases relative to DMSO control profile. The diagram showed that FDA-approved drugs induced a wide variation of cell cycle arresting patterns and may correlate with different anticancer mechanisms. Note that the y-axis represents the cell phase change relative to the DMSO control and each cell phase change was indicated by RG1, RS, RM, or RSG1 respectively.

3.3 Cell-based Evaluation of Repurposed Anticancer Drugs

We evaluated the antiproliferative effects of 91 selected FDA approved drugs in HeLa cell culture using cell viability assays, which quantified the ATP level in live cells. The 91 selected compounds were tested for cell inhibition for 48 hours using an endpoint assay at 50uM from which we identified 46 hits (~50%) with > 3SD reduction in cell viability in comparison to DMSO control. Among the 46 selected hits, 41 compounds (89%) have cell viability less than 50% (Figure 3.3A). In addition, CCI also showed a strong correlation with compound potency at this cutoff. In a secondary confirmatory assay, we evaluated compound potency (EC50) of 46 hits by dose-dependent titration under the same conditions. We subsequently confirmed 36 compounds with EC_{50} < 50uM that are potent cytotoxic agents (Figure 3.3A).

The identified known anticancer drugs include tubulin and DNA binding agents. Tubulin targeting agents include microtubule stabilizers, paclitaxel (EC_{50}=3.24nM), parthenolide (EC_{50}=17.33nM) and destabilizers, parbendazole (EC_{50}=0.53μM), febendazole (EC_{50}=2.27μM), mebendazole (EC_{50}=3.39μM), and colchicine (EC_{50}=8.79μM). On the other hand, DNA damaging agents include compounds that targeting topoisomerase machinery such as mitoxantrone (EC_{50}<0.01μM), camptothecine (EC_{50}=0.13μM), etoposide (EC_{50}=0.92μM), podophyllotoxin (EC_{50}=0.38μM), or DNA binders like cyclohexamide (EC_{50}=0.33μM), doxorubicin (EC_{50}=0.52μM), daunorubicin (EC_{50}=1.37μM), and chlorambucil (EC_{50}=1.37μM). Additional 26 compounds identified were not originally indicated for cancer treatments. Among these, several compounds had been previously demonstrated activity against multiple cancer cell lines including proscillaridin A (IC_{50}=5 nM), diethylstillbestrol (EC_{50} =0.15 μM), monesin (EC_{50} =0.38 μM), cardiac glycoside such as digoxigenin, digoxin and lanatoside C (EC_{50} = 0.26 μM, 0.73 μM, 1.84 μM), norgestrel (EC_{50} =1.45 μM), 17-beta estradiol (EC_{50} =2.62 μM), niclosamide (EC_{50} =2.81 μM), statin such as fluvastatin and simvastatin (EC_{50} =3.04 μM, 10.44 μM), eburnamonine (EC_{50} =8.74 μM), alexidine (EC_{50} =8.81 μM), methyl benzethonium (EC_{50} =9.98 μM), ciclopirox (EC_{50} =11.14 μM), and luteolin (EC_{50} =16.1 μM) [57-69]. The cell cycle profiling led to the discovery of novel anticancer activities from broad originally unrelated indications (Figure 3.3B).

Here, we report the discovery of 6 FDA-approved drugs with novel antiproliferative effects including methiazole (EC_{50} =1.37 µM), medrysone (EC_{50} =9.13 µM), nicergoline (EC_{50} =13.34 µM), tribenoside (EC_{50} =13.74 µM), primaquine (EC_{50} =16.16 µM), and GBR 12909 ((EC_{50} =17.91 µM) (Figure 3.3C). medrysone is corticosteroid commonly used in optometry to treat eye inflammation and has recently been shown to reverse lung cancer gene changes [70]. Nicergoline and GBR12909 (Vanoxerine) are psychotic agents for the treatment of senile dementia and cocaine dependence respectively while primaquine is effective against malaria.

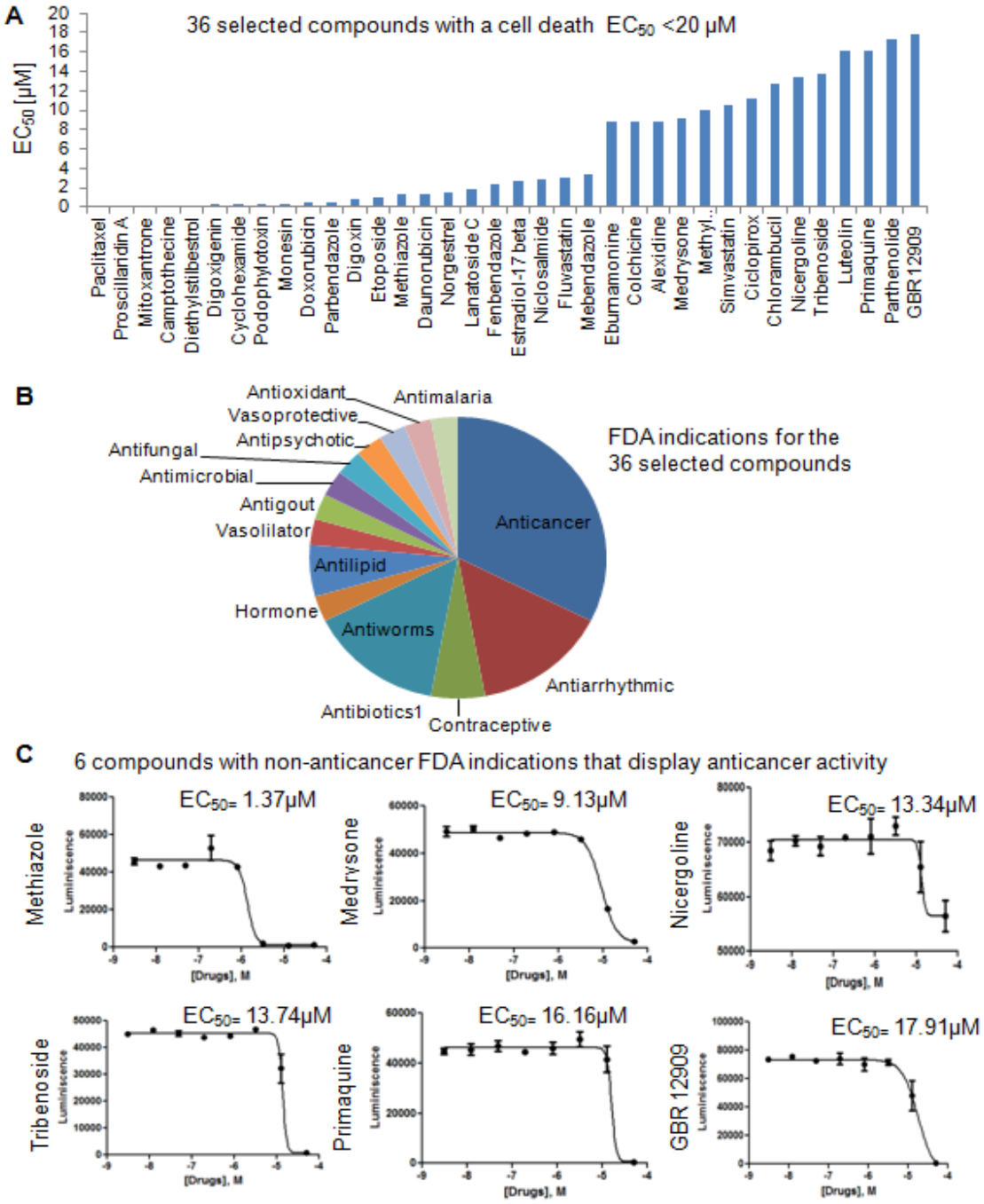

Figure 3.3 – **Repurposed FDA-approved drugs for anticancer effects.** (A) Cell-based viability assay evaluation of 91 FDA-approved drugs identified 36 FDA-approved drugs with anticancer effects. (B) The pie chart showed that the majority of the compounds were not originally indicated for cancer treatment. (C) 6 FDA-approved drugs with novel antiproliferative effects were discovered including methiazole (EC_{50} =1.37 µM), medrysone (EC_{50} =9.13 µM), nicergoline (EC_{50} =13.34 µM), tribenoside (EC_{50} =13.74 µM), primaquine (EC_{50} =16.16 µM), and GBR 12909 ((EC_{50} =17.91 µM).

3.4 Drug Mechanism Inference by Cell Cycle Profile Similarity

Next, we asked if the compounds drug mechanism can be inferred by cell cycle profile similarity. To test this, we clustered the cell cycle profile of the 36 selected FDA-approved drugs using a heatmap (Figure 3.4A). Interestingly, we found that compounds with similar mechanisms of action were clustered together based on unique cell cycle profile signatures. As a retrospective validation, doxorubicin and daunorubicin are structural analogs sharing similar DNA binding mechanism and the cell profile showed that both compounds induced a similar increase in subG1 cell population (Figure 3.4A). These two compounds also clustered in proximity with two additional DNA binding agents including mitoxantrone and cyclohexamide (Figure 3.4A). On the other hand, podophyllotoxin a dual agent that inhibits both DNA replication and tubulin polymerization were clustered with DNA targeting agent chlorambucil and tubulin destabilizing agent colchicines respectively (Figure 3.4A). We subsequently predicted the anticancer mechanism of 6 novel repurposed FDA-approved drugs based on the cell cycle profile of known anticancer compounds. For example, methiazole and primaquine are clustering with mebendazole and parbendazole and are predicted to be a tubulin destabilizing agents. On the other hand, nicergoline is clustered with DNA damaging agent camptothecine and is predicted to be a DNA binder. Similarly, GBR 12909 fall into the same clade as chlorambucil and podophyllotoxin and may be DNA binding agent. Medrysone is predicted to share similar anticancer effects with statin and cardiac glycoside based on cell cycle profile clustering.

To substantiate these predictions, we clustered these 36 FDA-approved drugs into a chemical similarity network (CSN) based on ligand structure similarity (Figure 3.4B). The basic assumption of this approach is the chemical similarity principle, stating that similar compounds will have similar bioactivities. As a validation, the CSNAP network showed that DNA binding agents doxorubicin, daunorubicin and mitoxantrone are all correctly clustering into a DNA binding sub network. Similarly, podophyllotoxin simultaneously linked to tubulin destabilizing agent colchicine and DNA binding agent, confirming its dual mechanism. Using CSNAP3D structural clustering algorithm, we were able to recapitulate predicted drug mechanism from cell profile similarity

comparison by ligand similarity association (Chapter 5). As predicted by cell cycle profile analysis, the vasodilator nicergoline is linked to camptothecine by structural similarity. On the other hand, methiazole was linked to all three tubulin destabilizing agents fenbendazole, mebendazole and parbendazole and are structural analogs to these compounds.

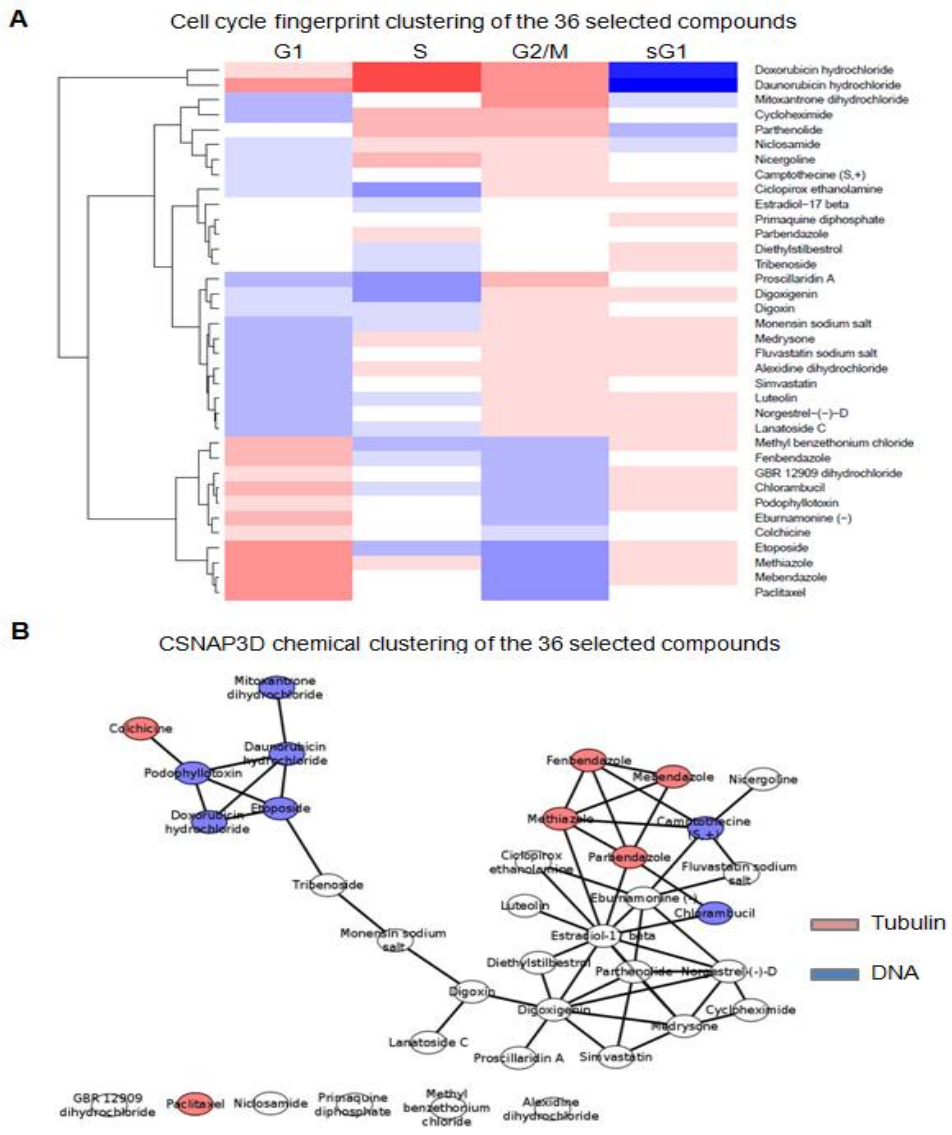

Figure 3.4 – **Mechanism inference by cell cycle profile similarity.** (A) The cell cycle profile of 36 FDA-approved drugs were clustered using heatmap and the anticancer mechanism of non-cancer drugs were inferred from the mechanism of known anticancer drugs in proximity. (B) Computational structure-based network clustering of 36 compounds based chemical similarity. Ligand similarity associations were able to recapitulate predicted drug mechanism from cell cycle profile similarity. Refer to chapter 4 and 5 for a complete description of this approach.

3.5 Discussion

Increased investment in cancer drug discovery to search for new chemical entities (NCE) has not translated into increased development of new anticancer drugs. The majority of NCEs failed in clinical trials due to lack of efficacy and associated toxicities. Consequently, repurposing old drugs for new anticancer indications represent a viable alternative in the new paradigm of polypharmacology. The approved drugs guarantee optimal pharmacological and safety profiles, and the discovery of a new indication for a known agent can be quickly approved for clinical use. In the past, drug repurposing is often the result of serendipitous discovery. While drug repurposing for anticancer indication has recently been attempted, many of which has been focusing on predefined mechanisms or drug targets, thus the diversity of repurposed drug classes is severely limited.

In this study, we presented a new approach for systematic drug repurposing using cell cycle profiling. The approach is based on the assumption that a cytotoxic compound will necessarily disturb the cell cycle profile, leading to a different distribution of cell phase populations. Thus, the cell cycle profile induced by a compound can be used as an indicator of their antiproliferative effect. To quantify the cell cycle profile, we use DMSO treated cell cycle profile as control, and the Euclidean distance of chemical treated profile was determined by cell cycle index (CCI). Using this approach, we repurposed 46 compounds from 1200 FDA-approved drugs that showed antiproliferative effect on HeLa cells. Among them, 36 compounds have an $EC_{50} < 50$ µM evaluated in a cell viability assay including 6 compounds with novel anticancer effects. Similarly, cell cycle profile clustering analysis of 46 approved drugs allowed us to infer anticancer drug mechanism of these drugs from known cell cycle profile based on profile similarity. The predicted mechanism of these drugs was subsequently validated by structural similarity network analysis and experimentally tested in genotoxicity assays.

Although cell cycle profiling has been demonstrated as a promising approach for anticancer drug repurposing and mechanism inference, our study is only limited to cell cycle study of HeLa cells. Since different cancer cell lines will necessarily acquire different drug sensitivity and resistance, thus drug-induced cell cycle profiles will need to be investigated on different cell lines for this approach to be universally applicable. Furthermore, while we have showed that cell cycle profile similarity can be used to infer drug action, the underlying mechanism of how drug induced subtle cell cycle profile variation remains to be elucidated. Potential mechanism includes multi-target drug binding that activate multiple cell phase checkpoint, leading to a different cell population distributions. Similarly, temporal difference in cell cycle progression as well as apoptotic kinetics by drug treatment also warrants further investigation.

In conclusion, we have developed a new approach for systematic drug repurposing and mechanism inference using cell cycle profiling. Cell cycle profiling can potentially be integrated as an integrated computational flow cytometry workflow for large-scale phenotypic-based drug discovery and mechanism inference. This can be expanded to include other cell lines to dissect drug mechanism for a broad array of cancers, including understanding their sensitivity and resistance. Our approach can be universally applied to repurpose a wide range of anticancer drugs for rapid clinical approval, thus increasing the quality of life and survival for cancer patients.

CHAPTER 4

Computational Cell Cycle Drug Target Prediction

4.1 Computational Drug Target Fishing

Computational approaches for predicting the targets, off-targets and poly-pharmacology of hit compounds have been used widely in recent years due to their speed, flexibility and ability to be easily couple to experimental validation techniques [71]. *In-silico* target inference methods include ligand-based and structure-based approaches. Ligand-based approaches, such as similarity ensemble approach (SEA), SuperPred, TargetHunter, HitPick, ChemMapper and others, compare hit compounds to a database of annotated compounds, and drug targets of hit compounds are inferred from the targets of the most similar annotated compounds, based on their chemical structure similarity (Figure 4.1) [25, 72]. The premise of the 2D chemical similarity inference approach is the "chemical similarity principle", which states that structurally similar compounds likely share similar biological activities [73]. The efficiency of 2D chemical search algorithms also led to the wide adoption of this target inference method in public bioactivity database searches including ChEMBL and PubChem [74]. On the other hand, structure-based target inference approaches, such a TarFisDock and INVDOCK, apply reverse panel docking and ranking of docking scores to predict protein targets from pre-annotated structures [75]. In comparison, ligand-based approaches are particularly advantageous due to their speed and algorithmic simplicity and they are not limited by structure availability. However, current ligand-based approaches analyze bioactive molecules in an independent sequential fashion, which has several disadvantages. For example, target inference is based on finding a single most similar annotated compound for a given query ligand, which may not provide consistent target prediction for a group of structurally similar ligands. Additionally, subtle structural changes in the functional groups of active molecules can alter their potency and specificity toward drug

targets; thus, analyzing each molecule independently may not offer a coherent SAR for a congeneric series. This suggests that a more global and systematic analysis of compound bioactivity is required to improve the current state of *in-silico* drug target prediction.

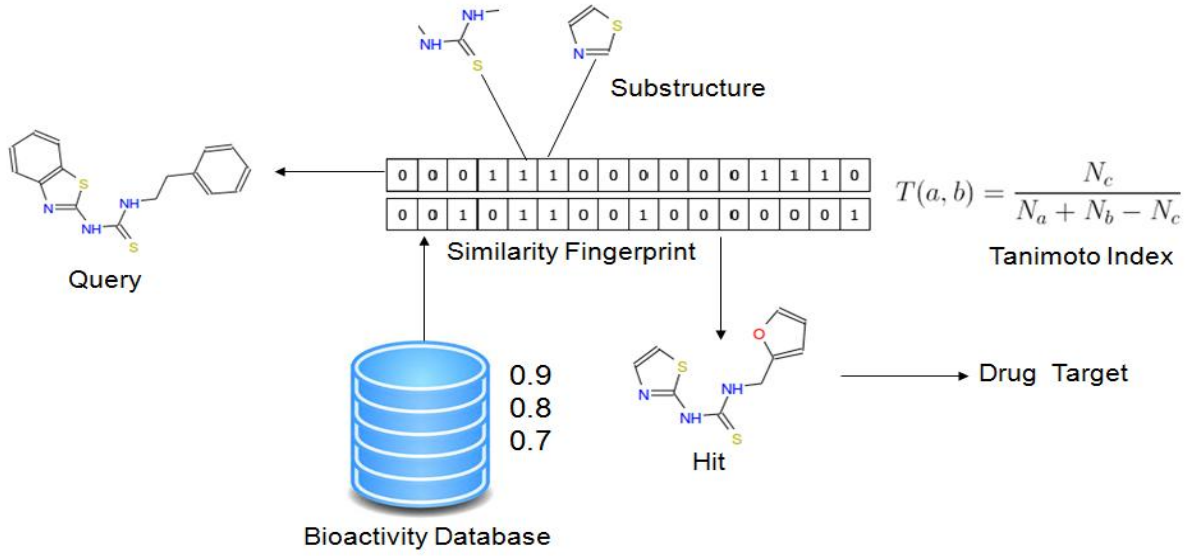

Figure 4.1 – Computational drug target prediction by 2D chemical similarity. Ligand-based drug target prediction approaches compare hit compounds to a database of annotated compounds, and drug targets of hit compounds are inferred from the targets of the most similar annotated compounds, based on their chemical structure similarity. For chemical similarity comparison, Tanimoto index was used to compare two fingerprints to determine the fraction of shared bits with values of 0 or 1, indicating the absence or presence of a particular substructure. The premise of the 2D chemical similarity inference approach is the "chemical similarity principle", which states that structurally similar compounds likely share similar biological activities.

4.2 Description of CSNAP algorithm

We have developed a new computational workflow for compound target deconvolution and prioritization of compounds based on chemical similarity networks that we have termed CSNAP (Chemical Similarity Network Analysis Pull-down) (Figure 4.2). In CSNAP, the Obabel FP2 fingerprints, which characterize molecules by a series of structural motifs as binary numbers (0 and 1), were utilized for structural comparison and compound retrieval from the ChEMBL database (version 16) containing more than 1 million annotated molecules with reported bioactivities (Figure 4.2A and Figure 4.2B)[76]. In comparison to other available fingerprints (FP3, FP4 and MACCS), the FP2 fingerprint uses a path-based algorithm, which has high specificity, is generally applicable to any ligand size and is not limited to pre-defined substructure patterns. To retrieve structurally similar ligands from the bioactivity database, two chemical similarity search functions were used: a threshold similarity search based on a Tanimoto coefficient (Tc) score and a Z-score [72]. The Tc score is one of the most commonly used metrics for chemical similarity comparison in chemoinformatics, which compares two chemical fingerprints to determine the fraction of shared bits with values of 0 or 1. However, a fixed similarity threshold search may not detect compounds with statistical significant scores; thus, a Z-score was also used to search database compounds based on the overall similarity score distribution of the hits [77]. The target annotations of the selected ChEMBL compounds (baits) most similar to input ligands were subsequently retrieved from the ChEMBL and PubChem databases. Based on the output of ligand similarity comparisons, a chemical similarity network was constructed by connecting pairs of ligands with similarity above a Tc threshold according to a weighted adjacency matrix. This resulted in weighted graphs (networks) in which nodes represent compounds and edges represent chemical similarity (Figure 4.2C).

Target inference of the query compounds within the CSNAP-generated network, which contains both query and reference nodes, is similar to the protein functional assignment in

protein-protein interaction (PPI) networks, where protein functional lineage between a characterized and an uncharacterized protein are used to assign shared protein functions [78]. Multiple scoring schemes have been developed to infer protein functions in PPI networks, including algorithms based on network connectivity, graph topology and modular recognition [79-81]. The most direct network-based scoring scheme is the neighbor counting method, where the annotation frequency in the immediate neighbors is ranked and assigned to the linked queries. Thus, the similarity between PPI networks and CSNs suggested that this approach could be effective for network-based drug target inference. As a proof-of-principle, we applied two neighbor-counting functions, Schwikowski score and Hishigaki score for drug target prediction in CSNAP networks. Specifically, a target consensus statistics score, Schwikowski score (S-score), was calculated by ranking the most common targets shared among the neighboring annotated ligands of each query compound within the network (Figure 4.2D) [79]. Additionally, a Hishigaki score (H-score), a chi-square like test based on the mean target annotation frequency distributed within the whole network, was also implemented to compute a significance value for each drug target assignment [81]. The rationale for applying Schwikowski and Hishigaki scoring functions in CSNAP target inference, apart from their algorithmic efficiency and scalability for large-scale network computation, was their accuracy. For example, it was shown that a Schwikowski score correctly predicted >70% of proteins with at least one functional category in a large-scale *S. cerevisiae* PPI network. Furthermore, a performance comparison in a *S. cerevisiae* network showed that these nearest neighbor approaches offer high specificity and prediction accuracy, making them competitive against more advanced statistical network models including Markov random field (MRF) and kernel logistic regression [78].

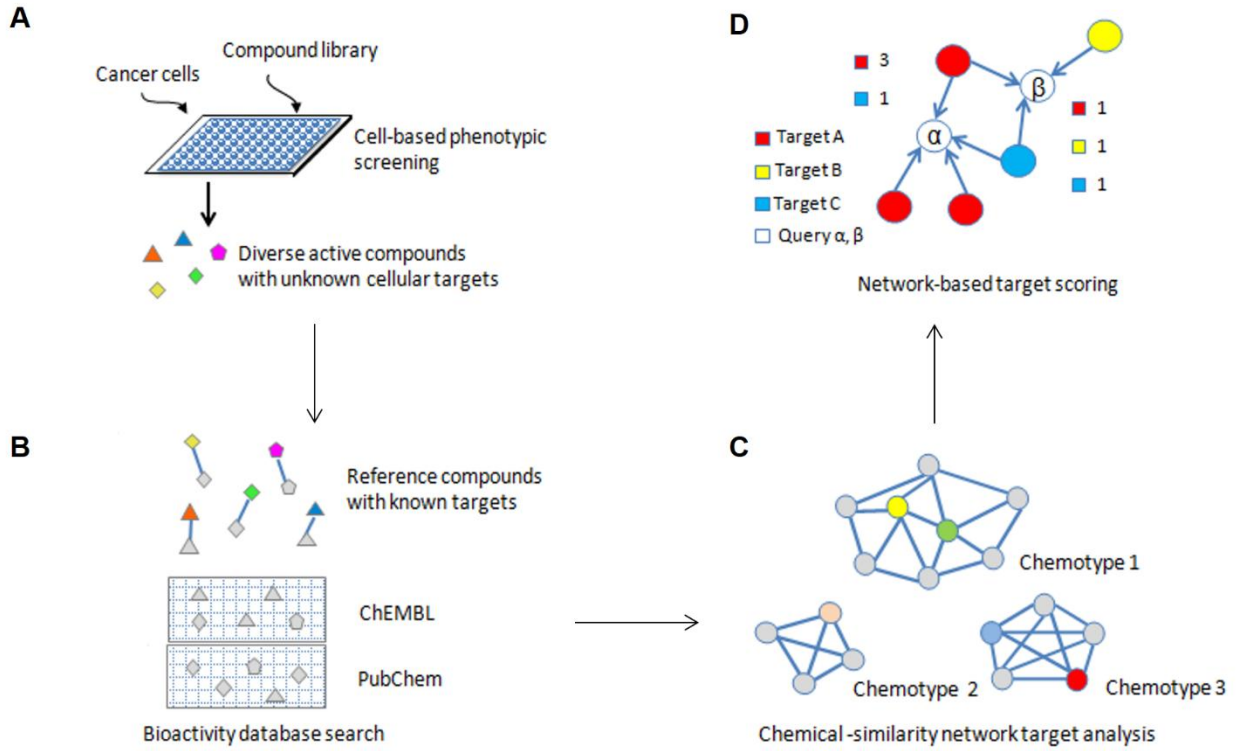

Figure 4.2 – Description of CSNAP algorithm. (A) Diverse hit compounds from a cell-based chemical screen were used as inputs for the CSNAP program to predict their drug targets. (B) In CSNAP, the Obabel FP2 fingerprints, which characterize molecules by a series of structural motifs as binary numbers (0 and 1), were utilized for structural comparison and compound retrieval from the ChEMBL database (version 16) containing more than 1 million annotated molecules with reported bioactivities. The target annotations of the selected ChEMBL compounds (baits) most similar to input ligands were subsequently retrieved from the ChEMBL and PubChem databases.(C) Based on the output of ligand similarity comparisons, a chemical similarity network was constructed by connecting pairs of ligands with similarity above a Tc threshold according to a weighted adjacency matrix. This resulted in weighted graphs (networks) in which nodes represent compounds and edges represent chemical similarity. (D) To infer drug targets from the networks, a target consensus statistics score, Schwikowski score (S-score), was calculated by ranking the most common targets shared among the neighboring annotated ligands of each query compound within the network. For example, compound alpha is predicted to have an S-score of 3 for target A and that of 1 for target C. Additionally, a Hishigaki score (H-score), a chi-square like test based on the mean target annotation frequency distributed within the whole network, was also implemented to compute a significance value for each drug target assignment.

4.3 CSNAP benchmark and methodology validation

To validate CSNAP computationally, we tested CSNAP's ability to correctly predict the assigned targets for annotated compounds as well as its ability to cluster compounds with similar target specificities using a diversity set retrieved from the directory of useful decoys (DUD LIB VS 1.0)[82]. The diversity set contained 206 ligands from 6 target-specific drug classes with known target annotations (including 46 angiotensin-converting enzyme (ACE), 47 cyclin-dependent kinase 2 (CDK2), 23 heat-shock protein 90 (HSP90), 34 HIV reverse-transcriptase (HIVRT), 25 HMG-CoA reductase (HMGA) and 31 Poly [ADP-ribose] polymerase (PARP) inhibitors). CSNAP was able to resolve 206 compounds into target specific chemical similarity sub-networks (Figure 4.3A). We then assessed the prediction accuracy (percentage of correctly predicted ligands) for each drug class by considering the top five consensus targets ranked by S-scores. The results indicated that CSNAP's overall prediction accuracy (recall-like score) for the benchmark compounds was 89% (S-score = 0 with no screen) and 80% (S-score > = 4) respectively. Of those compounds with a prediction, the precision-like score was 94% (S-score = 0 with no screen) and 85% (S-score > = 4) respectively.

To identify potential off-targets for these characterized drugs, we mapped the compound S-score for each drug class against the predicted targets using a ligand-target interaction fingerprint (LTIF), which allowed us to differentiate primary targets from off-targets on a heatmap (Figure 4.3). To further rank the most common targets within the whole compound set, we generated a target spectrum by summing the target prediction score, S-score for each predicted target, by which the heights of the target spectrum can be correlated with the total S-score ($\sum$ S-score). Next, we identified the most probable targets and off-targets from the top peaks above the average $\sum$ S-score. While we cannot exclude smaller peaks as false positives, as they may represent an experimentally verified interaction of the reference compounds in the ChEMBL database, the higher peaks nevertheless represent the most common targets and off-

targets among the analyzed ligands. Within the context of a chemical screen, additional target selection can be aided by gene ontology (GO) analysis, where molecular functions, cellular processes and pathway information can be used to verify the functional role of the predicted targets.

We subjected the diversity set to two different LTIF analyses, first by analyzing each drug class independently and then all drug classes combined (Figure 4.3). Independent LTIF analysis of HIVRT, HMGA and PARP compound sets revealed specific target binding patterns in contrast to CDK2 and ACE, which showed multiple interactions, suggesting potential off-target bindings (Figure 4.3B). From the target spectrum, we identified ENP and CDK1 as the major off-targets for ACE and CDK2 inhibitors respectively, which had been previously reported. For the combined analysis, the targets and off-targets of the 206 benchmark compounds were likewise successfully identified from the target spectrum (Figure 4.3C). Although these validated compounds were "drug-like" and had been optimized for target specificity and transport properties, CSNAP analysis nevertheless identified potential off-targets that were not originally intended for these ligands. This indicated that CSNAP could potentially be used for high-throughput target deorphanization and off-target prediction for bioactive compounds from any chemical screen.

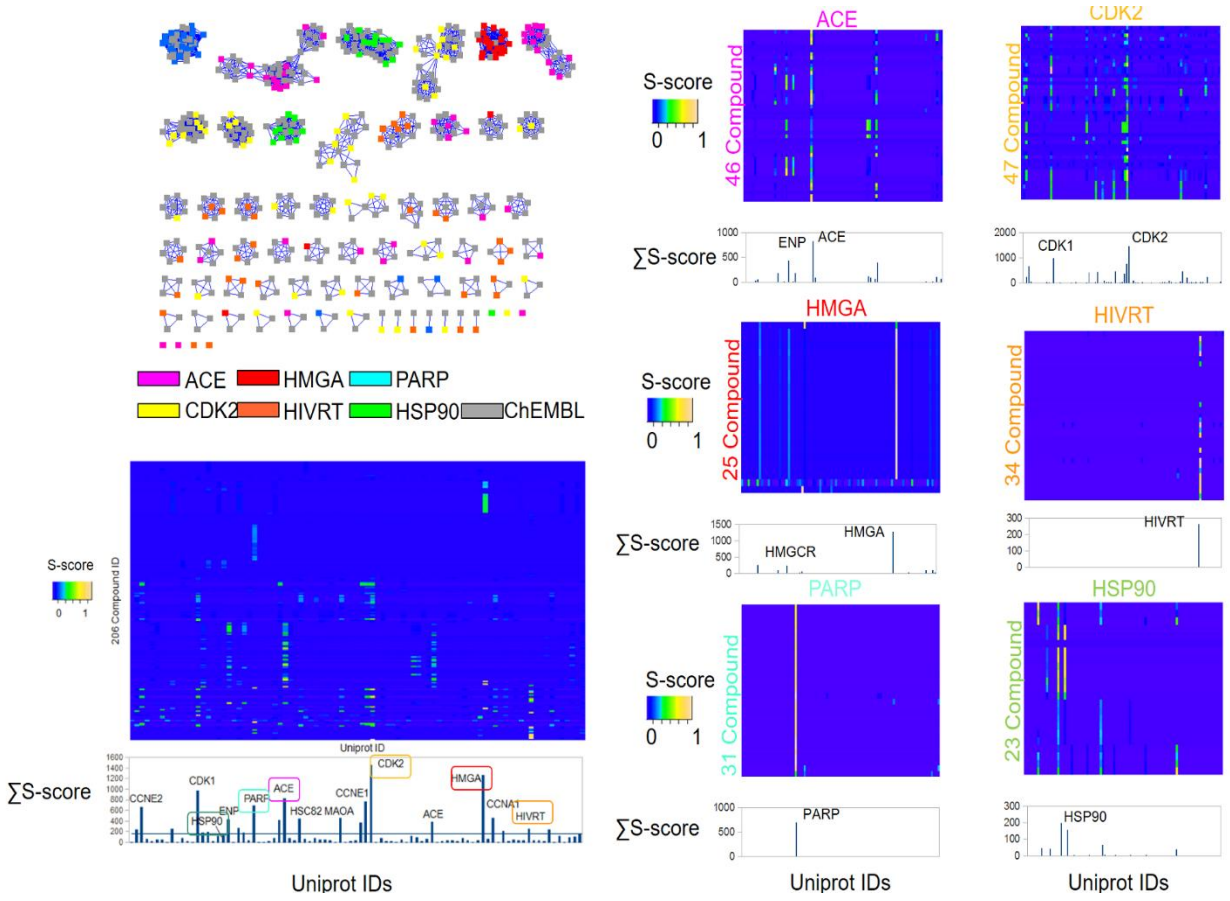

Figure 4.3 – CSNAP validation using benchmark sets (A) To validate CSNAP computationally, we tested CSNAP's ability to correctly predict the assigned targets for annotated compounds as well as its ability to cluster compounds with similar target specificities using a diversity set retrieved from the directory of useful decoys (DUD LIB VS 1.0). The diversity set contained 206 ligands from 6 target-specific drug classes with known target annotations (including 46 angiotensin-converting enzyme (ACE), 47 cyclin-dependent kinase 2 (CDK2), 23 heat-shock protein 90 (HSP90), 34 HIV reverse-transcriptase (HIVRT), 25 HMG-CoA reductase (HMGA) and 31 Poly [ADP-ribose] polymerase (PARP) inhibitors). CSNAP was able to resolve 206 compounds into target specific chemical similarity sub-networks. Diverse hit compounds from a cell-based chemical screen were used as inputs for the CSNAP program to predict their drug targets. (B) Independent LTIF analysis of HIVRT, HMGA and PARP compound sets revealed specific target binding patterns in contrast to CDK2 and ACE, which showed multiple interactions, suggesting potential off-target bindings. Note that the y-axis represents compounds while the x-axis represents predicted targets and the S-score is normalized from 0 to 1 for comparison. From the target spectrum, we identified ENP and CDK1 as the major off-targets for ACE and CDK2 inhibitors respectively, which had been previously reported. (C) For the combined analysis, the targets and off-targets of the 206 benchmark compounds were likewise successfully identified from the target spectrum.

Next, we compared CSNAP's target prediction accuracy with SEA (Similarity Ensemble Approach), a widely used ligand-based target prediction approach based on sequential chemical similarity comparisons, to correctly identify the annotated targets of the benchmark sets (Figure 4.4)[83]. CSNAP showed an overall improvement in prediction accuracy (80–94%) over SEA (63–75%) at identifying the labeled targets of each of the six drug classes from the top 1, top 5 and top 10 score rankings by each respective method. In particular, CSNAP provided substantially better target prediction for promiscuous ligands such as CDK2 and ACE inhibitors (92% and 96%) than the SEA approach (30% and 65%).

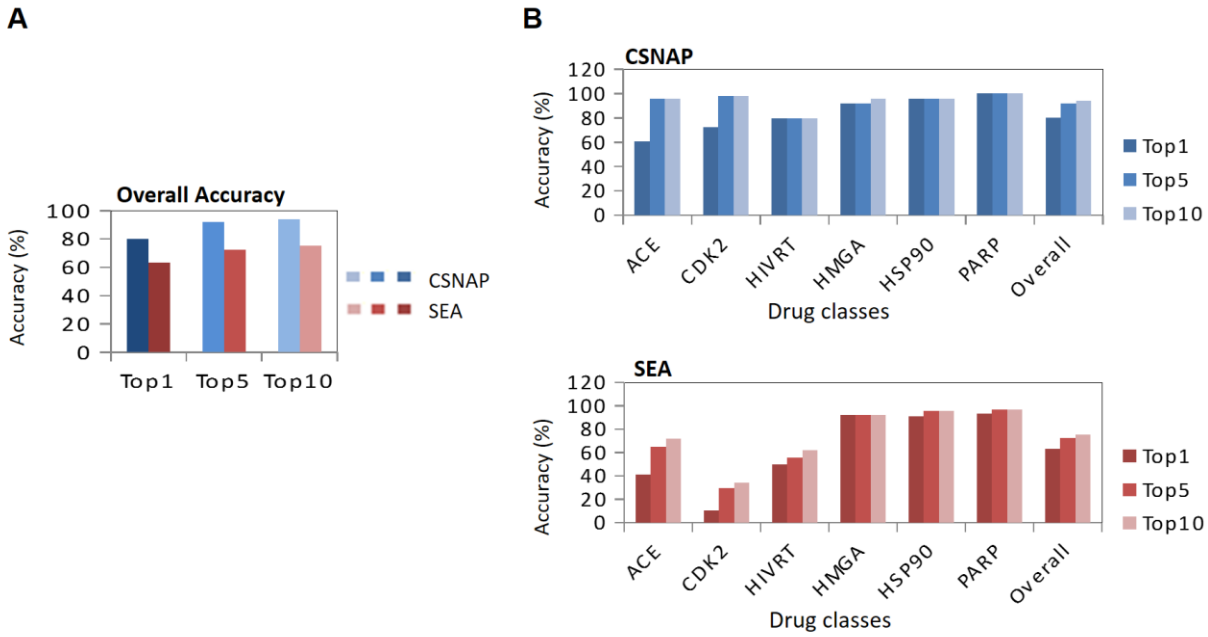

Figure 4.4 – CSNAP performance comparison (A) CSNAP showed an overall improvement in prediction accuracy (80–94%) over SEA (63–75%) at identifying the labeled targets of each of the six drug classes from the top 1, top 5 and top 10 score rankings by each respective method. (B) In particular, CSNAP provided substantially better target prediction for promiscuous ligands such as CDK2 and ACE inhibitors (92% and 96%) than the SEA approach (30% and 65%).

4.4 CSNAP target profiling of mitotic inhibitors

Recently, we performed a high-throughput cell-cycle modulator screen with a diverse, unbiased set of 90,000 drug-like compounds, which identified compounds arresting cancer cells in mitosis (212 compounds). We applied CSNAP to identify the potential targets of the 212 antimitotic compounds. CSNAP analysis generated 85 chemical similarity sub-networks representing diverse chemotypes and retrieved 116 UniProt target IDs from ChEMBL annotations (Figure 4.5A). These targets were analyzed using LTIF with a predefined cutoff ($\sum$ S-score >10) from which we identified 4 broad categories of putative mitotic targets (20 UniProt target IDs) (Figure 4.5B). These included 3 fatty acid desaturases (SCD, SCD1 and FADS2), 1 ABL1 kinase, 5 non-receptor type tyrosine phosphatases (PTPN7, PTPN12, PTPN22, PTPRC and ACP1) and 11 tubulin isoforms. Further compound deconvolution with respect to these targets identified 7 SCD inhibitors, 9 ABL1 inhibitors, 14 PTPN inhibitors and 7 TUBB inhibitors from 6 distinct clusters from the mitotic compound network (including SCD/ABL1: cluster 6, PTPN: cluster 3 and TUBB: clusters 1, 2, 4 and 5) and in which 4 compounds were shown to target both SCD and ABL1 (Figure 4.5C). Meanwhile, by querying the PubChem target annotations with respect to these four target categories, we identified an additional 19 tubulin-associated clusters (total 23), including 51 compounds with unknown bioactivities, which were predicted to be tubulin binders that covered ~20% of our mitotic set. Among the predicted targets were the tubulins (TUBB, including α and β-tubulin), which are the building blocks of microtubules that are essential for mitotic spindle assembly and are established anticancer drug targets[84]. Consistently, several well-known microtubule-targeting agents were identified in the TUBB clusters including mebendazole and nocodazole from cluster 5[84]. Although the compound chemotypes for ABL1, SCD1 and PTPN were known, either identical or analogous to reference compounds deposited in the bioactivity databases, the assay context from which these compounds were retrieved was not related to mitosis [85-87]. Additionally, the function of ABL1, SCD1 and PTPN in mitotic progression had not been explored. Thus, this analysis linked these proteins to important new roles during cell division.

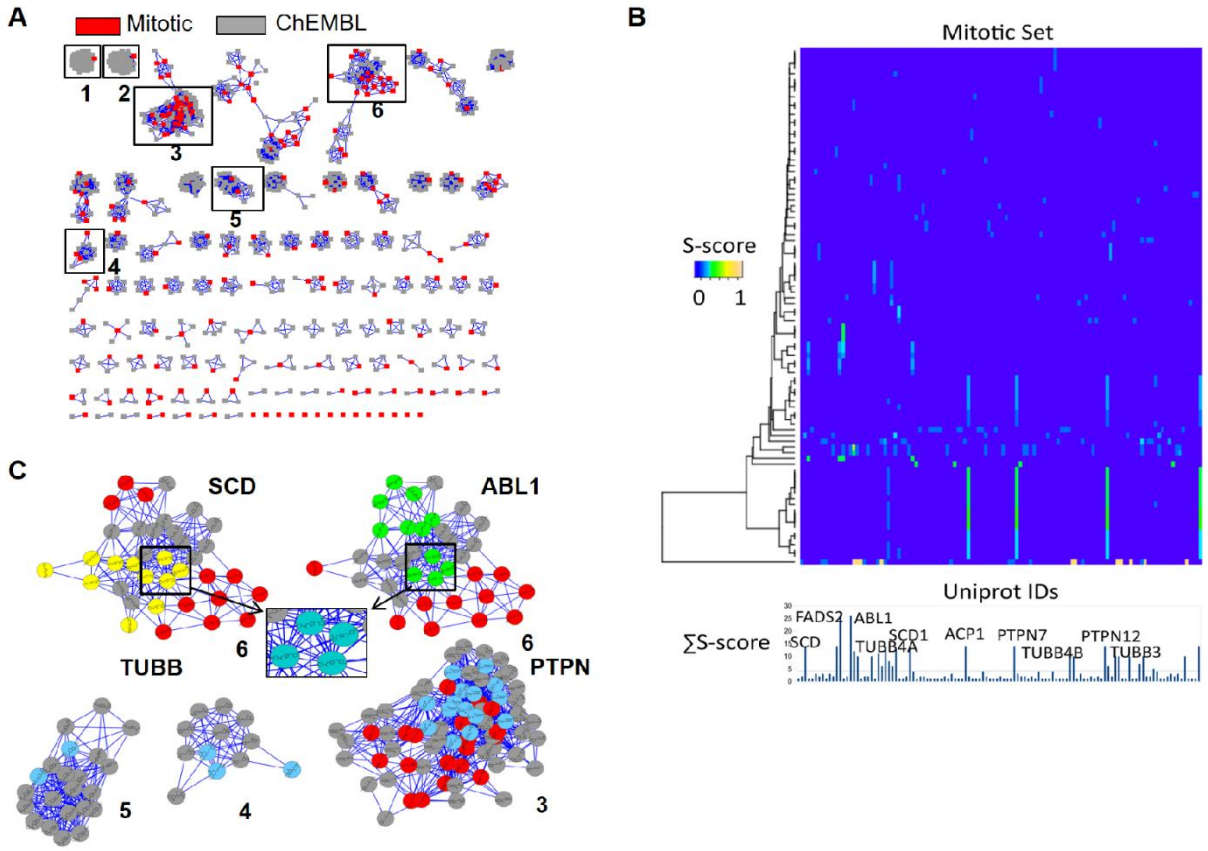

Figure 4.5 – CSNAP target profiling of mitotic sets (A) CSNAP analysis of 212 mitotic compounds generated 85 chemical similarity sub-networks representing diverse chemotypes and retrieved 116 UniProt target IDs from ChEMBL annotations. (B) These targets were analyzed using LTIF with a predefined cutoff (∑ S-score >10) from which we identified 4 broad categories of putative mitotic targets (20 UniProt target IDs). These included 3 fatty acid desaturases (SCD, SCD1 and FADS2), 1 ABL1 kinase, 5 non-receptor type tyrosine phosphatases (PTPN7, PTPN12, PTPN22, PTPRC and ACP1) and 11 tubulin isoforms.(C) Further compound deconvolution with respect to these targets identified 7 SCD inhibitors, 9 ABL1 inhibitors, 14 PTPN inhibitors and 7 TUBB inhibitors from 6 distinct clusters from the mitotic compound network (including SCD/ABL1: cluster 6, PTPN: cluster 3 and TUBB: clusters 1, 2, 4 and 5) and in which 4 compounds were shown to target both SCD and ABL1.

To further substantiate that these compounds were likely inhibiting these targets (ABL1, SCD, PTPN and TUBB), we compared the phenotypes induced by their siRNA knockdown (which often correlates with inhibition of protein activity) with the phenotypes induced upon

treatment with compounds from each target category using immunofluorescence (IF) microscopy (Figure 4.6A). To determine the target siRNA phenotype, we queried the MitoCheck database, which maintains data on the mitotic phenotypes observed upon siRNA knockdown of gene expression for most human genes (Figure 4.6B). As expected, all four target categories (SCD, ABL1, PTPN and TUBB) displayed diverse mitotic defects by siRNA treatment (Figure 4.6C). This included defects in spindle assembly, chromosome segregation and cytokinesis that led to mitotic delay, post-mitotic defects (binuclear and polylobed nucleus) and apoptosis (cell death), suggesting that these targets were critical for cell division. Next, five compounds from these target clusters were selected for phenotypic comparison including compound **1** from the SCD sub-cluster (cluster 6), compound **2** that overlapped with both SCD and ABL1 sub-clusters (cluster 6) and compound **3** from the ABL1 sub-cluster (cluster 6). Additionally, compound **4** and compound **5**, were retrieved from the PTPN cluster (cluster 3) and the TUBB cluster (cluster 4) respectively. All five compounds showed consistent cell phenotypes between siRNA knockdown and drug treatment. However, compound **1** (SCD sub-cluster) also displayed a "large nuclei" phenotype that was specific to ABL1 inhibitors, indicating that it may also target ABL1 based on chemical and phenotypic similarity. As expected, compound **2** (SCD/ABL1 sub-clusters) exhibited a "mixed" phenotype similar to compound **1** while compound **3** was ABL1 specific with very few mitotic delay and apoptotic cells that were specific to SCD inhibitors.

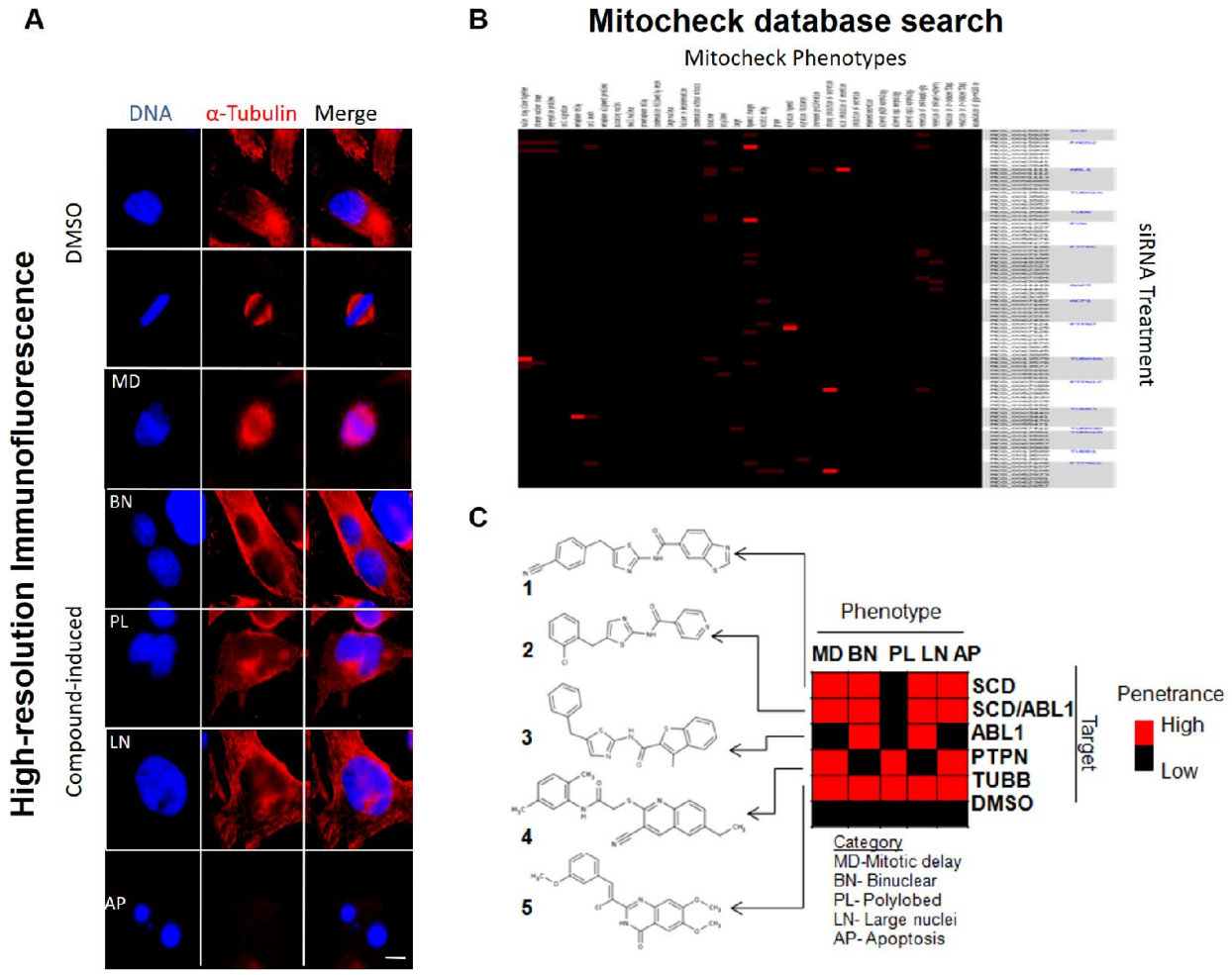

Figure 4.6 – Validation of CSNAP target predictions. (A) To further substantiate that these compounds were likely inhibiting these targets (ABL1, SCD, PTPN and TUBB), we compared the phenotypes induced by their siRNA knockdown (which often correlates with inhibition of protein activity) with the phenotypes induced upon treatment with compounds from each target category using immunofluorescence (IF) microscopy. (B) To determine the target siRNA phenotype, we queried the MitoCheck database, which maintains data on the mitotic phenotypes observed upon siRNA knockdown of gene expression for most human genes. This included defects in spindle assembly, chromosome segregation and cytokinesis that led to mitotic delay, post-mitotic defects (binuclear and polylobed nucleus) and apoptosis (cell death), suggesting that these targets were critical for cell division. (C) Next, five compounds from these target clusters were selected for phenotypic comparison including compound **1** from the SCD sub-cluster (cluster 6), compound **2** that overlapped with both SCD and ABL1 sub-clusters (cluster 6) and compound **3** from the ABL1 sub-cluster (cluster 6). Additionally, compound **4** and compound **5**, were retrieved from the PTPN cluster (cluster 3) and the TUBB cluster (cluster 4) respectively. All five compounds showed consistent cell phenotypes between siRNA knockdown and drug treatment.

4.5 Discovery of novel tubulin-targeting chemotypes

Based on target prediction, we selected microtubules (α and β-tubulin) as our target for *in-vitro* validation. To test CSNAP's prediction that 51 of the 212 mitotic compounds were targeting microtubules, we re-acquired all 212 compounds and tested their ability to perturb microtubule polymerization (stabilize or destabilize microtubules) in an *in-vitro* microtubule polymerization assay at 50μM concentration (Figure 4.7A). The end-point absorbance (dOD) was used to quantify the degree of microtubule polymerization and was converted to percent fold change (F) relative to DMSO drug vehicle (0%), as previously described[88]. Of the 51 compounds predicted to be targeting microtubules, 36 had more than 20% fold change in microtubule polymerization and 14 had no measurable effect. Thus CSNAP was able to predict the targets of this set with > 70% accuracy (Figure 4.7B). In addition, *in-vitro* testing led to the discovery of 96 additional compounds for a total of 132 anti-tubulin agents, including structurally diverse compounds covering ~54 novel chemotypes not discovered in previous chemical screens. Since CSNAP was able to cluster compounds into sub-networks with respect to target specificities, we asked if ligands within the same chemotypic cluster shared a consensus drug-target binding mechanism, as shape complementarity between receptor surface and ligand geometry is essential for inducing a specific cellular phenotype. To test this, we mapped the tubulin polymerization activity onto the mitotic chemical similarity network (Figure 4.7C). Overall, compounds with similar drug mechanisms, e.g. tubulin polymerization or depolymerization were clustered in close proximity within the CSN. However, a few compounds with opposing mechanisms of action were clustered within the same sub-network. This was expected as chemical similarity may not always correlate with compound bioactivity.

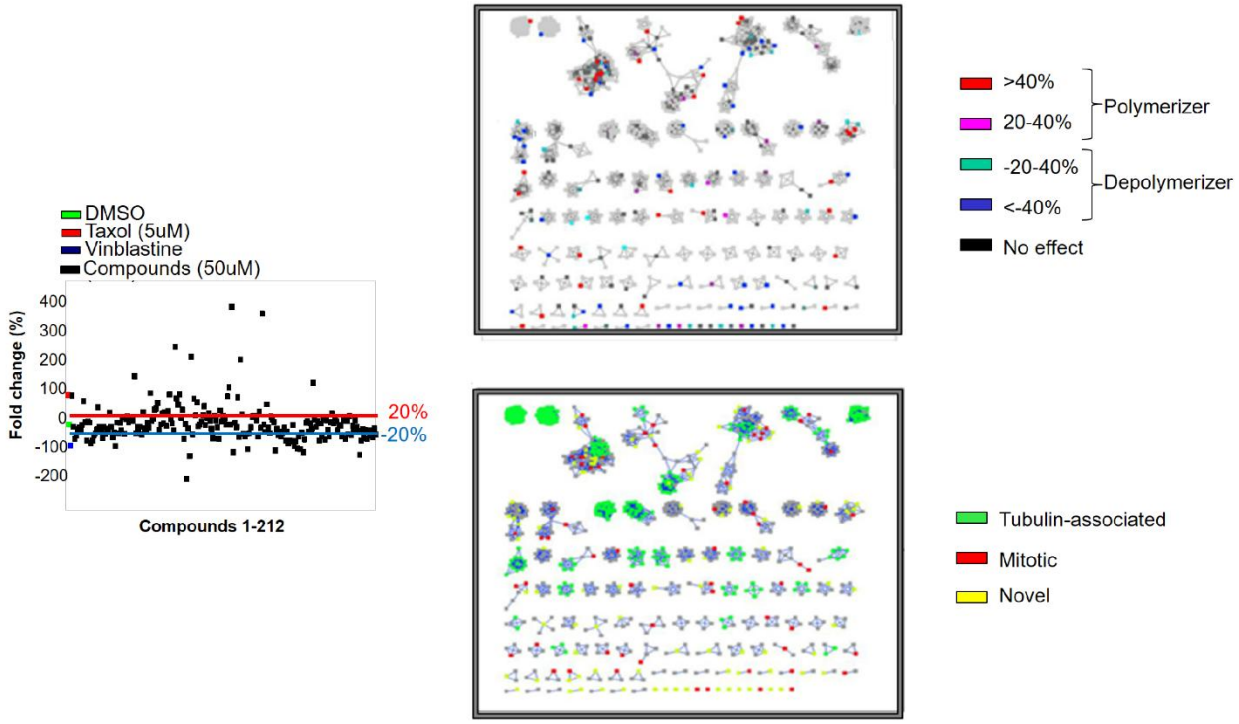

Figure 4.7 – *In-vitro* validation of tubulin as mitotic target. (A) Based on CSNAP target prediction, microtubules (α and β-tubulin) were selected as primary target for *in-vitro* validation. To test CSNAP's prediction that 51 of the 212 mitotic compounds were targeting microtubules, 212 m-phase inhibitors were tested for their ability to perturb microtubule polymerization (stabilize or destabilize microtubules) in an *in-vitro* microtubule polymerization assay at 50μM concentration. The end-point absorbance (dOD) was used to quantify the degree of microtubule polymerization and was converted to percent fold change (F) relative to DMSO drug vehicle (0%). (B) Of the 51 compounds predicted to be targeting microtubules, 36 had more than 20% fold change in microtubule polymerization and 14 had no measurable effect. Thus CSNAP was able to predict the targets of this set with > 70% accuracy. In addition, *in-vitro* testing led to the discovery of 96 additional compounds for a total of 132 anti-tubulin agents, including structurally diverse compounds covering ~54 novel chemotypes not discovered in previous chemical screens. (C) To test if ligands within the same chemotypic cluster shared a consensus drug-target binding mechanism, the tubulin polymerization activities were mapped onto the mitotic chemical similarity network. Overall, compounds with similar tubulin binding effects were clustered in close proximity within the CSN.

Here, we investigated a chemical similarity sub-network consisting of 7 novel anti-tubulin ligands based on a phenyl-sulfanyl-thiazol-acetamide scaffold (Figure 4.8A). Notably, all the connected ligands within the sub-network shared a similar microtubule destabilization effect.

By conducting SAR analysis on the network, we noticed that the addition of hydrophobic groups to the northern and eastern parts of the ligand enhanced microtubule depolymerization.

To provide a structural explanation for this SAR, we observed that compound **6** shared a common structural feature (tri-methoxyphenyl ring) with the microtubule depolymerizer colchicine, suggesting that compounds **6–12**, within the sub-network may share a common colchicine-like binding mechanism (Figure 4.8B). To test this hypothesis, we performed a structural alignment of compound **6** with colchicine and docked the aligned conformations onto the ligand-bound tubulin crystal structure (PDB: 1SA0) (Figure 4.8B). Surprisingly, the predicted binding modes of the two molecules were conserved despite low structural similarity. As further validation of this binding mode, the same binding conformation was also recovered from the top poses by re-docking compound **6** into the colchicine binding site of an apo beta tubulin structure (chain B, PDB: 1FFX), giving a score of-10.82 (London dG) based on free energy binding of the ligand to the receptor site points (Figure 4.8C). The docked structure revealed a consensus pharmacophore between the two aligned ligands including the 2 and 10-methoxy groups and a 9-keto group that interacted with Cys 241 of beta tubulin and Val 181 of alpha tubulin respectively, which had been previously reported[89]. The docking of compounds **7–12** using the same approach also yielded similar binding interactions. The discovery of this consensus-binding model for compounds **6–12** allowed us to link specific protein-ligand recognition features to compound network association and their SAR. For example, the receptor hydrophobicity map showed that the increased potency of compounds **7** and **8**, compared to **6**, could be attributed to the additional interaction of N-propyl group of compound **7** and the N-phenyl group of compound **8** within a sub-pocket enclosed between Leu 248 and Lys 352 of the colchicine-binding site, thus enhancing the protein-ligand interaction.

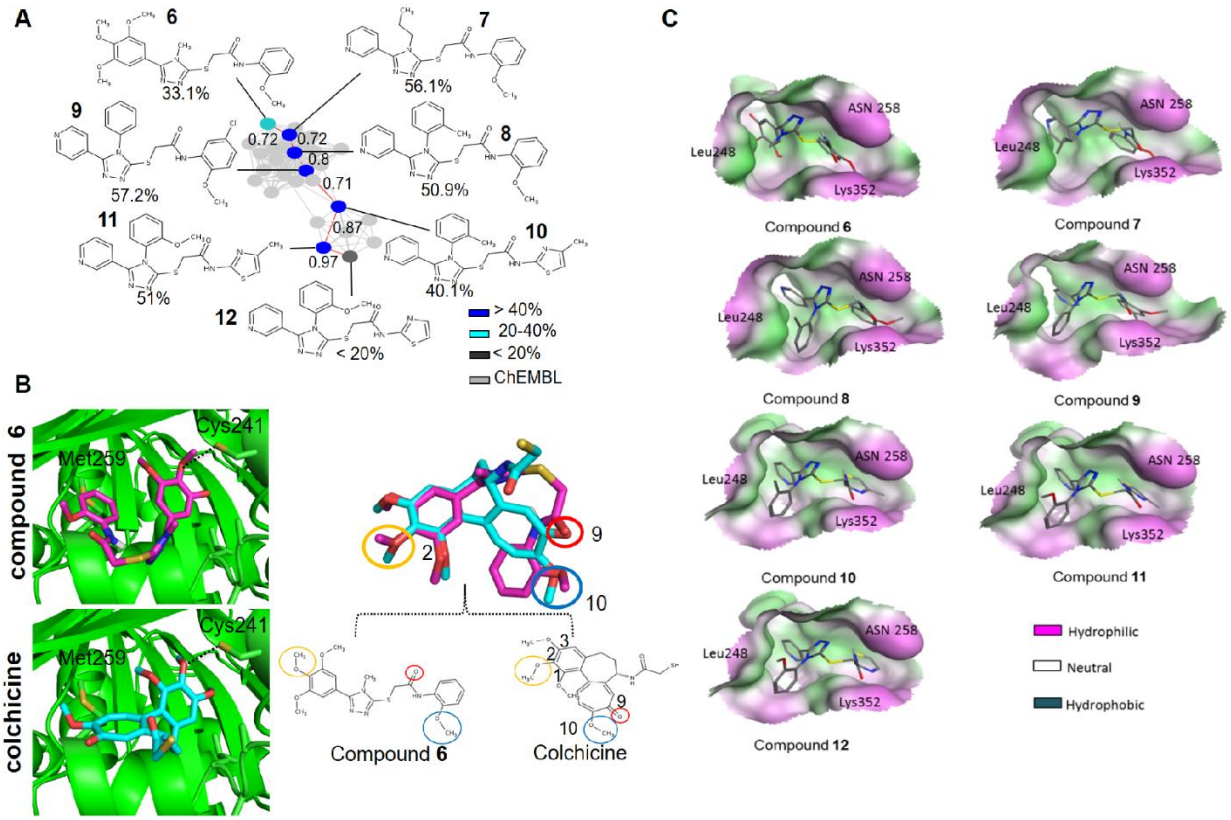

Figure 4.8 – **Discovery of a novel tubulin-targeting chemotype.** (A) A chemical similarity sub-network consisting of 7 novel anti-tubulin ligands based on a phenyl-sulfanyl-thiazol-acetamide scaffold. Notably, all the connected ligands within the sub-network shared a similar microtubule destabilization effect. Note that the addition of hydrophobic groups to the northern and eastern parts of the ligand enhanced microtubule depolymerization. Consistently, a similar SAR trend was observed by evaluating each compound's potency (EC_{50}) in HeLa cells with regards to their ability to arrest cells in G2/M-phase and induce cell death. This identified compound 8 ($EC_{50:G2/M}$ = 33 nM; $EC_{50:\ cell\ death}$ = 60 nM) as the most potent compound in the series. (B) compound **6** shared a common structural feature (tri-methoxyphenyl ring) with the microtubule depolymerizer colchicine, suggesting that compounds **6–12**, within the sub-network may share a common colchicine-like binding mechanism. Structural alignment of compound **6** with colchicine is docked into the ligand-bound tubulin crystal structure (PDB: 1SA0). The predicted binding modes of the two molecules were conserved despite low structural similarity. (C) As further validation of this binding mode, the same binding conformation was also recovered from the top poses by re-docking compound **6** into the colchicine binding site of an apo beta tubulin structure (chain B, PDB: 1FFX), giving a score of-10.82 (London dG) based on free energy binding of the ligand to the receptor site points. The docked structure revealed a consensus pharmacophore between the two aligned ligands including the 2 and 10-methoxy groups and a 9-keto group that interacted with Cys 241 of beta tubulin and Val 181 of alpha tubulin respectively, which had been previously reported. The docking of compounds **7–12** using the same approach also yielded similar binding interactions.

Consistently, a similar SAR trend was observed by evaluating each compound's potency (EC_{50}) in HeLa cells with regards to their ability to arrest cells in G2/M-phase and induce cell death. This identified compound 8 ($EC_{50:G2/M}$ = 33 nM; $EC_{50:\ cell\ death}$ = 60 nM) as the most potent compound in the series (Figure 4.9A and Figure 4.9B). To validate the binding of these compounds to the colchicine site, we used a mass spectrometry-based competition assay where compound **8** competed with colchicine for tubulin binding, similar to the positive control podophyllotoxin (colchicine site binder), and the negative control vincristine (vinca site binder) was unable to compete this interaction (Figure 4.9C) [90]. To test if tubulin was the primary target, we treated HeLa cells with compounds **6–12** and analyzed their effects by IF microscopy. As expected, compounds **6–12** induced a microtubule depolymerization phenotype in HeLa cells (Figure 4.9D) [72]. Thus, the structural binding analysis within a specific sub-network identified a relationship between network connectivity and consensus mechanism, likely due to shape complementarity between protein and ligands. Most importantly, this could be generalized as an effective strategy for structure-based target validation following CSNAP drug target prediction.

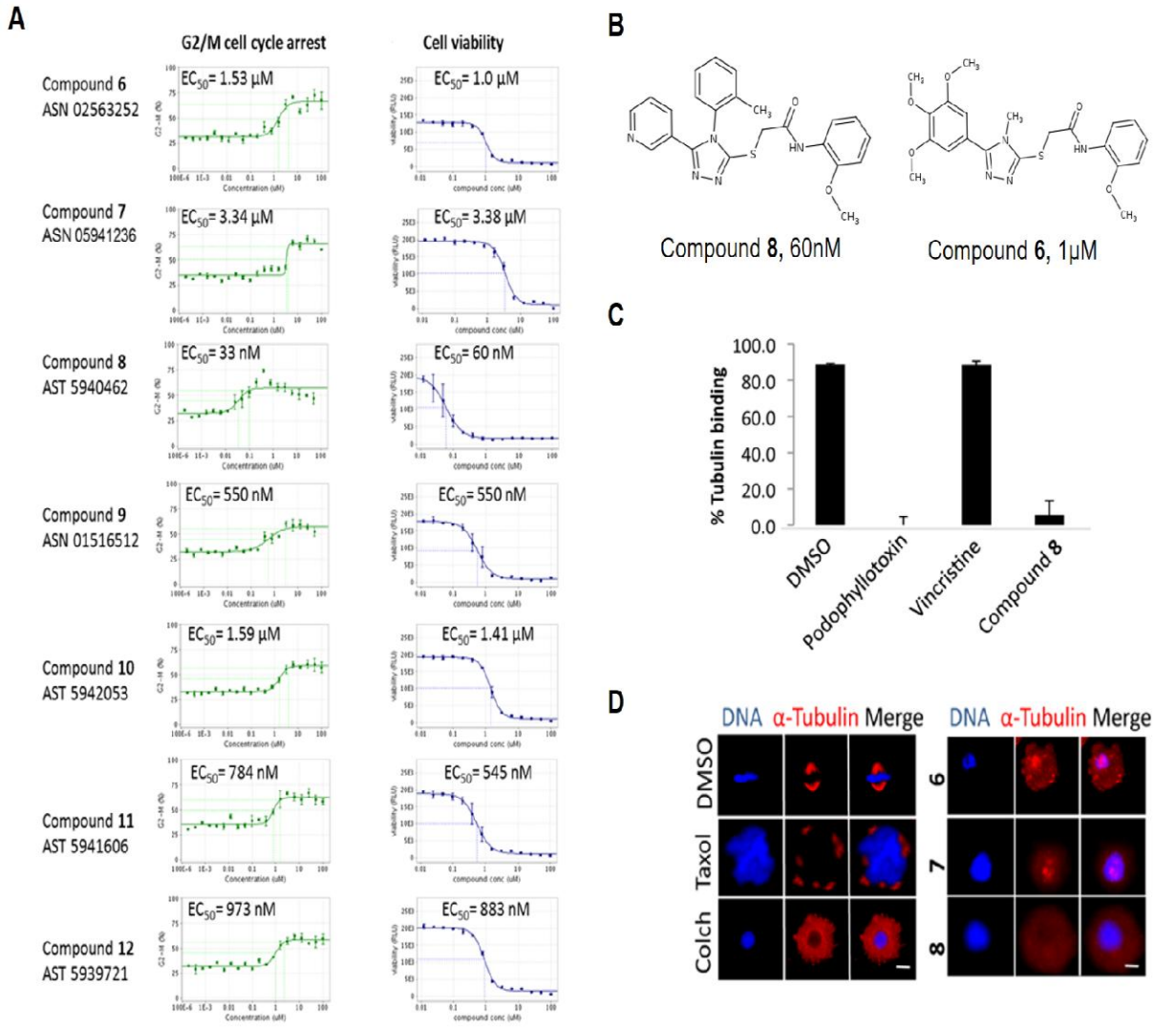

Figure 4.9 – Evaluation of a novel tubulin-targeting chemotype (A)Each compound's potency (EC_{50}) in HeLa cells with regards to their ability to arrest cells in G2/M-phase and induce cell death were evaluated. (B) This identified compound 8 ($EC_{50:G2/M}$ = 33 nM; $EC_{50:\ cell\ death}$ = 60 nM) as the most potent compound in the series. (C) To validate the binding of these compounds to the colchicine site, a mass spectrometry-based competition assay were used where compound **8** competed with colchicine for tubulin binding, similar to the positive control podophyllotoxin (colchicine site binder), and the negative control vincristine (vinca site binder) was unable to compete this interaction. (D) To test if tubulin was the primary target, HeLa cells were treated with compounds **6–8** and analyzed their effects by IF microscopy. As expected, compounds **6–8** induced a microtubule depolymerization phenotype in HeLa cells.

4.6 Discussion

At the completion of cell-based chemical screening efforts researchers are faced with the daunting task of understanding drug mechanism of action and selecting lead compounds from a large number of structurally diverse hits to pursue further. To date, researchers have relied on experimental secondary screens, like multiparametric phenotypic profiling, to select a small number of compounds to validate, which is often costly to conduct and has reduced throughput. On the other hand, computational approaches like simple chemical similarity searches do not capture the bioactivity correlation among the analyzed ligands, leading to prediction inconsistencies and low prediction accuracy. Our study demonstrated that CSNAP, a new computational target prediction methodology that uses chemical similarity networks coupled to a consensus-scoring scheme, improves the current state of the art in *in-silico* drug target identification. First, our benchmark study showed that CSNAP achieved a higher success rate than SEA, an approach based on sequential ligand similarity searches, at identifying pre-annotated drug targets from six major drug classes, especially for promiscuous ligands like CDK2 and ACE inhibitors. Since hit compounds from large chemical screens usually possess sub-optimal target specificity, CSNAP is particularly suitable for deconvolving these compounds compared to the existing approaches. Second, we applied CSNAP to predict and validate the drug targets of 212 mitotic compounds, whose drug binding mechanisms were previously unknown. Here, CSNAP was used in both a positive selection strategy to identify known compounds associated with three new categories of mitotic targets and in a negative selection strategy to identify novel chemotypes targeting microtubules, a major target in cancer drug discovery. Thus, we have demonstrated that CSNAP can achieve accurate large-scale drug target profiling of any compound set without relying on absolute chemical similarity or pre-conditioning from training sets.

However, CSNAP has several limitations. For instance, our tubulin polymerization assays indicated that around 30% of the tubulin targeting compounds were not predicted by

CSNAP. This highlights the general limitation of any ligand-based approach, in that target annotation of the intended chemotype has to be deposited in the bioactivity database *a-priori*. Nevertheless, our structural studies of the novel microtubule depolymerizer compound **6**, whose pharmacophore aligned with the known microtubule targeting agent colchicine, suggests that a chemical similarity measure based on the three-dimensional structure of the compounds could potentially improve CSNAP's prediction power. Likewise, the similarity between CSNAP networks and PPI networks provides further opportunities to apply different PPI network scoring schemes to improve CSNAP prediction. For instance, neighbor counting functions could be readily expanded to consider second-order network neighbors, which has been shown to improve the prediction accuracy of PPI networks. Finally, we showed that incorporating multiple databases, for example PubChem in conjunction with ChEMBL, improved the prediction range of the mitotic compounds by CSNAP. Thus, the simultaneous integration of multiple chemogenomic and bioinformatic knowledge databases can potentially aid the ability of CSNAP to predict the targets of any compound set.

In conclusion, we have developed a new network-based compound target identification method called CSNAP that can be used for large-scale profiling of hit compounds from chemical screens. To further extend the applicability of CSNAP for compound target prediction in a broad array of disciplines, we have made the CSNAP algorithm freely accessible as a CSNAP web server. The web server allows users to analyze up to 300 ligands in parallel, where each ligand can be processed in less than a minute on average. We envision that CSNAP will be instrumental for deconvolving bioactive compounds from past and future cell-based studies relating to the discovery of antiproliferative agents and other processes related to cell division. More broadly, the flexibility of CSNAP to incorporate a wide variety of databases enables it to analyze any active compound set identified from any cell-based high throughput screen, thus expanding its utility across disciplines. Finally, CSNAP should expedite target identification and validation, while limiting costs associated with conventional target identification approaches.

CHAPTER 5

Structure-based Cell Cycle Drug Target Prediction

5.1 Structure-based drug target fishing

In chapter 4, we showed that CSNAP algorithm can be used to infer putative drug targets by similarity comparison. Nonetheless, the targets of the majority of compounds cannot be inferred from 2D structure similarity. For example, a large-scale target identification study of compounds from a zebrafish phenotypic screen using the SEA approach indicated that more than 60% of compounds had no predicted activities against known targets using more stringent criteria [83]. Similarly, CSNAP target profiling of cell cycle inhibitors showed that more than 30% of the compounds do not resemble any known chemotypes [72]. Indeed, one of the major challenges in ligand-based drug target prediction is to deorphanize novel compounds, particularly those ligands with unknown molecular binding partners that normally share low chemical similarity to existing small molecules in bioactivity databases. Consequently, these orphan compounds do not conform to an existing structure-activity relationship (SAR) and their targets cannot be predicted by simple similarity comparisons. While orphan ligands can potentially interact with novel receptor pockets and their targets are difficult to predict computationally, recent studies indicate many of these compounds are capable of interacting with the receptors of known ligands with which it shares a similar 3D environment, known as “scaffold hopping”. This higher level of degeneracy beyond 2D chemical similarity can be applied to improve target prediction based on the 3D structural similarity of compounds.

Several approaches have been proposed to address the limitations of 2D chemical similarity searches. 2D similarity fingerprints like ECFP4 and MACCS have been investigated and are capable of identifying scaffold hoppers. However, the specificity and scaffold enrichment rates are low. While 2D pharmacophore fingerprints have been able to improve novel chemotype detection based on triplets or quartets of pharmacophore features through graph matching, the similarity comparisons do not yield direct molecular alignments and may not indicate true protein-ligand interactions[91]. Recently, 3D molecular similarity comparison programs like PharMapper, ROCS, Piramid, and Pharao predict chemical similarity by 3D conformational alignment based on either shape or pharmacophore descriptors [92, 93]. These approaches align two ligands by optimizing the overlapping molecular volumes between query and annotated compounds to determine a 3D similarity score and have recently been applied for *in-silico* target identification [94]. Nevertheless, target predictability in these studies was only limited for compounds sharing high chemical similarity, whereas for structurally diverse compounds like HIVRT inhibitors that are known scaffold hoppers, the target prediction was suboptimal. Furthermore, since drug targets were predicted based on either molecular shape or pharmacophore scoring alone, the inference of putative targets may not be accurate due to poor shape complimentary with the binding pocket. Most importantly, target predictions were based on pair-wise similarity inference, which is not amenable for large-scale drug target profiling.

Here we describe a new shape-based similarity network approach for large-scale drug target inference called CSNAP3D (Figure 5.1). First, we performed an unbiased computational screen of 28 3D chemical similarity metrics to identify the optimal chemical descriptors for 3D similarity comparison. We showed that 3D similarity metrics based on a combination of shape and pharmacophore scoring provide the strongest improvement in target prediction compared to those based shape or pharmacophore scoring alone. To this end, we have developed the ShapeAlign protocol, which can be used effectively to identify scaffold hoppers using shape alignment followed by a combination of shape, pharmacophore and 2D similarity scoring. In our validation study, we showed that CSNAP3D achieved a high true positive prediction rate of up to

95% for six labeled targets using a benchmark set of 206 compounds. Significant improvement in target prediction was also observed for HIVRT, a challenging drug class consisting of diverse chemical structures that are scaffold hoppers to the nucleotidyltransferase binding site. To test the utility this approach, we performed a CSNAP3D analysis of antimitotic compounds from a cell-based chemical screen and have successfully identified several low molecular weight microtubule-stabilizing agents that mimic the Taxol binding mode and exhibit potent anticancer activity. The Taxol-like mechanism of these compounds was validated experimentally using *in-vitro* microtubule polymerization assays and cell-based assays.

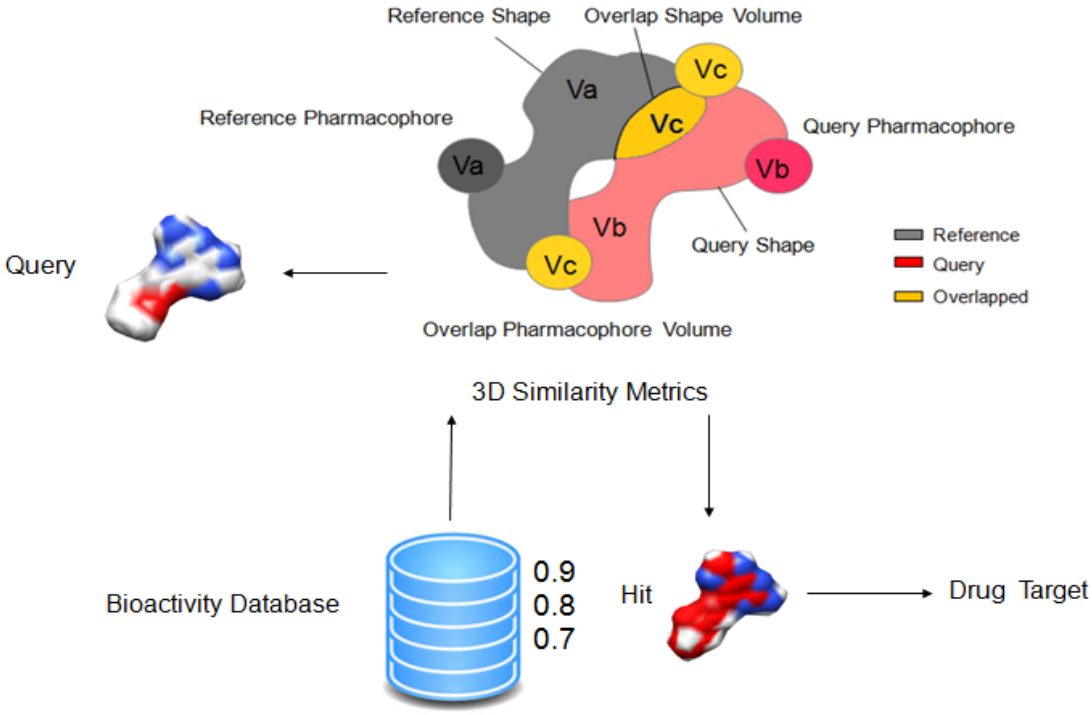

Figure 5.1 – Drug target fishing by 3D chemical similarity. Ligand-based drug target prediction approach compares 3D structure of hit compounds to a database of annotated compounds and drug targets of hit compounds are inferred from the targets of the most similar annotated compounds, based on diverse 3D similarity metrics including pharmacophore and volume similarity.

5.2 Evaluation of 3D ligand similarity metrics

To identify the optimal 3D similarity metrics for ligand structure comparison, we performed an unbiased screen of 28 3D similarity scoring functions generated by four popular 3D similarity comparison programs including Piramid (Shape-it), Pharao (Align-it) and ROCS[95]. These programs aligned molecules based on either molecular shape (Piramid and ROCS) or pharmacophore (Pharao) features using a Gaussian-like function and the ligand alignments were scored using 28 diverse 3D similarity metrics. 3D shape similarity metrics were determined by the percentage of overlapped molecular volume between two aligned molecules including 3D Tanimoto index, Tversky index or maximum overlapped volume[96]. On the other hand, pharmacophore similarity metrics were defined by chemical matching of atomic pairs including hydrogen donor, hydrogen receptor, lipophilic and aromatic center as well as atomic charges[97]. The aligned pharmacophore features were scored using the percentage of overlapped pharmacophore volumes or the number of overlapped pharmacophore points (np). In addition, scoring functions based on a combination of shape and pharmacophore similarity such as ComboScore and ScaledCombo were also provided by the ROCS program. Here, we proposed a new ligand alignment and scoring procedure called "ShapeAlign." ShapeAlign performed an initial shape alignment between query and reference compounds using the Piramid program. The aligned conformations were used as input to the Pharao program, which generated consensus pharmacophore features and the optimal ligand alignment was identified using ComboScore and ScaledCombo based on a combination of shape Tanimoto index and number of matching pharmacophore points (np).

To identify the optimal 3D similarity metrics for scaffold hopping, we retrieved 6 drug classes from the DUD (Directory of Useful Decoy) consisting of 46 ACE inhibitors, 47 CDK2 inhibitors, 25 HMGA inhibitors, 23 HSP90 inhibitors, 31 PARP inhibitors and 34 HIVRT inhibitors and one active conformation for each compound was generated using the MOE

program[82]. Notably, the DUD compounds were known actives for the assigned drug class and were validated binders to the same respective receptor binding sites. One representative compound from each of the six target categories was randomly selected as query and used to test the ability of each 3D similarity metric to enrich class-specific chemical scaffold to the top rank by structure-based similarity search (Figure 5.2A and Figure 5.2B). The heatmap and enrichment curves indicated that several 3D metrics were indeed able to enrich each target specific conformer (Figure 5.2C). To quantify the performance of 3D similarity metrics, the enrichment factors (EF) of each metric was estimated by the area-under-curve (AUC) where 1, indicates high enrichment rate while 0.5, indicates random selection. Overall, ROCS and ShapeAlign similarity metrics based on a combination of shape and pharmacophore scoring gave the highest EF value compared to Piramid that relied on simple molecular shape scoring. While the Pharao program also produced the highest overall AUC values, visual inspection of the aligned conformations revealed that many class-specific compounds did not yield correct superposition perhaps due to the difficulty in resolving multiple alignments from redundant pharmacophore arrangements.

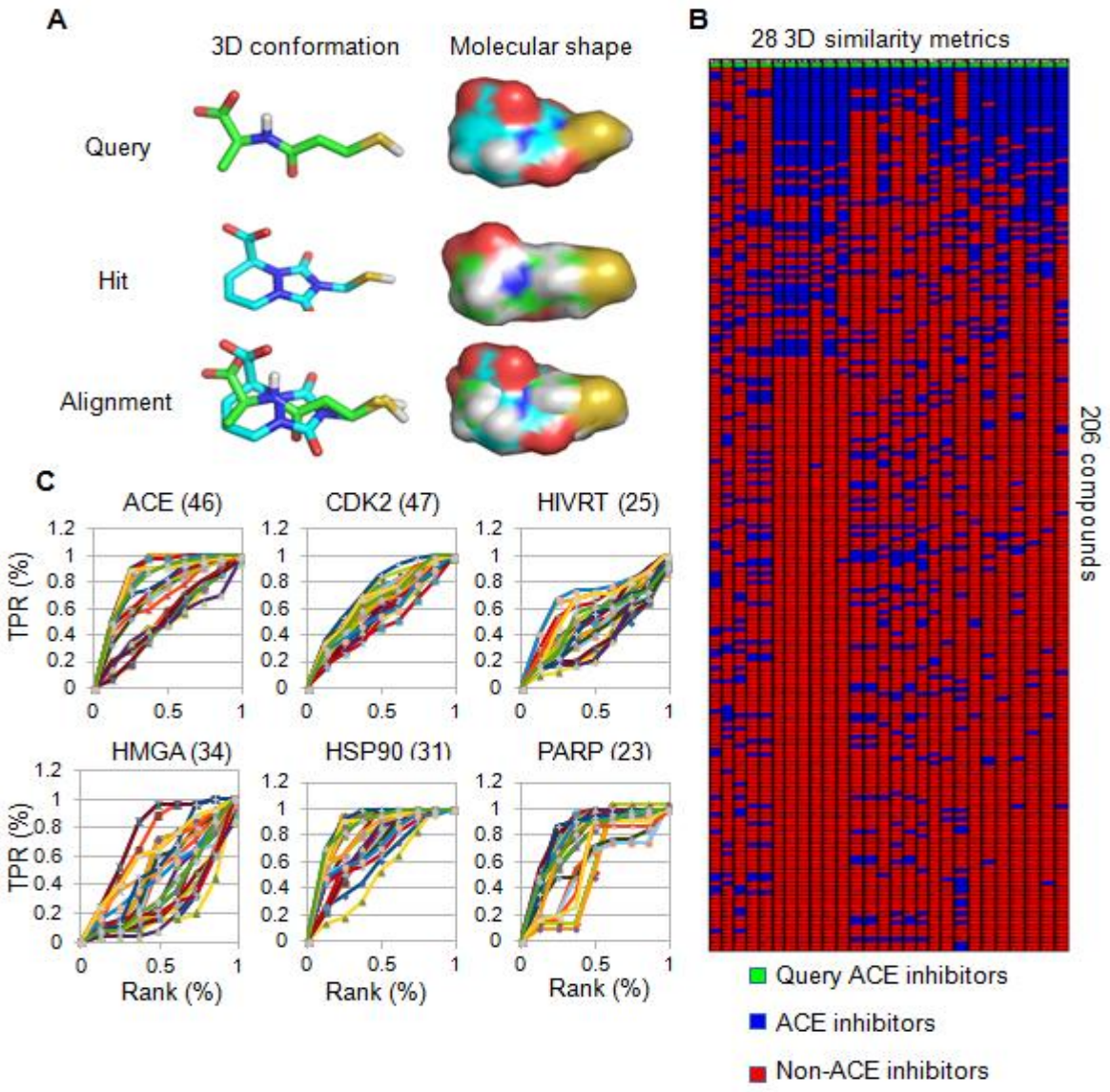

Figure 5.2 – Unbiased screen of 3D chemical similarity metrics. (A) An example of a structural alignment between a query ACE inhibitor and a hit ACE compound with a distinct chemical scaffold generated by the structural superposition algorithm. The 3D chemical similarity metrics were used to measure the molecular shape and overlapping chemical features. (B) Unbiased screen of 28 3D chemical similarity metrics from Piramid, Pharao and ROCS programs. Representative ACE inhibitors were used as a query to test the ability the each 3D chemical similarity metric to enrich for class-specific scaffolds to the top rank from a combined set of 206 benchmark compounds consisting of 6 drug classes. Heatmap shows that the query (green) was retrieved as the top hit for all metrics. Additionally, each metric demonstrated a different ability to enrich for ACE-specific scaffolds (blue) from other drug classes. (C) The percentage of retrieved class-specific scaffolds were plotted against the ranking by each respective similarity score. TPR denotes true positive rate. To determine the performance of each metric, the area under the curve (AUC) was used to compute an enrichment factor (EF).

To further evaluate the ability of these 3D similarity metrics for large-scale compound target profiling, 8 selected 3D similarity metrics including Piramid: ShapeTanimoto, Pharao: PharmTanimoto, ShapeAlign: ScaledCombo, ShapeAlign: ComboScore or in combination with FP2 Tanimoto score were incorporated into the CSNAP3D program and the target prediction accuracy (TPR, FPR) for 206 benchmark DUD compounds was assessed at different similarity threshold levels (0.6-0.9). In addition, the computational efficiency, measured by the amount of processing time per ligand, was also determined. The overall CSNAP3D algorithm is summarized. Briefly, the 206 benchmark compounds were used as query to search the ChEMBL database version 18 using FP2 2D similarity fingerprints and compounds sharing similar structural features evaluated by 3D similarity metrics were clustered into shape-based chemical similarity networks. The TPR and FPR of each compound were assessed based on the predicted targets ranked by a S-score. As expected, the results showed that 3D similarity metrics that used a combination of shape and pharmacophore scoring such as ShapeAlign:ScaledCombo or ROCS:ComboScore provided the highest overall true positive rate (TPR) and lowest false positive rate (FPR) than scorings based on either shape or pharmacophore features (Figure 5.2A and Figure 5.2B). The optimal performance was achieved by combining ShapeAlign:ScaledCombo with 2D FP2 Tanimoto score, which gives a TPR> 95% at Tc of 0.85. On the other hand, ROCS demonstrated the highest computational efficiency among the four programs. A comparison of CSNAP3D and CSNAP2D chemical similarity networks showed that diverse drug chemotypes from the 2D similarity analysis were clustered into a smaller number of 3D shape-based networks with high target specificity, indicating CSNAP3D potentially allowed automated scaffold hopper recognition based on diverse ligand 3D structures (Figure 5.2C and Figure 5.2D). CSNAP3D also improved the target predictability of orphan compounds that normally do not share high 2D similarity in bioactivity databases.

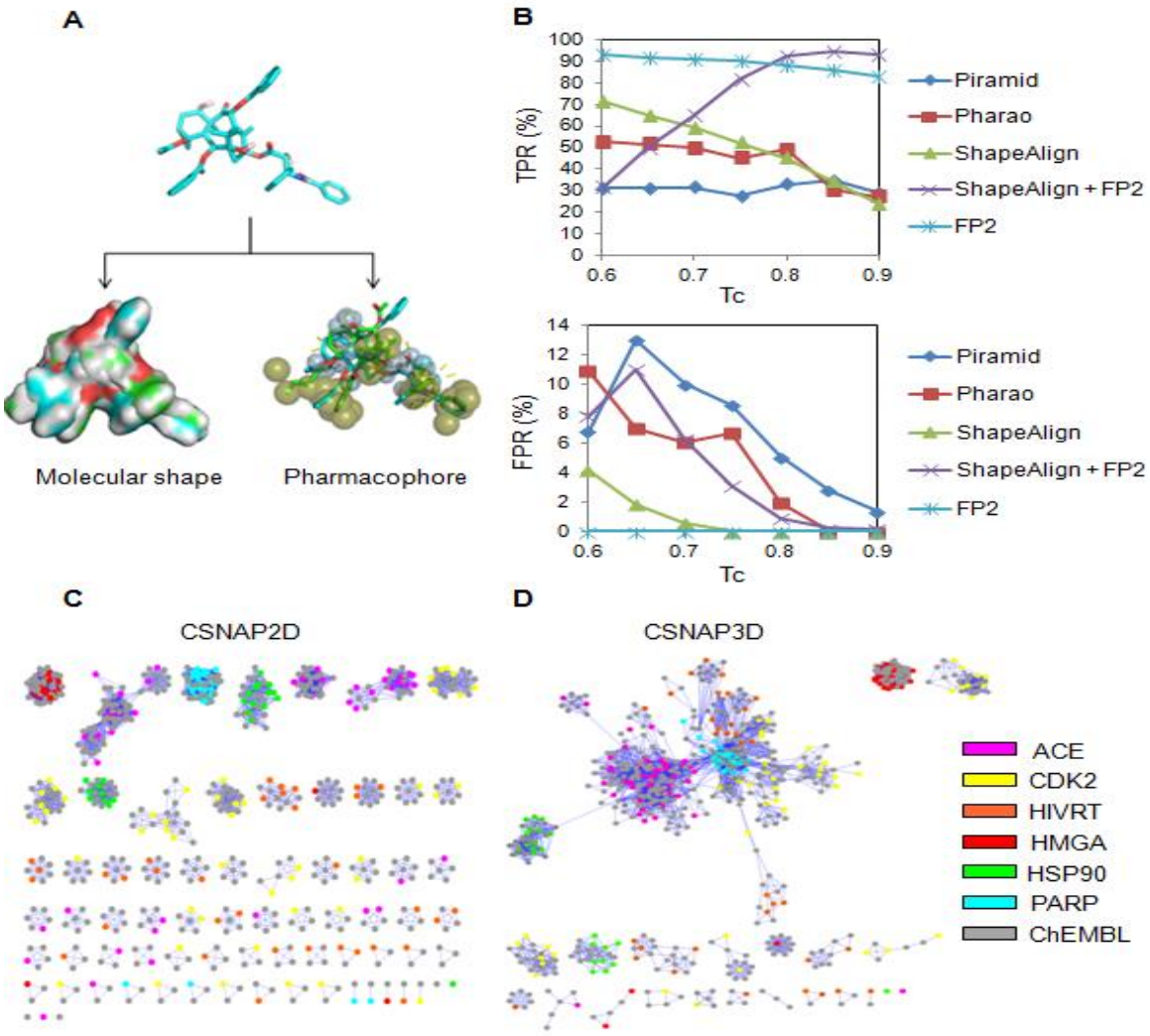

Figure 5.2 – Evaluation of 3D chemical similarity metrics. (A) 3D chemical similarity metrics based on molecular shape (Piramid: Tanimoto), pharmacophore features (Pharao: np) or a combination of shape and pharmacophore scoring (ShapeAlign: ScaledCombo) were incorporated into the CSNAP algorithm.(B) Drug target prediction of 206 benchmark compounds using selected 3D chemical similarity metrics. The percent true positive rate (TPR) and false positive rate (FPR) were evaluated at different chemical similarity thresholds. The results showed that the optimal 3D metrics were ShapeAlign using a combination of shape and pharmacophore scoring. In combination with 2D FP2 score, the ShapeAlign metrics achieved a TPR value > 85% and FPR ~0% at threshold of 0.85. The performance surpassed those based on 2D or 3D chemical similarity metrics alone. (C-D) Comparison between CSNAP2D and CSNAP3D target profiling of 206 benchmark compounds of six known drug classes. 2D CSNAP analysis partitioned 206 known drugs into multiple chemotype-based similarity networks using 0.6 Tanimoto cut off (left). The CNSAP3D algorithm linked the 2D chemotype-based networks into a smaller number of 3D shape-based networks using threshold of 0.85. The results indicate that many benchmark compounds are class-specific scaffold hoppers.

5.3 CSNAP3D target profiling of HIVRT inhibitors

HIV reverse transcriptase is a well-known molecular target for antiviral therapy. Many HIVRT drugs like NNRTIs are known scaffold hoppers to a common nucleotidyltransferase binding site and the binding modes have been well characterized by extensive crystallography and molecular modeling studies. However, recent studies using alignment or non-alignment approaches based on 3D conformation or 2D similarity analysis of the HIVRT DUD set yielded low predictability [98]. The challenge may be attributed to the structural diversity of the NNRTI chemical scaffolds and the lack of target annotations in bioactivity database. This also suggested that current 3D target similarity inference approaches have limited capabilities for detecting scaffold hoppers. Therefore, we performed a CSNAP3D target profiling analysis of the 34 HIVRT inhibitors from the DUD set and the results were compared to that from the CSNAP2D approach. Indeed, our initial CSNAP2D analysis of 34 HIVRT inhibitors from the DUD set resulted in 20 chemical similarity subnetworks corresponding to diverse chemotypes (Figure 5.3A). Further 2D target profiling by mapping the predicted S-scores of each compound on a heatmap indicated that 12 of the HIVRT inhibitors did not yield a prediction, giving a TPR of 65% (Figure 5.3A). In contrast, CSNAP3D analysis using ShapeAlign similarity metrics with 2D FP2 Tanimoto score was able to cluster 34 HIVRT inhibitors into a single shape-based chemical network that relates compounds by 3D structural similarity (Figure 5.3B). The network connectivity supports that many of NNRTIs are in fact scaffold hoppers. Furthermore, 3D target profiling analysis correctly predicted HIVRT as the primary target for 20 HIVRT inhibitors, thus improving the TPR values to 97%. To test if network connectivity was due to 3D scaffold hopping, the aligned ligand conformations generated by the CSNAP3D algorithm were retrieved. CSANP3D correctly identify three FDA-approved HIVRT drugs Enfavirenz, Nevirapine and Tivirapine as scaffold hoppers and the structural alignments were in agreement with previous pharmacophore modeling and SAR studies (Figure 5.3C). In addition, CSNAP3D analysis also identified multiple novel HIVRT scaffold hopping pairs including compound **4** and **5** that shared high 3D similarity and were not previously known.

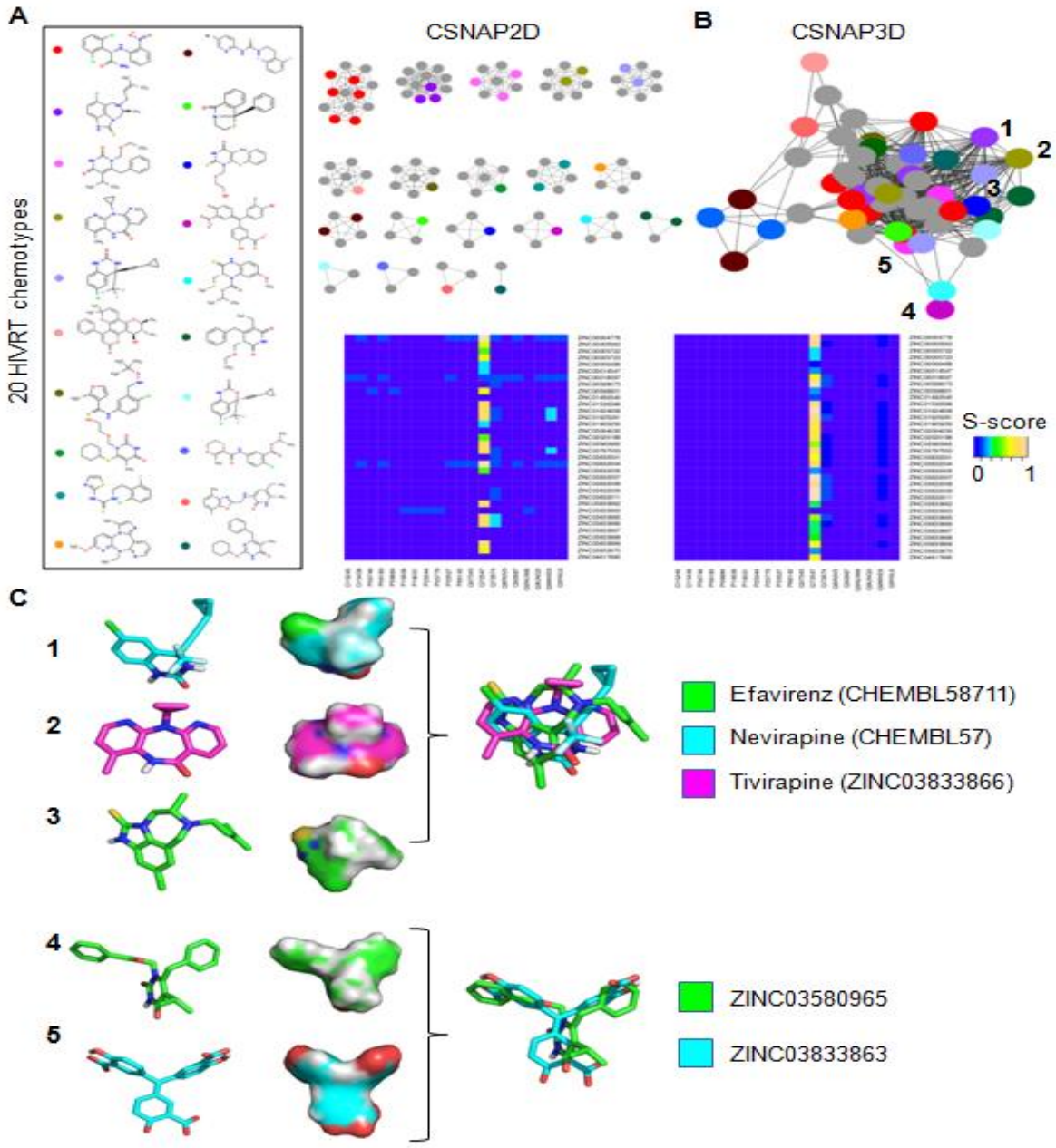

Figure 5.3 – CSNAP3D target profiling of HIVRT sets. (A) CSNAP2D analysis partitioned 20 diverse HIVRT inhibitors into 20 chemotype-based chemical similarity subnetworks. One representative chemotype from each network is shown. Target profiling indicated only 10 compounds were correctly predicted, giving a TPR of 60%. (B) CSNAP3D analysis linked 20 diverse HIVRT inhibitors into a single shape-based chemical similarity network, indicating compound scaffold hoppers. Target profiling correctly predicted 20 compounds, thus improving TPR value up to 80%. (C) Selected HIVRT scaffold hopping pairs predicted by the CSNAP3D algorithm. Three FDA-approved drugs Efavirenz, Nevirapine, and Tivirapine were predicated to be scaffold hoppers and the structure alignment agreed with a previous study.

5.4 Discovery of Taxol mimetics by 3D scaffold hopping

Taxol is a potent anticancer natural product that has been widely applied in the clinics to treat a broad array of cancers. Taxol exerts its major anticancer mechanism by binding and stabilizing microtubules during cell division, leading to M-phase cell cycle arrest and subsequently apoptosis.[99] Although Taxol is one of the most widely-used anticancer drugs used to treat multiple types of cancers including breast, ovarian and lung, the drug is plagued by several issues related to its synthesis and solubility, despite a long history of attempts to circumvent these limitations by structural modifications [100, 101]. Furthermore, many treatment failures can be attributed to its exclusion from the central nervous system due its large molecular weight, acquired drug resistance and dose limiting toxicities. Thus, the discovery of small molecular weight Taxol mimetics that bind to the taxane site with improved transport properties and resistance profiles will be particularly important.

We recently performed a high-throughput cell cycle small molecule screen and identified ~200 compounds that arrested HeLa cells in M-phases of the cell cycle [28]. Further CSNAP target profiling and tubulin polymerization assays indicated that more than 20 compounds strongly stabilized microtubules and were potential Taxol mimetics [72]. However, 2D similarity comparison suggested that these compounds were low molecular weight ligands that shared low structural similarity to Taxol. Thus, we hypothesized that many of these compounds were potential scaffold hoppers to the taxane site. To test this, we performed CSNAP3D target profiling of the 200 antimitotic compounds using ShapeAlign: ComboScore similarity metrics in combination with FP2 fingerprints. CSNAP3D target profiling revealed that the compound predictability had increased in comparison to CSNAP2D where many compounds were predicted to target tubulin due to their network linkage with tubulin stabilizers or destabilizers, consistent with previous tubulin polymerization assay results (Figure 5.4). To identify potential Taxol mimetics, 30 Taxol structural analogs with different 3D conformations were identified from the mitotic compound network and their first order neighbors were extracted followed by 2D similarity clustering (Figure 5.4A). Among the 44 predicted Taxol mimetics, 7 mitotic

compounds were true positives and were validated by end-point tubulin polymerization assays with > 25% fold increases in microtubule mass relative to DMSO control (Figure 5.4B). We therefore focused on analyzing compounds **6-9**, which co-localized onto a 3D similarity subnetwork. Surprisingly, preliminary structural alignment with the Taxol conformation retrieved from CSNAP3D algorithm indicated that these compounds displayed a T-shape 3D conformation similar to Taxol despite possessing a much simpler molecular scaffold.

We further evaluated the anticancer effects of compounds **6-9** in cell viability and mitotic arrest assays to determine their potency (EC_{50}) in cell culture. Compounds **6**, **7** and **9** showed cytotoxic and antimitotic effects with EC_{50} < 5 μM for either cell viability or G2/M mitotic arrest. Our previous study indicated that compound **6** had a >2 fold increase in microtubule polymerization in our tubulin polymerization assays and was the most potent tubulin polymerizing agent in the series. As further validation, we tested compound **6** in tubulin polymerization kinetic assays by monitoring the rate of drug-induced tubulin polymerization as well as the total amount of drug-induced microtubule formation by end-point readings. The results indicated that compound **6** achieved a faster polymerization kinetic at 50uM in comparison to Taxol at 5μM (Figure 5.4D). To further test that compound **6** was exerting a mechanism similar to Taxol, we compared the mitotic arrest phenotype that was induced by the two compounds in HeLa cells. As expected, high-resolution immunofluorescence microscopy showed that both compound **6** and Taxol induced the formation of similar mitotic microtubule asters, a characteristic phenotype of Taxol-like compounds (Figure5.4E).

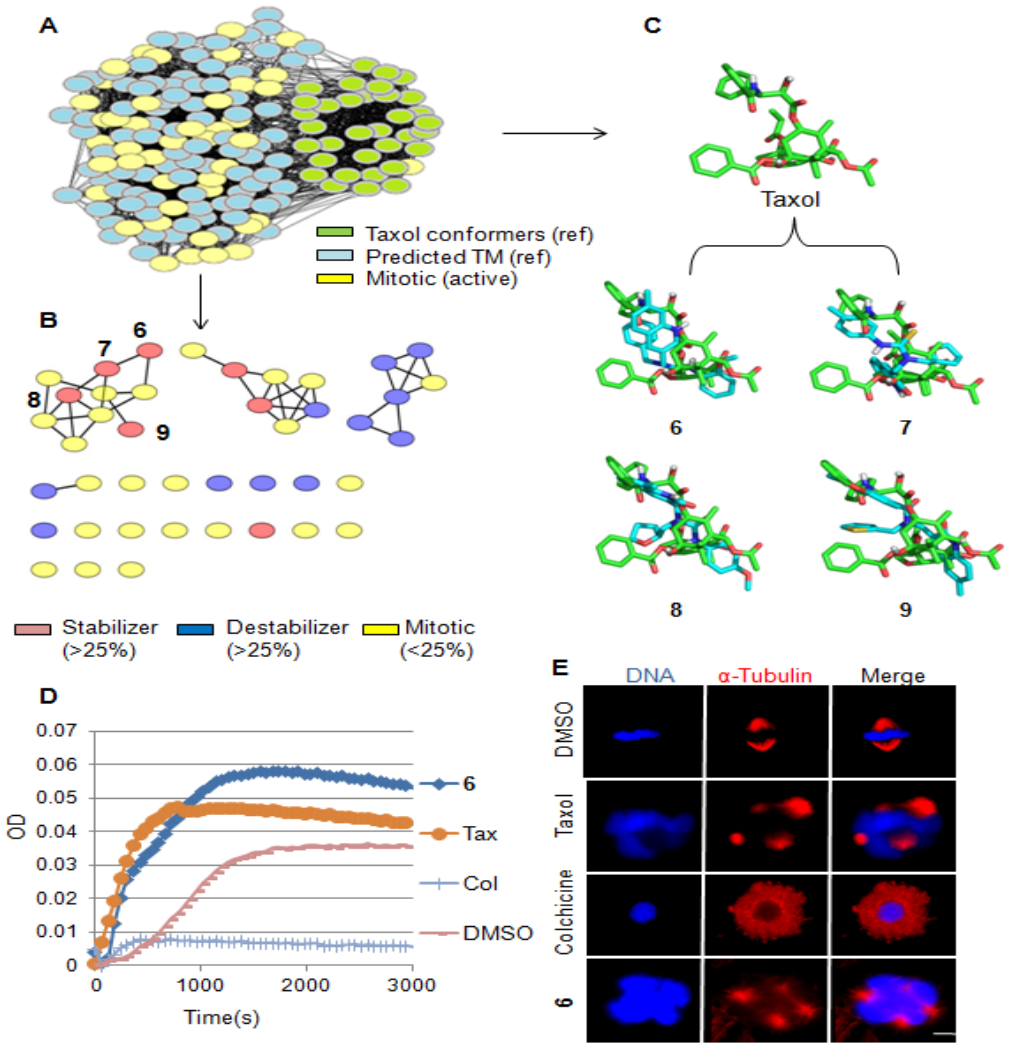

Figure 5.4 – CSNAP3D identify novel taxol mimetics. (A) CSNAP3D analysis of 212 antimitotic compounds identified 36 Taxol structural analogs (green). 60 first order neighbors were retrieved and among them, 44 antimitotic active compounds were predicted to be Taxol mimetics. (B) CSNAP2D analysis partitioned 44 predicted Taxol mimetics into three major chemotype subnetworks. 7 mitotic compounds were true positives and were validated by end-point tubulin polymerization assays with a > 25% fold increase in microtubule mass relative to the DMSO control.4 compounds (compound 6-9) formed a microtubule-stabilizing network. (C) Structural alignment between Taxol mimetics and Taxol. Compound 6-9 shared a similar T-shape conformation similar to Taxol and may interact with the Taxane site of beta tubulin. (D) The polymerization curve showed that compound 6 has a fast polymerization kinetic similar to Taxol and has higher end-point polymerization mass than that of Taxol. (D) Drug-treated mitotic phenotype of compound 6 was compared with Taxol in HeLa cells using immunofluorescence microscopy. Both compound 6 and Taxol induced similar mitotic aster formation, a characteristic phenotype of Taxol-like compounds. In contrast, the negative control colchicine displayed a distinct microtubule depolymerization effect.

5.5 Discussion

We have developed a new structure-based target prediction approach called CSNAP3D that incorporates 3D similarity metrics for large-scale drug target profiling. Our initial computational screen of 28 3D similarity metrics indicates that ligand structural comparisons based on initial shape alignment followed by a combination of shape, pharmacophore and 2D similarity scores provides higher enrichment rates and specificity than those based on shape or pharmacophore alignment alone. This allowed us to develop the ShapeAlign protocol, which applies both the shape alignment program Piramid and the pharmacophore alignment program Pharao to achieve optimal 3D alignment and structure-based target fishing. Interestingly, when the ShapeAlign protocol was further combined with 2D FP2 similarity fingerprints for rescoring, we achieved a 95% positive prediction rate at Tanimoto cutoff of 0.85, which surpassed the performance by CSNAP2D and any other 3D similarity metric. The widely used shape alignment program ROCS was also evaluated in our CSNAP3D program. Similarly, we observed that 3D similarity scores that used a combination of shape and pharmacophore features to rank ligand alignment such as ComboScore provided a robust and highly efficient measure for large-scale target profiling of low-weight small molecular ligands. Notably, the ROCS program achieved the highest processing speed (time/ligand) among the four programs and was also highly competitive against CSNAP2D similarity target profiling. To analyze high-molecular weight compounds including natural products, we also proposed a 3D similarity metric called ShapeAlign:ComboScore. ShapeAlign ComboScore is an unscaled scoring function that simply adds the number of overlapped pharmacophores to the shape Tanimoto score and allows effective detection of multiple pharmacophore features in large molecular alignments including natural products.

We subsequently applied the CSNAP3D algorithm to identify several novel low molecular weight Taxol mimetics from a list of antimitotic compounds identified in a cell-based small molecule screen. While several small molecules that stabilizing microtubules have been identified including Synstab B and GS-164, many of which were discovered by stochastic chemical screens, the binding modes of these ligands is unknown. Automated structural

alignments generated by CSNAP3D showed that compound **6** displayed a T-shape conformation similar to that of Taxol. Further docking studies indicated that a similar conformation was observed in the Taxane pockets from which we were be able to identify a consensus pharmacophore for Taxol mimetics. The identification of this pharmacophore model will motivate further pharmacophore-based virtual screening and structure-based design of small molecule Taxol mimetics. This pharmacophore model also allowed us to explain the increased tubulin polymerization effect of compound **6** that is due to the presence of phenyl moieties (position 2), which are capable of making a pi-pi stacking interactions with residue His229 in the taxane site. The binding mechanism of compound **6** was experimentally validated by kinetic tubulin polymerization assays, where it showed fast microtubule polymerization kinetics similar to Taxol and in end-point microtubule polymerization assays based on OD measurements. Similarly, phenotypic comparison of drug-induced phenotypes indicated that compound **6** was capable of inducing microtubule aster formation in HeLa cells, a signature characteristic of Taxol-like compounds.

Although CSNAP3D substantially improves compound target predictability, particularly for orphan ligands, there are several limitations to this approach. First, structural alignment is based on a single rigid low-energy conformation. Thus, it potentially excludes the possibility that a single ligand can interact with multiple binding sites by adopting different conformations and induce multiple on and off target activities. One potential approach to address this is to consider multi-conformers CSNAP3D networks by structural enumeration. Alternatively, chemical descriptor similarity fingerprints capable of capturing ligand flexibility can likewise be applied. Secondly, while CSNAP3D is capable of detecting scaffold hoppers of ligands that share low 2D chemical similarity but bind in a similar 3D conformation to the same pocket environment, CSNAP3D is not able to deorphanize compounds that interact with orphan receptors that have unknown binding ligands. This challenge can potentially be addressed by considering pseudo-ligands, which can be extracted as a mirror image of receptor pocket with annotated chemical features. We are currently pursuing this approach for large-scale structural polypharmacological target profiling.

In conclusion, we have developed a new computational program CSNAP3D, a 3D upgrade of the CSNAP framework for large-scale network-based drug target prediction based on ligand superposition. Specifically, we have addressed the challenges of target prediction by 2D similarity comparison by using 3D similarity metrics for automated identification of scaffold hoppers and ligand structural alignment for large-scale network target prediction. In addition, we have successfully applied this approach to elucidate the binding mode of small molecule Taxol mimetics and identified a consensus pharmacophore for structure-based discovery. We expect that CSNAP3D will stimulate further methodology development and application on drug target prediction in the new paradigm of structural polypharmacology.

CHAPTER 6

Conclusion

6.1 Experimental Perspectives

Cell-based drug discovery remains the *de facto* approach for identifying potent inhibitors for a wide range of disease including aging, diabetes and cancers. The potential advantage of cell-based screens is the possibility of discovering diverse compounds targeting diverse targets and pathways that does not necessitate specific knowledge of disease mechanism. This has been demonstrated from our cell cycle small molecule screens to identify compounds targeting multiple phases of the cell cycle. Furthermore, we showed that cell cycle profile can be used as a predictor of cytotoxicity. Using this approach, we have successfully repurposed FDA-approved drugs with novel anticancer associations.

Still, the major challenge of cell-based small molecule screen is the identification of drug targets and the mechanism of compounds. Classical proteomics approach using affinity chromatography requires structural modification and can potentially inactivate the compound activity. Furthermore, these approaches are currently costly and have low success rate. Thus, more effective approach will need to be developed. Here, we have explored a novel label-free approach for drug target identification based on cell cycle profile similarity. This approach combines cell flow cytometry and computational analysis to infer molecular drug action and showed high accuracy and precision. The approach is based on a simple assumption that each compound will induce a distinct cell cycle profile signature and this can be used to differentiate drug mechanism of actions. While majority of the compounds can be classified based on the cell phases in which they induce the highest cell cycle arrest, the subtle variation in other phases of

the cell cycle also play a role in uniquely defined the compound mechanism by clustering analysis. Thus, future aspects of computational flow cytometry should be further investigated to elucidate the mechanism of such variations. We hypothesized that many of the variations were due to off-target binding by polypharmacology as well as temporal responses of the cell cycle due to differences in apoptotic kinetics. Similarly, drug-induced cell cycle profiling in different cell lines can be constructed to provide a predictor of drug sensitivity and resistance. Furthermore, it would be interesting to see if drug combinations can induce superpositional cell cycle profiles from individual cell cycle states.

6.2 Computational Perspectives

Computational target identification is by-far the most efficient approach to identify drug targets if the structure of the compound is known. Computational target inference relies on chemical similarity based on the assumption that if two compounds are structurally similar, then they are likely to share similar bioactivity. The strength of similarity-based target inference is that the possibility of devising diverse similarity metrics suitable for a wide range of target identification problems, ranging from 2D/3D similarity, molecular descriptor similarity, shape/pharamacophore similarity. Despite this, current ligand-based approaches have relied on pair-wise comparisons and do not apply to large-scale compound analyses, particularly for hits derived from cell-based chemical screens.

Here, we have developed a novel approach for large-scale drug target profiling called CSNAP. The CSNAP approach uses chemical similarity networks to cluster diverse compounds into sub-networks based on specific chemotypes and the drug targets are inferred using a consensus statistics similar to that used for protein functional network prediction. We have shown that the CSNAP approach can effectively be used to differentiate on and off targets and the prediction accuracy has substantially improved over traditional ligand-based SEA approach. Upon incorporation of 3D chemical similarity metrics, we showed the CSNAP3D's target predictability is further improved and can be used to identify scaffold hoppers, including

compounds that share low 2D chemical similarity but bind to the same receptor site. Nevertheless, similarity-based target inference necessitates a known ligand in a bioactivity database, and the predication cannot be made for compounds that bind to an orphan receptor. Thus, one potential direction is to incorporate structural information into the CSNAP framework. For instance, receptor cavities of all the protein binding pockets, known as the pocketome can be used as a pseudo-ligand for similarity search. This combined approach can potentially circumvent the limitations of ligand-based target prediction.

6.3 Discussion

Cancer remains one of the most critical health challenges worldwide that affects millions of people; consequently, identifying novel anticancer drugs and molecular targets that can be followed up by HTS or virtual screening becomes particularly important. In this thesis, we have developed a combined strategy that tackles drug discovery and target identification through experimentation in conjugation with *in-silico* prediction. Specifically, we focused on the discovery of cell cycle inhibitors which potentially encompassed diverse anticancer mechanisms. To tease out these mechanisms, we applied computational flow cytometry to analyze cell cycle profile similarity as well as developing a new *in-silico* approach for large-scale target profiling based on chemical similarity networks. These approaches can be effectively integrated to follow up any cell cycle small molecule screen or cell-based chemical screen to discover and dissect the anticancer mechanism of potent compounds. Similarly, these approaches can be effectively used to repurpose anticancer drugs from known compounds, thus facilitating the approval of clinical medicine. Thus, chemical dissection of the cell cycle for the discovery of anticancer drugs and identification of novel targets is likely to have an impact on cancer patients' quality of life and survival.

References

1. Diaz-Moralli S, Tarrado-Castellarnau M, Miranda A, Cascante M. Targeting cell cycle regulation in cancer therapy. Pharmacology & therapeutics. 2013;138(2):255-71. doi: 10.1016/j.pharmthera.2013.01.011. PubMed PMID: 23356980.
2. Schwartz GK, Shah MA. Targeting the cell cycle: a new approach to cancer therapy. J Clin Oncol. 2005;23(36):9408-21. doi: 10.1200/JCO.2005.01.5594. PubMed PMID: 16361640.
3. Lee JH, Paull TT. Activation and regulation of ATM kinase activity in response to DNA double-strand breaks. Oncogene. 2007;26(56):7741-8. doi: 10.1038/sj.onc.1210872. PubMed PMID: 18066086.
4. Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu Rev Biochem. 2004;73:39-85. doi: 10.1146/annurev.biochem.73.011303.073723. PubMed PMID: 15189136.
5. Rudner AD, Murray AW. The spindle assembly checkpoint. Curr Opin Cell Biol. 1996;8(6):773-80. PubMed PMID: 8939672.
6. Funk JO. Cancer cell cycle control. Anticancer Res. 1999;19(6A):4772-80. PubMed PMID: 10697591.
7. Dash BC, El-Deiry WS. Cell cycle checkpoint control mechanisms that can be disrupted in cancer. Methods in molecular biology. 2004;280:99-161. doi: 10.1385/1-59259-788-2:099. PubMed PMID: 15187251.
8. Dickson MA, Schwartz GK. Development of cell-cycle inhibitors for cancer therapy. Curr Oncol. 2009;16(2):36-43. PubMed PMID: 19370178; PubMed Central PMCID: PMCPMC2669234.
9. Adamsen BL, Kravik KL, De Angelis PM. DNA damage signaling in response to 5-fluorouracil in three colorectal cancer cell lines with different mismatch repair and TP53 status. Int J Oncol. 2011;39(3):673-82. doi: 10.3892/ijo.2011.1080. PubMed PMID: 21674128.
10. Islam MN, Iskander MN. Microtubulin binding sites as target for developing anticancer agents. Mini Rev Med Chem. 2004;4(10):1077-104. PubMed PMID: 15579115.
11. Peltier S, Oger JM, Lagarce F, Couet W, Benoit JP. Enhanced oral paclitaxel bioavailability after administration of paclitaxel-loaded lipid nanocapsules. Pharmaceutical research. 2006;23(6):1243-50. doi: 10.1007/s11095-006-0022-2. PubMed PMID: 16715372.
12. Carnero A. High throughput screening in drug discovery. Clin Transl Oncol. 2006;8(7):482-90. PubMed PMID: 16870538.
13. Broach JR, Thorner J. High-throughput screening for drug discovery. Nature. 1996;384(6604 Suppl):14-6. doi: 10.1038/384014a0. PubMed PMID: 8895594.
14. Borkin D, Pollock JW, Kempinska K, Purohit T, Li X, Wen B, et al. Property focused structure-based optimization of small molecule inhibitors of the protein-protein interaction between menin and Mixed Lineage Leukemia (MLL). Journal of medicinal chemistry. 2016. doi: 10.1021/acs.jmedchem.5b01305. PubMed PMID: 26744767.
15. Lenart P, Petronczki M, Steegmaier M, Di Fiore B, Lipp JJ, Hoffmann M, et al. The small-molecule inhibitor BI 2536 reveals novel insights into mitotic roles of polo-like kinase 1. Current biology : CB. 2007;17(4):304-15. doi: 10.1016/j.cub.2006.12.046. PubMed PMID: 17291761.
16. Rollinger JM, Stuppner H, Langer T. Virtual screening for the discovery of bioactive

natural products. Prog Drug Res. 2008;65:211, 3-49. PubMed PMID: 18084917.
17. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct. 2003;32:335-73. doi: 10.1146/annurev.biophys.32.110601.142532. PubMed PMID: 12574069.
18. Wang HY, Cao ZX, Li LL, Jiang PD, Zhao YL, Luo SD, et al. Pharmacophore modeling and virtual screening for designing potential PLK1 inhibitors. Bioorganic & medicinal chemistry letters. 2008;18(18):4972-7. doi: 10.1016/j.bmcl.2008.08.033. PubMed PMID: 18762425.
19. Zon LI, Peterson RT. In vivo drug discovery in the zebrafish. Nature reviews Drug discovery. 2005;4(1):35-44. doi: 10.1038/nrd1606. PubMed PMID: 15688071.
20. Futamura Y, Muroi M, Osada H. Target identification of small molecules based on chemical biology approaches. Molecular bioSystems. 2013;9(5):897-914. doi: 10.1039/c2mb25468a. PubMed PMID: 23354001.
21. Rix U, Superti-Furga G. Target profiling of small molecules by chemical proteomics. Nature chemical biology. 2009;5(9):616-24. doi: 10.1038/nchembio.216. PubMed PMID: 19690537.
22. Lomenick B, Olsen RW, Huang J. Identification of direct protein targets of small molecules. ACS chemical biology. 2011;6(1):34-46. doi: 10.1021/cb100294v. PubMed PMID: 21077692; PubMed Central PMCID: PMCPMC3031183.
23. Cereto-Massague A, Ojeda MJ, Valls C, Mulero M, Pujadas G, Garcia-Vallve S. Tools for in silico target fishing. Methods. 2015;71:98-103. doi: 10.1016/j.ymeth.2014.09.006. PubMed PMID: 25277948.
24. Wang L, Xie XQ. Computational target fishing: what should chemogenomics researchers expect for the future of in silico drug design and discovery? Future medicinal chemistry. 2014;6(3):247-9. doi: 10.4155/fmc.14.5. PubMed PMID: 24575960; PubMed Central PMCID: PMCPMC4508123.
25. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. Nature biotechnology. 2007;25(2):197-206. doi: 10.1038/nbt1284. PubMed PMID: 17287757.
26. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nature chemical biology. 2008;4(11):682-90. doi: 10.1038/nchembio.118. PubMed PMID: 18936753.
27. Rajcevic U, Niclou SP, Jimenez CR. Proteomics strategies for target identification and biomarker discovery in cancer. Front Biosci (Landmark Ed). 2009;14:3292-303. PubMed PMID: 19273274.
28. Senese S, Lo YC, Huang D, Zangle TA, Gholkar AA, Robert L, et al. Chemical dissection of the cell cycle: probes for cell biology and anti-cancer drug development. Cell Death Dis. 2014;5:e1462. doi: 10.1038/cddis.2014.420. PubMed PMID: 25321469; PubMed Central PMCID: PMCPMC4237247.
29. Ruegg UT, Burgess GM. Staurosporine, K-252 and UCN-01: potent but nonspecific inhibitors of protein kinases. Trends in pharmacological sciences. 1989;10(6):218-20. PubMed PMID: 2672462.
30. Osherov N, Gazit A, Gilon C, Levitzki A. Selective inhibition of the epidermal growth factor and HER2/neu receptors by tyrphostins. The Journal of biological chemistry. 1993;268(15):11134-42. PubMed PMID: 8098709.
31. Abbott BJ, Fukuda DS, Dorman DE, Occolowitz JL, Debono M, Farhner L. Microbial transformation of A23187, a divalent cation ionophore antibiotic. Antimicrobial agents and chemotherapy. 1979;16(6):808-12. PubMed PMID: 119484; PubMed Central PMCID:

PMCPMC352958.
32. Simon VR, Moran MF. SERCA activity is required for timely progression through G1/S. Cell Prolif. 2001;34(1):15-30. PubMed PMID: 11284916.
33. Tsimberidou AM, Alvarado Y, Giles FJ. Evolving role of ribonucleoside reductase inhibitors in hematologic malignancies. Expert Rev Anticancer Ther. 2002;2(4):437-48. doi: 10.1586/14737140.2.4.437. PubMed PMID: 12647987.
34. Takahashi-Yanaga F, Sasaguri T. GSK-3beta regulates cyclin D1 expression: a new target for chemotherapy. Cell Signal. 2008;20(4):581-9. doi: 10.1016/j.cellsig.2007.10.018. PubMed PMID: 18023328.
35. Clifford B, Beljin M, Stark GR, Taylor WR. G2 arrest in response to topoisomerase II inhibitors: the role of p53. Cancer research. 2003;63(14):4074-81. PubMed PMID: 12874009.
36. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. J Pharmacol Toxicol Methods. 2000;44(1):235-49. PubMed PMID: 11274893.
37. McNamara DE, Senese S, Yeates TO, Torres JZ. Structures of potent anticancer compounds bound to tubulin. Protein science : a publication of the Protein Society. 2015;24(7):1164-72. doi: 10.1002/pro.2704. PubMed PMID: 25970265; PubMed Central PMCID: PMCPMC4500315.
38. Gupta SC, Sung B, Prasad S, Webb LJ, Aggarwal BB. Cancer drug discovery by repurposing: teaching new tricks to old dogs. Trends in pharmacological sciences. 2013;34(9):508-17. doi: 10.1016/j.tips.2013.06.005. PubMed PMID: 23928289.
39. Shim JS, Liu JO. Recent advances in drug repositioning for the discovery of new anticancer drugs. International journal of biological sciences. 2014;10(7):654-63. doi: 10.7150/ijbs.9224. PubMed PMID: 25013375; PubMed Central PMCID: PMCPMC4081601.
40. Pammolli F, Magazzini L, Riccaboni M. The productivity crisis in pharmaceutical R&D. Nature reviews Drug discovery. 2011;10(6):428-38. doi: 10.1038/nrd3405. PubMed PMID: 21629293.
41. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nature reviews Drug discovery. 2004;3(8):673-83. doi: 10.1038/nrd1468. PubMed PMID: 15286734.
42. Jin G, Wong ST. Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. Drug discovery today. 2014;19(5):637-44. doi: 10.1016/j.drudis.2013.11.005. PubMed PMID: 24239728; PubMed Central PMCID: PMCPMC4021005.
43. Boolell M, Allen MJ, Ballard SA, Gepi-Attee S, Muirhead GJ, Naylor AM, et al. Sildenafil: an orally active type 5 cyclic GMP-specific phosphodiesterase inhibitor for the treatment of penile erectile dysfunction. Int J Impot Res. 1996;8(2):47-52. PubMed PMID: 8858389.
44. Swamidass SJ. Mining small-molecule screens to repurpose drugs. Brief Bioinform. 2011;12(4):327-35. doi: 10.1093/bib/bbr028. PubMed PMID: 21715466.
45. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE. Drug Discovery Using Chemical Systems Biology: Repositioning the Safe Medicine Comtan to Treat Multi-Drug and Extensively Drug Resistant Tuberculosis. PLoS computational biology. 2009;5(7). doi: ARTN e1000423
10.1371/journal.pcbi.1000423. PubMed PMID: WOS:000269220100028.
46. Xie L, Evangelidis T, Xie L, Bourne PE. Drug Discovery Using Chemical Systems Biology: Weak Inhibition of Multiple Kinases May Contribute to the Anti-Cancer Effect of

Nelfinavir. PLoS computational biology. 2011;7(4). doi: ARTN e1002037 10.1371/journal.pcbi.1002037. PubMed PMID: WOS:000289973600018.
47. Chong CR, Xu J, Lu J, Bhat S, Sullivan DJ, Jr., Liu JO. Inhibition of angiogenesis by the antifungal drug itraconazole. ACS chemical biology. 2007;2(4):263-70. doi: 10.1021/cb600362d. PubMed PMID: 17432820.
48. Platz EA, Yegnasubramanian S, Liu JO, Chong CR, Shim JS, Kenfield SA, et al. A novel two-stage, transdisciplinary study identifies digoxin as a possible drug for prostate cancer treatment. Cancer Discov. 2011;1(1):68-77. doi: 10.1158/2159-8274.CD-10-0020. PubMed PMID: 22140654; PubMed Central PMCID: PMCPMC3227223.
49. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS computational biology. 2012;8(5):e1002503. doi: 10.1371/journal.pcbi.1002503. PubMed PMID: 22589709; PubMed Central PMCID: PMCPMC3349722.
50. Bakkenist CJ, Kastan MB. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature. 2003;421(6922):499-506. doi: 10.1038/nature01368. PubMed PMID: 12556884.
51. Ree AH, Stokke T, Bratland A, Patzke S, Nome RV, Folkvord S, et al. DNA damage responses in cell cycle G2 phase and mitosis--tracking and targeting. Anticancer Res. 2006;26(3A):1909-16. PubMed PMID: 16827124.
52. Torres JZ, Summers MK, Peterson D, Brauer MJ, Lee J, Senese S, et al. The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. Cell. 2011;147(6):1309-23. doi: 10.1016/j.cell.2011.11.020. PubMed PMID: 22153075; PubMed Central PMCID: PMCPMC4180425.
53. Musacchio A, Salmon ED. The spindle-assembly checkpoint in space and time. Nat Rev Mol Cell Biol. 2007;8(5):379-93. doi: 10.1038/nrm2163. PubMed PMID: 17426725.
54. Albanese C, Alzani R, Amboldi N, Avanzi N, Ballinari D, Brasca MG, et al. Dual targeting of CDK and tropomyosin receptor kinase families by the oral inhibitor PHA-848125, an agent with broad-spectrum antitumor efficacy. Molecular cancer therapeutics. 2010;9(8):2243-54. doi: 10.1158/1535-7163.MCT-10-0190. PubMed PMID: 20682657.
55. Cappella P, Tomasoni D, Faretta M, Lupi M, Montalenti F, Viale F, et al. Cell cycle effects of gemcitabine. Int J Cancer. 2001;93(3):401-8. PubMed PMID: 11433406.
56. Carpinelli P, Ceruti R, Giorgini ML, Cappella P, Gianellini L, Croci V, et al. PHA-739358, a potent inhibitor of Aurora kinases with a selective target inhibition profile relevant to cancer. Molecular cancer therapeutics. 2007;6(12 Pt 1):3158-68. doi: 10.1158/1535-7163.MCT-07-0444. PubMed PMID: 18089710.
57. Denicolai E, Baeza-Kallee N, Tchoghandjian A, Carre M, Colin C, Jiglaire CJ, et al. Proscillaridin A is cytotoxic for glioblastoma cell lines and controls tumor xenograft growth in vivo. Oncotarget. 2014;5(21):10934-48. PubMed PMID: WOS:000348036900054.
58. Koike H, Ito K, Takezawa Y, Oyama T, Yamanaka H, Suzuki K. Insulin-like growth factor binding protein-6 inhibits prostate cancer cell proliferation: implication for anticancer effect of diethylstilbestrol in hormone refractory prostate cancer. Brit J Cancer. 2005;92(8):1538-44. doi: 10.1038/sj.bjc.6602520. PubMed PMID: WOS:000228554600027.
59. Tumova L, Pombinho AR, Vojtechova M, Stancikova J, Gradl D, Krausova M, et al. Monensin Inhibits Canonical Wnt Signaling in Human Colorectal Cancer Cells and Suppresses Tumor Growth in Multiple Intestinal Neoplasia Mice. Molecular cancer therapeutics. 2014;13(4):812-22. doi: 10.1158/1535-7163.Mct-13-0625. PubMed PMID:

WOS:000334342300005.
60. Kepp O, Menger L, Vacchelli E, Adjemian S, Martins I, Ma YT, et al. Anticancer activity of cardiac glycosides At the frontier between cell-autonomous and immunological effects. Oncoimmunology. 2012;1(9):1640-2. doi: 10.4161/onci.21684. PubMed PMID: WOS:000316281700031.
61. Caronti B, Palladini G, Bevilacqua MG, Petrangeli E, Fraioli B, Cantore G, et al. Effects of 17 beta-estradiol, progesterone and tamoxifen on in vitro proliferation of human pituitary adenomas: correlation with specific cellular receptors. Tumour Biol. 1993;14(1):59-68. PubMed PMID: 8493452.
62. Wieland A, Trageser D, Gogolok S, Reinartz R, Hofer H, Keller M, et al. Anticancer effects of niclosamide in human glioblastoma. Clin Cancer Res. 2013;19(15):4124-36. doi: 10.1158/1078-0432.CCR-12-2895. PubMed PMID: 23908450.
63. Chan KK, Oza AM, Siu LL. The statins as anticancer agents. Clin Cancer Res. 2003;9(1):10-9. PubMed PMID: 12538446.
64. Woods JR, Riofski MV, Zheng MM, O'Banion MA, Mo H, Kirshner J, et al. Synthesis of 15-methylene-eburnamonine from (+)-vincamine, evaluation of anticancer activity, and investigation of mechanism of action by quantitative NMR. Bioorganic & medicinal chemistry letters. 2013;23(21):5865-9. doi: 10.1016/j.bmcl.2013.08.095. PubMed PMID: 24055047; PubMed Central PMCID: PMCPMC3867447.
65. Yip KW, Ito E, Mao X, Au PY, Hedley DW, Mocanu JD, et al. Potential use of alexidine dihydrochloride as an apoptosis-promoting anticancer agent. Molecular cancer therapeutics. 2006;5(9):2234-40. doi: 10.1158/1535-7163.MCT-06-0134. PubMed PMID: 16985057.
66. Yip KW, Mao X, Au PY, Hedley DW, Chow S, Dalili S, et al. Benzethonium chloride: a novel anticancer agent identified by using a cell-based small-molecule screen. Clin Cancer Res. 2006;12(18):5557-69. doi: 10.1158/1078-0432.CCR-06-0536. PubMed PMID: 17000693.
67. Zhou H, Shen T, Luo Y, Liu L, Chen W, Xu B, et al. The antitumor activity of the fungicide ciclopirox. Int J Cancer. 2010;127(10):2467-77. doi: 10.1002/ijc.25255. PubMed PMID: 20225320; PubMed Central PMCID: PMCPMC2888914.
68. Lin Y, Shi R, Wang X, Shen HM. Luteolin, a flavonoid with potential for cancer prevention and therapy. Curr Cancer Drug Targets. 2008;8(7):634-46. PubMed PMID: 18991571; PubMed Central PMCID: PMCPMC2615542.
69. Lesiak K, Koprowska K, Zalesna I, Nejc D, Duchler M, Czyz M. Parthenolide, a sesquiterpene lactone from the medical herb feverfew, shows anticancer activity against human melanoma cells in vitro. Melanoma Res. 2010;20(1):21-34. doi: 10.1097/CMR.0b013e328333bbe4. PubMed PMID: 19949351.
70. Fortney K, Griesman J, Kotlyar M, Pastrello C, Angeli M, Sound-Tsao M, et al. Prioritizing therapeutics for lung cancer: an integrative meta-analysis of cancer gene signatures and chemogenomic data. PLoS computational biology. 2015;11(3):e1004068. doi: 10.1371/journal.pcbi.1004068. PubMed PMID: 25786242; PubMed Central PMCID: PMCPMC4364883.
71. Schenone M, Dancik V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. Nature chemical biology. 2013;9(4):232-40. doi: 10.1038/nchembio.1199. PubMed PMID: 23508189.
72. Lo YC, Senese S, Li CM, Hu Q, Huang Y, Damoiseaux R, et al. Large-scale chemical similarity networks for target profiling of compounds identified in cell-based chemical screens. PLoS computational biology. 2015;11(3):e1004153. doi: 10.1371/journal.pcbi.1004153. PubMed

PMID: 25826798; PubMed Central PMCID: PMCPMC4380459.
73. Maggiora G, Vogt M, Stumpfe D, Bajorath J. Molecular similarity in medicinal chemistry. Journal of medicinal chemistry. 2014;57(8):3186-204. doi: 10.1021/jm401411z. PubMed PMID: 24151987.
74. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic acids research. 2012;40(Database issue):D1100-7. doi: 10.1093/nar/gkr777. PubMed PMID: 21948594; PubMed Central PMCID: PMCPMC3245175.
75. Nettles JH, Jenkins JL, Bender A, Deng Z, Davies JW, Glick M. Bridging chemical and biological space: "target fishing" using 2D and 3D molecular descriptors. Journal of medicinal chemistry. 2006;49(23):6802-10. doi: 10.1021/jm060902w. PubMed PMID: 17154510.
76. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. Journal of cheminformatics. 2011;3:33. doi: 10.1186/1758-2946-3-33. PubMed PMID: 21982300; PubMed Central PMCID: PMCPMC3198950.
77. Baldi P, Benz RW. BLASTing small molecules--statistics and extreme statistics of chemical similarity scores. Bioinformatics. 2008;24(13):i357-65. doi: 10.1093/bioinformatics/btn187. PubMed PMID: 18586735; PubMed Central PMCID: PMCPMC2718662.
78. Sharan R, Ulitsky I, Shamir R. Network-based prediction of protein function. Molecular systems biology. 2007;3:88. doi: 10.1038/msb4100129. PubMed PMID: 17353930; PubMed Central PMCID: PMCPMC1847944.
79. Schwikowski B, Uetz P, Fields S. A network of protein-protein interactions in yeast. Nature biotechnology. 2000;18(12):1257-61. doi: 10.1038/82360. PubMed PMID: 11101803.
80. Vazquez A, Flammini A, Maritan A, Vespignani A. Global protein function prediction from protein-protein interaction networks. Nature biotechnology. 2003;21(6):697-700. doi: 10.1038/nbt825. PubMed PMID: 12740586.
81. Hishigaki H, Nakai K, Ono T, Tanigami A, Takagi T. Assessment of prediction accuracy of protein function from protein--protein interaction data. Yeast. 2001;18(6):523-31. doi: 10.1002/yea.706. PubMed PMID: 11284008.
82. Jahn A, Hinselmann G, Fechner N, Zell A. Optimal assignment methods for ligand-based virtual screening. Journal of cheminformatics. 2009;1:14. doi: 10.1186/1758-2946-1-14. PubMed PMID: 20150995; PubMed Central PMCID: PMCPMC2820492.
83. Laggner C, Kokel D, Setola V, Tolia A, Lin H, Irwin JJ, et al. Chemical informatics and target identification in a zebrafish phenotypic screen. Nature chemical biology. 2012;8(2):144-6. doi: 10.1038/nchembio.732. PubMed PMID: 22179068; PubMed Central PMCID: PMCPMC3262069.
84. Dumontet C, Jordan MA. Microtubule-binding agents: a dynamic field of cancer therapeutics. Nature reviews Drug discovery. 2010;9(10):790-803. doi: 10.1038/nrd3253. PubMed PMID: 20885410; PubMed Central PMCID: PMCPMC3194401.
85. Sun S, Zhang Z, Pokrovskaia N, Chowdhury S, Jia Q, Chang E, et al. Discovery of triazolone derivatives as novel, potent stearoyl-CoA desaturase-1 (SCD1) inhibitors. Bioorganic & medicinal chemistry. 2015;23(3):455-65. doi: 10.1016/j.bmc.2014.12.014. PubMed PMID: 25555732.
86. Manetti F, Falchi F, Crespan E, Schenone S, Maga G, Botta M. N-(thiazol-2-yl)-2-thiophene carboxamide derivatives as Abl inhibitors identified by a pharmacophore-based database screening of commercially available compounds. Bioorganic & medicinal chemistry

letters. 2008;18(15):4328-31. doi: 10.1016/j.bmcl.2008.06.082. PubMed PMID: 18621522.
87. Stanford SM, Krishnamurthy D, Falk MD, Messina R, Debnath B, Li S, et al. Discovery of a novel series of inhibitors of lymphoid tyrosine phosphatase with activity in human T cells. Journal of medicinal chemistry. 2011;54(6):1640-54. doi: 10.1021/jm101202j. PubMed PMID: 21341673; PubMed Central PMCID: PMCPMC3086468.
88. Hu W, Dong H, Li YZ, Hu XT, Han GJ, Qu YB. A high-throughput model for screening anti-tumor agents capable of promoting polymerization of tubulin in vitro. Acta Pharmacol Sin. 2004;25(6):775-82. PubMed PMID: 15169631.
89. Nguyen TL, McGrath C, Hermone AR, Burnett JC, Zaharevitz DW, Day BW, et al. A common pharmacophore for a diverse set of colchicine site inhibitors using a structure-based approach. Journal of medicinal chemistry. 2005;48(19):6107-16. doi: 10.1021/jm050502t. PubMed PMID: 16162011.
90. Li CM, Lu Y, Ahn S, Narayanan R, Miller DD, Dalton JT. Competitive mass spectrometry binding assay for characterization of three binding sites of tubulin. J Mass Spectrom. 2010;45(10):1160-6. doi: 10.1002/jms.1804. PubMed PMID: 20814887.
91. Jenkins JL, Glick M, Davies JW. A 3D similarity method for scaffold hopping from known drugs or natural ligands to new chemotypes. Journal of medicinal chemistry. 2004;47(25):6144-59. doi: 10.1021/jm049654z. PubMed PMID: 15566286.
92. Liu X, Ouyang S, Yu B, Liu Y, Huang K, Gong J, et al. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. Nucleic acids research. 2010;38(Web Server issue):W609-14. doi: 10.1093/nar/gkq300. PubMed PMID: 20430828; PubMed Central PMCID: PMCPMC2896160.
93. Hawkins PC, Skillman AG, Nicholls A. Comparison of shape-matching and docking as virtual screening tools. Journal of medicinal chemistry. 2007;50(1):74-82. doi: 10.1021/jm0603365. PubMed PMID: 17201411.
94. AbdulHameed MD, Chaudhury S, Singh N, Sun H, Wallqvist A, Tawa GJ. Exploring polypharmacology using a ROCS-based target fishing approach. Journal of chemical information and modeling. 2012;52(2):492-505. doi: 10.1021/ci2003544. PubMed PMID: 22196353.
95. Bemis GW, Kuntz ID. A fast and efficient method for 2D and 3D molecular shape description. Journal of computer-aided molecular design. 1992;6(6):607-28. PubMed PMID: 1291629.
96. Heifets A, Lilien RH. LigAlign: flexible ligand-based active site alignment and analysis. Journal of molecular graphics & modelling. 2010;29(1):93-101. doi: 10.1016/j.jmgm.2010.05.005. PubMed PMID: 20713281.
97. Taminau J, Thijs G, De Winter H. Pharao: pharmacophore alignment and optimization. Journal of molecular graphics & modelling. 2008;27(2):161-9. doi: 10.1016/j.jmgm.2008.04.003. PubMed PMID: 18485770.
98. Cortes-Cabrera A, Morris GM, Finn PW, Morreale A, Gago F. Comparison of ultra-fast 2D and 3D ligand and target descriptors for side effect prediction and network analysis in polypharmacology. Br J Pharmacol. 2013;170(3):557-67. doi: 10.1111/bph.12294. PubMed PMID: 23826885; PubMed Central PMCID: PMCPMC3791994.
99. Horwitz SB. Taxol (paclitaxel): mechanisms of action. Ann Oncol. 1994;5 Suppl 6:S3-6. PubMed PMID: 7865431.
100. Konno T, Watanabe J, Ishihara K. Enhanced solubility of paclitaxel using water-soluble and biocompatible 2-methacryloyloxyethyl phosphorylcholine polymers. Journal of biomedical materials research Part A. 2003;65(2):209-14. doi: 10.1002/jbm.a.10481. PubMed PMID:

12734814.
101. Kingston DG, Bane S, Snyder JP. The taxol pharmacophore and the T-taxol bridging principle. Cell Cycle. 2005;4(2):279-89. PubMed PMID: 15611640.