CRC Platform: A Colorectal Cancer Domain-specific Chemogenomics Knowledgebase for Polypharmacology and Target Identification Research

by

Xiaole Yang

Bachelor of Science, Liaoning University of TCM, 2012

Submitted to the Graduate Faculty of

School of Pharmacy in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

SCHOOL OF PHARMACY

This dissertation was presented

by

Xiaole Yang

It was defended on

March 27th, 2014

and approved by

Xiang-Qun (Sean) Xie, PhD, Professor

Lirong Wang, PhD, Research Assistant Professor

Vinayak Sant, PhD, Assistant Professor

Dissertation Advisor: Xiang-Qun (Sean) Xie, PhD, Professor

Copyright © by Xiaole Yang
2015

CRC Platform: A Colorectal Cancer Domain-specific Chemogenomics Knowledgebase for Polypharmacology and Target Identification Research

Xiaole Yang, M.S.

University of Pittsburgh, 2015

Colorectal cancer (CRC) is the third most common cancer, causing more than 600,000 deaths worldwide annually. Due to the involvement of complicated signaling pathways, epigenetic changes and genetic/genomic alterations, it is still challenging to develop effective treatments to reverse CRC progression. In order to facilitate developing new drugs for CRC treatment and revealing the mechanisms of CRC drug action at molecular level, we have constructed a computational CRC Platform (http://www.cbligand.org/CRC/), a domain-specific chemogenomics knowledgebase.

The CRC platform consists of four database modules, e.g. 762 CRC related genes and proteins, 411 known CRC drugs and chemicals, 168383 CRC related bioassays, and 269 CRC pathways, as well as searching tools for multi-function retrieval. It is also featured with powerful cloud computation technologies and computational tools to expedite target identification, polypharmacology and drug synergy analysis for CRC research.

We have also demonstrated the application of the CRC platform in the case studies: (1) computational exploration of FDA-approved CRC drugs for polypharmacology and drug

synergy analysis; (2) *in silico* target identification of small chemical molecules from natural products with anti-CRC bioactivity; and (3) target identification and experimental validation for our in-house compounds. CRC platform will not only enrich our knowledge of CRC target identification, polypharmacology analysis, and biomarkers investigation, but also enhance the CRC chemogenomics data sharing and information exchange globally, and assist new drug design discovery and development for CRC treatment.

Keywords: colorectal cancer (CRC); chemogenomics database; cloud computation; target identification; polypharmacology; natural product; drug discovery.

TABLE OF CONTENTS

PREFACE..XI

1.0 INTRODUCTION..1

1.1 COLORECTAL CANCER..1

1.2 RISKS AND PREVENTIVE FACTORS..2

1.3 GENOMICS AND RELATED PATHWAYS OF COLORECTAL CANCER..3

1.3.1 Genetic pathways of colorectal cancer..3

1.3.2 Inherited forms of CRC genomics..6

1.4 EXAMPLE OF CRC RELATED TARGET..6

1.5 ANTICANCER DRUG REPURPOSE..8

1.6 POLYPHARMACOLOGY AND DRUG DISCOVERY..11

2.0 MATERIALS AND METHODS..13

2.1 COLORECTAL CANCER DOMAIN SPECIFIC KNOWLEDGEBASE (CRC-KB).....13

2.2 DATABASE INSFRASTRUCTURE AND WEB INTERFACE..14

2.2.1 Web-interface..14

2.2.2 Data collection and contents..14

2.3 POLYPHARMCOLOGY ANALYSIS TOOLS..15

2.3.1 HTDocking..16

2.3.2 TargetHunter..17

2.4 EXPERIMENT..17

2.4.1 Cell culture and treatment..17

2.4.2 Cell growth and analysis of apoptosis 18

2.4.3 Antibodies and western blotting 18

3.0 RESULTS 20

3.1 COLORECTAL CANCER RELATED TARGETS, PATHWAYS, AND DRUGS 20

3.2 NATURAL PRODUCTS FOR COLORECTAL CANCER 26

3.2.1 New targets identification for curvularin 31

3.2.2 New targets identification for aloe emodin 34

3.2.3 New target identification for kaempferol 39

3.3 POLYPHARMACOLOGY ANALYSIS OF ANTI-CRC DRUGS AND THEIR TARGETS 42

3.4 VIRTUAL SCREENING AND EXPERIMENTAL VALIDATION 47

3.4.1 Virtual screening of in-house compounds 47

3.4.2 Experimental validation of in-house compounds 49

4.0 DISCUSSION 52

5.0 SUMMARY **57**

6.0 FUTURE SPECULATIONS 58

APPENDIX ABBREVIATIONS 60

BIBLIOGRAPHY 62

LIST OF FIGURES

Figure 1. The classical model of CRC development from polyp to cancer. 5

Figure 2. IGF1R targeted strategies... 7

Figure 3. Summary of existing drug-repurposing methods.. 10

Figure 4. The Polypharmacology wide target networks... 11

Figure 5. The homepage of CRCPlatform.. 20

Figure 6. The classification of CRC targets and anti-CRC drugs collected in CRCPlatform. 21

Figure 7. Statistical analysis of CRC drugs and their targets in different developing clinical trial phases. 23

Figure 8. The top list pathways with more than 20 CRC related targets involved................. 25

Figure 9. Overview of targets predictions for curvularin by TargetHunter............................ 32

Figure 10. Docking results of CHEMBL4802050 (A) and Curvularin(B).. 33

Figure 11. Crosstalk between Wnt and Prostaglandin signalling... 36

Figure 12. The potential compounds from NCI database are screened for Tdp1 based on chemical similarity. 39

Figure 13. Docking analysis of Kaempferol and hTRPV1 protein model. 41

Figure 14. Polypharmacology analysis of five FDA approved anti-CRC drugs.................... 43

Figure 15. The detailed information of docking analysis between erlotinib and CDK2......... 47

Figure 16. The detailed docking analysis of XIE62-1032 in IGF1R (PDB: 3NW6). 49

Figure 17. HCT116 cells were treated with indicated compounds for 72 h. Cell viability was determined by MTS assay. Results were expressed as means ± S.D. of three independent experiments. 50

Figure 18. HCT116 cells were treated with XIE62-1032 at different concentrations for 72 h. Cell viability was determined by MTS assay. Results were expressed as means ± S.D. of three independent experiments. IC50 was calculated by Prism. 50

Figure 19. Cells were treated with XIE62-1032 and XIE62-1044 for 24h at indicated concentrations. The indicated proteins were analyzed by western blotting. β-actin is a loading control. 51

Figure 20. The proposed mechanism of anti-CRC activity of our in-house compound XIE62-1032. 56

LIST OF TABLES

Table 1 Summary of risks and preventive factors of colorectal cancer 3

Table 2. Summary of non-cancer drugs repurposed as anti-CRC drugs 9

Table 3. Detailed information of KEGGID and pathway names 26

Table 4. Compounds from natural products with anti-CRC activities 27

Table 5. Potential collaborators found by Bioassay Geomap function of TargetHunter 37

Table 6. The comparison of the experimental pKi or pKd data and the predicted pKd values for the FDA-approved CRC drugs 44

Table 7. Verification of other predicted targets of approved CRC drugs by experiment....... 45

Table 8. Top seven in-house compounds with docking score above 6 in complex with IGF1R 48

Table 9. The list of potential collaborators for XIE62-1032 and IGF1R binding assay 55

PREFACE

I would like to express my special appreciation and gratitude to my advisor Dr. Xiang-Qun (Sean) Xie, you have been a tremendous mentor for me. I would like to thank you for accepting me as a member of this excellent research group and encouraging my research. Your advice on both research as well as on my career have been priceless. Your continually and persuasively conveyed a spirit of hard working and independence in regard to research and an excitement in regard to teaching, which inspired me a lot.

I would like to especially thank you my co-advisor Dr. Lirong Wang, who has shown a large and consistent interest in my project during the times. Our numerous scientific discussions and his many constructive comments have greatly improved this work. I consider it a great honor to work with Dr. Wang. Also, I would like to take the opportunity to say thank you to all the committee members: Dr. Xiang-Qun Xie, Dr. Lirong Wang, and Dr. Vinayak Sant.

Furthermore, I would like to thank Dr. Yu's lab from cancer institute, school of medicine, University of Pittsburgh. Thank you very much for helping us do the experimental validations: MTS assay and apoptosis analysis for our predications and helping us revising the experimental procedures, results and discussion part. I appreciated that your input efforts and your precious suggestions for the manuscript.

I would like to specially thank Shifan Ma who gave me a lot of suggestions when I came to this lab last year. What's more, Shifan Ma's great contributions to the construction of

CRCPlatform and the literature validations for the FDA-approved drugs are very appreciated. It is very pleasant to work with her to do research.

I would also like to thank all of my lab members who supported me in writing, and incented me to strive towards my goal. Thank you for Dr. Zhiwei Feng providing the hTRPV1 protein model and teaching me how to do the virtual screening, manual docking, pharmacophores etc. with SYBYL. Thank Dr. Peng Yang for teaching us how to identify a good molecule for potential drugs.

At the end, I would like to express appreciation to my mother, father and brother who supported me to further study here. Also, my boyfriend Joel Oller always supports me and encourages me to do my best. Thank you very much.

1.0 INTRODUCTION

1.1 COLORECTAL CANCER

Colorectal cancer (CRC) is the third most common aggressive cancer causing more than 600,000 deaths worldwide annually and incidence rate is increasing in developing countries [1]. According to the report, approximately 1 in 20 Americans will be diagnosed with colon or rectum cancer in their lifetime [2]. In 2014, the American Cancer Society estimated that 136,830 individuals were newly diagnosed with colorectal cancer and 50,310 colorectal cancer deaths in United States [3]. Furthermore, the incidence and mortality rates range from 30% to 40% higher among males than females overall. As usual, the incidence and death rates for colorectal cancer increase with age. The median age of male patients diagnosed of colon cancer is 69 and for female patients is 73, which is older than the median age of patients diagnosed of rectal cancer, 63 in men and 65 in women [4]. However, the rates increased by 1.1% per year among men and women aged younger than 50 years since 2002. This trend was confined to tumors in the distal colon (1.3% annually) and rectum (1.8% annually) [3].

1.2 RISKS AND PREVENTIVE FACTORS

CRC is also known as colon cancer, rectal cancer or bowel cancer. Signs and symptoms of CRC may include a change in bowel habits, such as diarrhea, constipation, rectal bleeding, blood in the stool, weakness and fatigue, unexpected weight loss [5]. Many risk factors are associated to the development of CRC like age, male, family history, inflammatory bowel disease, obesity, diabetes, tobacco smoking, alcohol, and diet style such as high intake of red and processed meat less consumption of fiber (**Figure 1**) [1]. However, the attributable effects of diet on molecular subtypes of CRC are not yet completely clear [6]. People with the family history of CRC or with inflammatory bowel disease are more likely to develop CRC. According to the literature, infection with *Helibobacter pylori* and other potential infectious agents contribute to the increased risk of CRC [7]. Additionally, there are a number of case-control studies that have shown different gastrointestinal (GI) microbial compositions in normal subjects versus adenoma- and CRC-affected patients [8]. Hence, it is reasonable to suggest that there is a potential association between compositions of GI microbiota and CRC epidemiology.

It is estimated that the number of CRC cases will continue to increase dramatically from 1.2 to 2.2 million worldwide; the increased cases from developing countries are accounting for 62% over the next two decades [9]. It is crucial to take some preventative measures to control CRC before the CRC rate increased. Population-based screening is to detect the CRC at early stage like colorectal neoplasms, amendable for curative treatment, and it is shown to be an efficient and cost-effective method [10]. Meanwhile, it has been proposed that consumption of unrefined grains, fish, and legumes as sources of protein, could lower CRC risks. As mentioned before, obesity, diabetes, physical inactivity, excessive alcohol intake

and smoking are all associated to increasing the risk. Therefore, it is a primary strategy to control these factors at population level.

Table 1 Summary of risks and preventive factors of colorectal cancer

Risk factors	Preventative factors
Family history +++	Large bowel endoscopy
Inflammatory bowel disease ++	Hormone replacement therapy
Diabetes +	Aspirin
Helicobacter pylori infection+	Statins
Other infections +	Fruit and vegetables
Smoking +	Cereal fiber and whole grain
Excessive alcohol consumption+	Fish
Obesity+	Dairy products
Sedentary+	Physical activity
High consumption of red and processed meat+	

Note: +++ means very strong risk, + + means strong risk, + means moderate risk.

1.3 GENOMICS AND RELATED PATHWAYS OF COLORECTAL CANCER

1.3.1 Genetic pathways of colorectal cancer

Colorectal cancer contributed to cancer mortality and morbidity, and its distinction between the colon and the rectum is largely anatomical [11]. The molecular mechanisms of colorectal cancer are clinically important because they are related to the disease prognosis and potential treatment development [12]. Researches on the molecular pathogenesis and therapy response for colorectal cancer increase greatly during the past decades years including the identification of the molecular mechanisms, and genetic changes causing the hereditary forms of colorectal cancer [13]. Different molecular alterations contribute to the apparent heterogeneity of early-onset CRC and the subgroups can be distinguished with distinct histopathology and familial characteristics [14].

It is important to know that most of the CRC cases (70-85%) are sporadic and the patients without identifiable genetic risk factors [15]. The development of CRC from begin to malignant lesions is usually more than 10 years, and dysplastic adenomas are the most common form of premalignant precursor lesions [16]. The classical model of colorectal carcinogenesis for the adenoma-carcinoma sequence has been evolving since its original formulation, and it links genetic alterations and their order of introduction, to different stages in tumor development [17]. This model was reported to involve adenomatous polyposis coli (APC) gatekeeping mutations, which provides a selective growth advantage to a normal epithelial cell, allowing it to outgrow the cells that surround it and become a microscopic clone (**Figure 1**) [18]. The APC gene mutation will contribute to the slow growth of small adenoma representing the target lesion for prevention and intervention of colorectal cancer [19]. Alternatively, not only APC gene but also KRAS gene will lead to uncontrollable new clonal growth with the expansion of cell number. Interestingly, the number of cells with only APC gene mutation is smaller than with both mutations. With the expansion of cells, the mutations in genes increase like BRAF, PTEN, BAX, SMAD4/TGF-β and TP53. In the end, they generate a malignant tumor invading through the membrane and metastasize to lymph nodes [20]. These "driver" gene mutations are with chromosomal instability, i.e., changes in number of chromosomes and structural changes of the chromosomes. This classical model of adenoma-carcinoma sequence is also named Chromosomal Instability (CIN) pathway.

In addition to this CIN pathway, around 10-15% of sporadic CRC cases are associated to the Microsatellite Instability (MSI) with the condition of genetic hyper mutability that results from impaired DNA mismatch repair (MMR). Cells with abnormally functioning MMR are more likely to accumulate mutations (insertion or deletions) in microsatellites located in DNA coding regions, generating frame shift mutations and subsequently leading to sporadic CRCs [21]. Another genetic pathway is serrated pathway, which is different from

the CIN and MSI pathways initiated through classical APC mutations. The name of this pathway is attributed to the morphologically serrated appearance of the precursor lesions highlighted by the presence of BRAF (protein kinase B-Raf) mutation and epigenetic silencing of genes involved in cell differentiation, DNA repair, and cell-cycle control [22, 23]. Most of the sporadic CRC cases with a lot of mutations combined these above-mentioned genetic pathways. Therefore, these pathways will cross talk with each other and modify these routes to the carcinogenesis, which will make the cancer development more complicated.

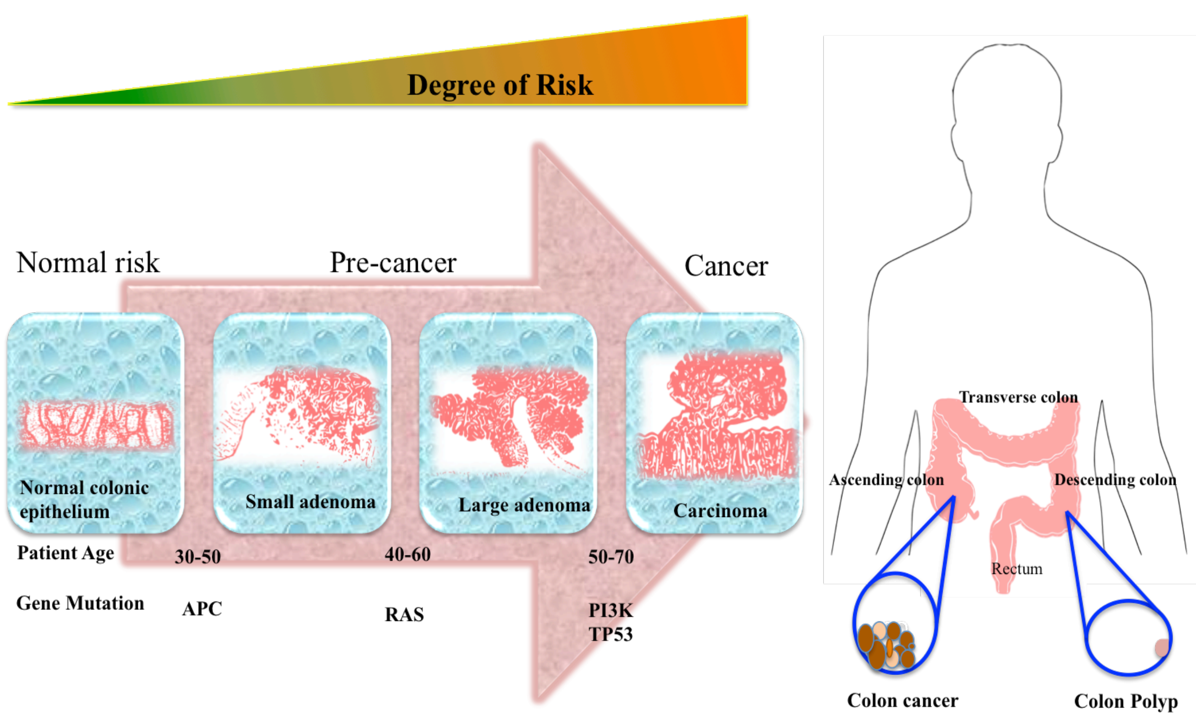


Figure 1. The classical model of CRC development from polyp to cancer. The APC gene mutation will contribute to the slow growth of small adenoma representing the target lesion for prevention and intervention of colorectal cancer. With the expansion of cells, the mutations in genes increase like BRAF, PTEN, BAX, SMAD4/TGF-βT, and P53. In the end, they generate a malignant tumor invading through the membrane and metastasize to lymph nodes. The green color indicates low risk and the orange indicates high risk of colorectal cancer.

1.3.2 Inherited forms of CRC genomics

It is estimated that between 3-6% of all CRC cases are associated with highly penetrant hereditary GI cancer syndromes and another 25-30% of individuals with CRC report having one or more relatives diagnosed with CRC [24]. The two most frequent appeared forms of hereditary colorectal cancers are hereditary non-polyposis colon cancer (Lynch syndrome, estimated allele frequency 1:350 to 1:1,700) and familial adenomatous polyposis coli (estimated allele frequency 1:10,000) [25]. Lynch syndrome results from germ line mutations in a class of genes involved in DNA mismatch repair, including MSH2, MLH, MSH6, and PMS2 [26]. The second most common hereditary CRC syndrome is familial adenomatous polyposis (FAP). However, only a small fraction of all CRCs are associated to FAP caused by germ line mutations in APC gene, which encodes a tumor-suppressor protein that is part of the WNT signaling pathway [27].

1.4 EXAMPLE of CRC RELATED TARGET

The insulin-like growth factor-1 receptor (IGF1R) is a transmembrane receptor tyrosine kinase (RTK) that transduces IGF1 and IGF2 signals which play an important role in growth, differentiation and survival of cells [28]. Two extracellular α-subunits form disulphide-bonds to two transmembrane-spanning β-subunits with cytoplasmic tyrosine kinase activity. Binding of IGF1 or IGF2 to the ectodomain of IGF1R induces a structural rearrangement resulting in phosphorylation of specific tyrosine residues in the cytoplasmic domains, stimulating catalytic (tyrosine kinase) activity and generating recruitment sites of Insulin

Receptor substrate proteins and adapter protein Src homology 2 domain containing (Shc), among other signaling proteins. Phosphorylation of these substrates by IGF1R leads to activation of mitogen-activated protein kinase (MAPK) and phosphoinositide 3'-kinase (PI3K) signaling cascades [29]. The over expression of the IGF1R has been implicated in different types of tumor systems including CRC [30]. A correlation between the IGF1/2 expression levels and tumor progression has been consistently documented and extensively studied with various approaches to down-regulate the IGF1R pathway. These target approaches include a reduction of ligand levels of bioactivity and an inhibition of receptor function using receptor-specific antibodies or small-molecule tyrosine kinase inhibitors [31] (**Figure 2**).

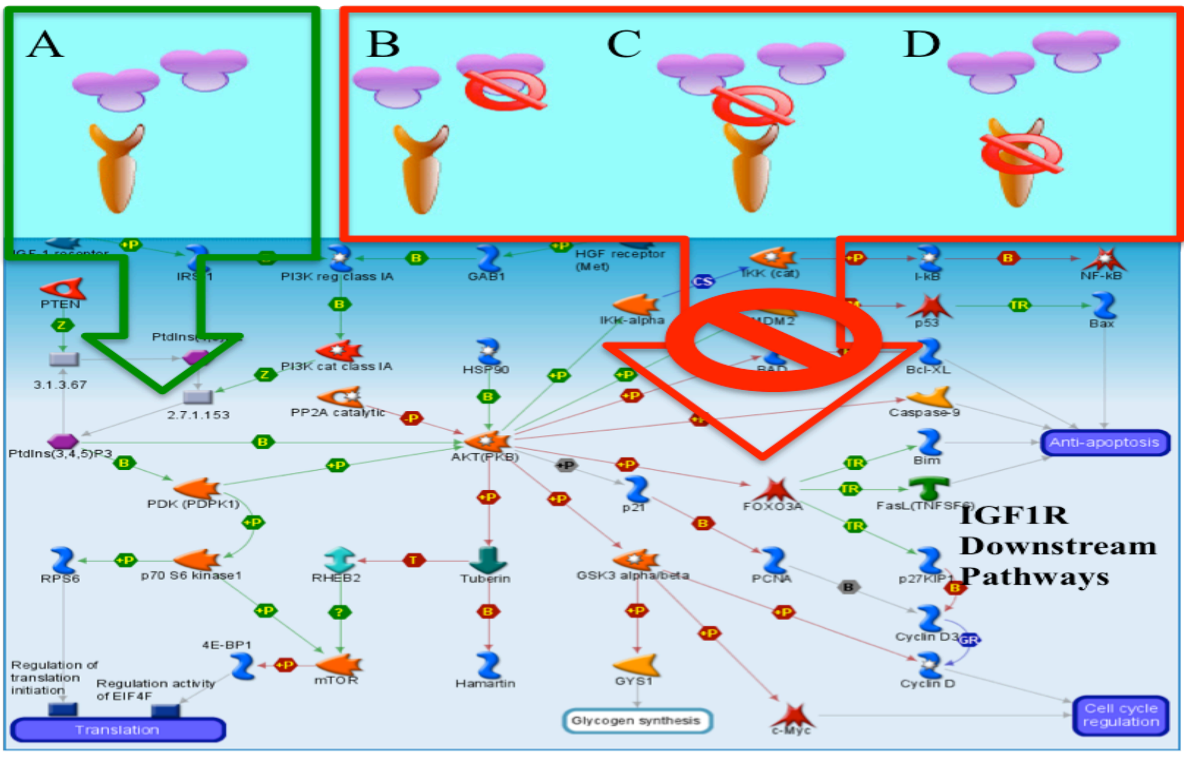


Figure 2. IGF1R targeted strategies. (A) The binding of insulin to its receptor triggers a diversity of downstream pathways (Green arrow: inducing downstream pathways). (B) Anti-ligand approaches including lower ligand concentration or the use of ligand-specific antibodies. (C) Anti-receptor approach suggests that most receptor-specific antibodies

effectively block the IGF1R as well as IGF1R hybrids but don't act on insulin receptors. (D) Receptor inhibitor such as tyrosine kinase inhibitors (Red arrow: inhibiting downstream pathways).

Recently, many small molecule inhibitors and antibodies specifically target the IGF1R are in preclinical and early clinical development such as CP-751, 871 from Pfizer clinical trial phase I-III [32]. It is estimated that at least 10 different drug candidates are being evaluated in clinical trials and early results have justified expansion of clinical trial programs [31] These small molecules inhibitors have high selectivity and orally active. A recent study summarized that IGF-1R inhibitor PQIP disrupted abnormal IGF1R signaling in GEO human colon cancer cells resulting in the induction of cell death. Studies with orthotopic colon carcinoma animal models in vivo suggested antitumor activity without significant weight loss and toxicity [33]. Therefore, based on various therapeutic strategies that target IGF1R has demonstrated impressive antineoplastic activity in laboratory models and clinical trials, IGF1R has the potential to serve as a target for colorectal cancer therapy.

1.5 ANTICANCER DRUG REPURPOSE

Drug development requires an average of 13 years of research and an investment of US$1.8 billion to bring a single drug from the bench to a patient's bedside [34]. Researchers and clinicians have adopted numerous strategies to reduce the cost of the drug discovery. One of the strategies is to find potential new use in established non-cancer drugs already approved and demonstrate an acceptable level of safety and tolerability. "Drug repurposing" or "repositioning" refers to the use of an old drug for a new indication. The major advantage of this approach is that their pharmacokinetic pharmacodynamics and toxicity profiles are

generally well known.

There are some strategies to effectively identify and implement current non-cancer drugs for cancer-related treatment [35]. The first idea is that almost all drugs used in human therapy will produce off-target side effects in addition to their original indications. The second idea is based on the finding that many different diseases share common molecular pathways and targets in the cell. Therefore, it is highly possible that the same drug could be therapeutic for more than one disease. Here are some examples of non-cancer drugs are successfully repurposed as anti-cancer drugs for colorectal cancer (**Table 2**)

Table 2. Summary of non-cancer drugs repurposed as anti-CRC drug

Drug	Original indication	New anticancer indications	Reference
Aspirin	Analgesic, antipyretic	CRC	[1]
Celecoxib	Osteoarthritis, rheumatoid arthritis	CRC, lung cancer	[2]
Metformin	Diabetes mellitus	Breast, adenocarcinoma, prostate cancer, CRC	[3]
Rapamycin	Immunosuppressant	CRC, lymphoma, leukemia	[4]
Noscapine	Antitussive, antimalarial, analgesic	Multiple cancer types	[5]
Troglitazone	Diabetes mellitus	Metastatic colorectal cancer	[6]
Enilconazole	Antifungal	Metastatic colorectal cancer	[7]
Citalopram	Depression	Metastatic colorectal cancer	[7]
Mebendazole	Helmintic	Colon cancer	[8]

In recent years, the number of drug-repositioning methods has increased greatly. It is of great importance to better understand existing methods and prioritize them based on specific studies. Transcriptomic approaches can relate a drug to an expression-based phenotype. For example, the CMap approach was one of the first attempts to take a more holistic view of these transcriptomic data and apply them to link expression profiles across conditions [36].

Genetically, genome-wide association studies (GWAS) have shown the association between genetic variants and polygenic diseases, resulting in the identification of genes proximal to these variants being linked to numerous complex diseases. There are some challenges by using GWAS data [37]. Pathway- or network-based drug-repurposing methods utilize disease omics data, available signaling or metabolic pathways, which are still the key ways for the repositioned drugs [38]. From a methodological point of view, the computational drug-repurposing methods can be classified into target-based, knowledge-based, signature-based, pathway-based, and targeted-mechanism-based methods. The following figure shows a summary of the repurposing methods (**Figure 3**).

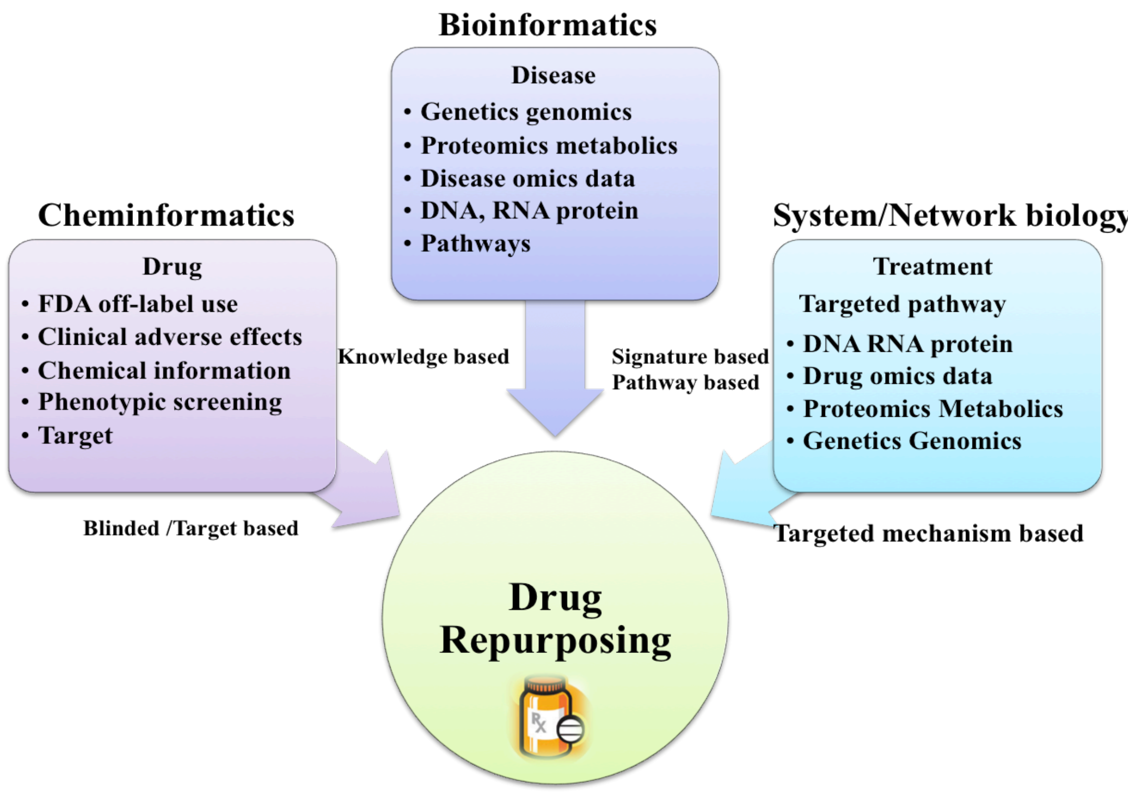


Figure 3. Summary of existing drug-repurposing methods. The computational drug-repurposing methods can be classified into target-based, knowledge-based, signature-based

pathway-based, and targeted-mechanism-based methods. Different disciplines are also integrated to do the drug repurposing research such as Cheminformatics, Bioinformatics, System or Network biology.

1.6 POLYPHARMACOLOGY AND DRUG DISCOVERY

Recently, designing a single molecule able to simultaneously and specifically interact with multiple targets is gaining more interests in drug discovery, which is referred as "polypharmacology" differed from combination therapy [39, 40]. Current research focuses on two aspects of polypharmacology: (1) unintended polypharmacology can lead to adverse effects; and (2) polypharmacology across several disease-relevant targets can enhance therapeutic efficacy, prevent drug resistance, or reduce therapeutic-target-related adverse effects [41]. Interestingly, it is by now generally recognized that several approved drugs elicit their therapeutic effect through complex polypharmacology [42]. **Figure 4** is well illustrated the concept of polypharmacology, which is already published by our lab in 2014 [43].

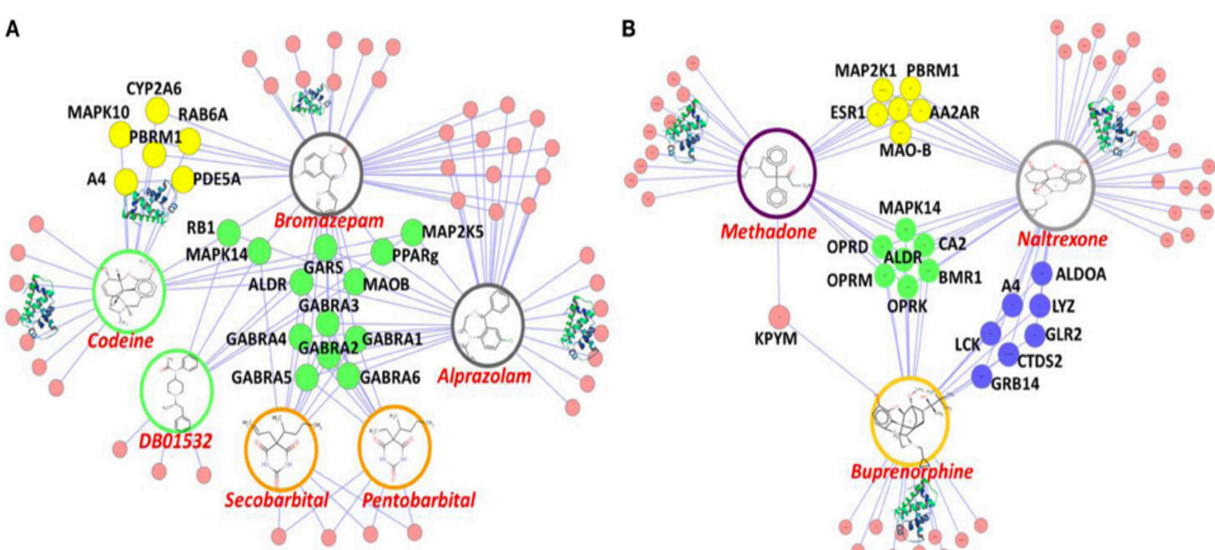


Figure 4. The Polypharmacology wide target networks. (A) The predicted targets of six known abused/approved drugs (*opioids*: codeine and DB01532; *benzodiazepines*: bromazepam and alprazolam; *barbiturates*: secobarbital and pentobarbital). (B) The predicted

targets of 3 approved drugs for Drug Abuse treatments (methadone, naltrexone, and buprenorphine). This figure was from our lab previously published paper [43].

Polypharmacology is largely relevant for diseases associated with wide target networks and cellular pathways such as neurodegenerative diseases and different types of cancer with complicated cellular pathways. The excessive proliferation and survival of cancer cells can be sustained by deregulation in the expression or activity of different proteins. Tumors secrete vascular endothelial growth factor (VEGF) and platelet-derived growth factors triggering angiogenesis to form endothelial cells and pericytes. Also, activated fibroblasts also secrete enzymes, which degrade the extracellular matrix and make room for the tumor growth. Today, many effective kinase targeted therapies for cancer treatment can hit on the multiple sensitive target nodes, which illustrate the observed efficacy of marketed kinase inhibitors in the treatment of several forms of cancer. Simultaneous inhibition of multiple kinases is now considered as a promising therapeutic strategy. In addition, polypharmacological anticancer drugs are also believed to prevent anti-cancer drug resistance [44].

The major challenge of polypharmacology is the ability to rationally design multi-target ligands. In addition, further methodological advancements are of great importance. To meet this demand, dedicated efforts are made by integrated approaches involving medicinal chemistry, genetics, chemical biology, and computational chemistry [45].

2.0 MATERIALS AND METHODS

2.1 COLORECTAL CANCER DOMAIN SPECIFIC KNOWLEDGEBASE (CRC-KB)

To meet the demands of CRC drug development, an integrated cloud computing server, CRCPlatform, has been constructed with a large collection of CRC relevant chemogenomics data, including genes, protein targets, and chemical molecules with their bioactivity records, bioassays and references as well as CRC drugs of FDA approved or in clinical trial. CRC Platform (http://cbligand.org/CRC) also provides powerful computational tools such as our established TargetHunter and HTDocking for new targets identification, drug repurposing, polypharmacology analysis associated with CRC. On the other hand, our group has already developed several disease domain specific knowledgebase, including a drug abuse knowledgebase (DA-KB) [43], an Alzheimer's disease knowledgebase (AlzPlatform: http://www.cbligand.org/AD) [46]. To summarize, it is of great significance to build the CRCPlatform for investigating CRC targets, small chemical molecules, which will facilitate the understanding the mechanism of polypharmacology and anti-CRC drug discovery.

2.2 DATABASE INSFRASTRUCTURE AND WEB INTERFACE

CRCPlatform was rooted from our established web-interface molecular database prototype CBID (www.CBLIgand.org/CBID) [47]. It is constructed with a MySQL (http://www.mysql.com) database and an apache (http://www.apache.org/) web server, and implemented with our developed chemogenomics tools.

2.2.1 Web-interface

An accessible and easy-to-use web interface was offered with effective and efficient search engine for the detailed data on CRC, written in PHP language (http://www.php.net/). CRCPlatform provides two types of structure query function for retrieval of bioactivity information: substructure and similarity. Open Babel [48] is utilized as the structural searching engine and JME [49] supplies the input interface for search. Users can either draw a chemical structure, or upload and submit a file containing the structure of a small molecule in the format of SMILES, sdf, mol, or cdx. After submission, the search will be performed automatically at the server side with GPU accelerating computing approach, and the results can be retrieved and shown to the user side on a new page, containing structure of compound, target name, and the corresponding reference link.

2.2.2 Data collection and contents

We have data-mined CRC-related genes, proteins, pathways, compounds with bioassays from public databases and literatures, and have integrated them into our CRCPlatform. The current version of CRCPlatform is comprised of the following features

and we will keep updating our information for CRC, and involve additional sections of data in the future.

CRC genes/proteins. We collected proteins as potential targets for CRC based on the target information of CRC drugs that are FDA-approved or currently in the clinical trials. These targets information were collected from literature, patents, and some public databases, such as DrugBank [50] (www.drugbank.ca/), ClinicalTrials.gov (clinicaltrials.gov/), PubChem [51] (pubchem.ncbi.nlm.nih.gov/), PubMed (www.ncbi.nlm.nih.gov/pubmed), Cancer.gov [52] (www.cancer.gov/) and NCI (National Cancer Institute). These CRC related target genes and proteins were then mapped to UniportKB proteins and genes.

CRC related Pathways. The CRC related pathways were achieved via the publicly available database KEGG [53, 54] (http://www.genome.jp/kegg/).

Chemicals of CRC target proteins. Besides CRC drugs of FDA approved or in clinical trials, chemicals that directly interact with the CRC targets were also archived in our database. The compounds were from journal articles and archived in CHEMBL [55] (www.ebi.ac.uk/chembl/). The corresponding bioassays to validate the interactions between these molecules and the target proteins were also collected.

2.3 POLYPHARMCOLOGY ANALYSIS TOOLS

At present, one of the challenges is how to identify bonafide and viable targets and efficiently exploit these for the development of selective nontoxic cancer therapies, to overcome the major drawbacks of conventional cytotoxic cancer chemotherapy [56]. On the other hand, small molecules with desired properties are considered as top priorities for CRC drug design. The CRCPlatform provides chemogenomics and cheminformatics data to

explore the potential CRC targets or off-targets, ADME (absorption, distribution, metabolism, excretion), and toxicity prediction (http://www.cbligand.org/Tox), as well as molecular properties, the properties explorer (http://www.cbligand.org/OP/Property_Explorer.php) and drug-likeness calculations. The chemogenomics tools based on state-of-the-art machine learning algorithms developed by our group have been implemented on the CRCPlatform to assist CRC drug design and target identification as showed below.

2.3.1 HTDocking

We constructed online high-throughput docking program (HTDocking, http://www.cbligand.org/CRC/docking_search.php) in our CRC*Platform*. HTDocking online program aims to explore multiple druggable protein targets and small molecule interactions and their potential pharmacology. In our HTDocking for CRC, files for the structures of CRC target proteins were downloaded from Protein Data Bank. AutoDock Vina is utilized in HTDocking program, which offers a multi-facet capability, high performance rate and enhanced accuracy to facilitate effortless usage [57]. It can provide 3-5 predicted binding affinity values (ΔG values) from different poses for each compound in a binding pocket of a protein [46]. The best binding affinity value is considered as docking score. The calculation of docking score is pKi=-log (predicted Ki) where the predicted $Ki=exp^{(\Delta G*1000/(1.987191*298.15))}$. According to the docking score of a queried compound from each protein structure, we can rank the potential CRC targets. A higher docking score means a good binding affinity indicating that the protein could be a candidate target for a small molecule.

2.3.2 TargetHunter

A web-interfaced target identification program, TargetHunter (http://cbligand.org/TargetHunter), was built in CRC*Platform* to predict the biotargets or off-targets of compounds [58]. TargetHunter predicts target for small molecules by powerful data-mining algorithm (TAMOSIC), which is based on the principle that compounds with structural similarities have similar physicochemical properties and potentially similar biological profiles. What's more, Tanimoto threshold as an important parameter of TAMOSIC can exclude irrelevant targets [59]. There are five prominent features of TargetHunter: (a) User-friendly interface; (b) History data retrieval; (c) Multiple options of specific databases and fingerprints; (d) High accuracy; (e) Bioassay finder or Bioassay GeoMap function to easily find the potential collaborators who may already have the bioassays established for predicted targets validation.

2.4 EXPERIMENT

2.4.1 Cell culture and treatment

Human colorectal cancer cell line HCT116 was obtained from American Type Culture Collection (Manassas, VA, USA). Cells were cultured in McCoy's 5A modified medium (Invitrogen, Carlsbad, CA, USA) supplemented with 10% defined fetal bovine serum (Hyclone, Logan, UT, USA), 100 units/mL penicillin, and 100µg/mL streptomycin (Invitrogen). Cells were maintained in a 37 °C incubator at 5% CO2.

Cells were plated in 12-well plates at 20% to 30% density 24 hours before treatment. The DMSO (cat# D2650, Sigma, St Louis, MO, USA) stocks of compounds were diluted to appropriate concentrations with the cell culture medium before adding to cells.

2.4.2 Cell growth and analysis of apoptosis

Following treatment, floating and adhering cells were collected at 72 hr. For analysis of apoptosis by nuclear staining, cells were resuspended and fixed in PBS solution containing 3.7% formaldehyde, 0.5% Nonidet P-40 and 10 μg/ml Hoechst 33258 (Molecular Probes). Apoptosis was assessed through microscopic visualization of condensed chromatin and micro nucleation as previously described [60]. A minimum of 300 cells was analyzed in triplicate.

One thousand HCT 116 cells/well were plated in 96-well plates 24 hours prior to treatment and treated for 72 hours. Cell proliferation in triplicates was measured using Cell-Titer 96 Aqueous One Solution Cell Proliferation Assay (G3581, Promega) according to manufacturer's recommendations. A490 nm was measured with a Victor III (Perkin-Elmer/Wallace) plate reader. Each experiment was done in triplicate and repeated at least twice. The values were normalized to that of the vehicle control and displayed with the means with one standard deviation (SD) [61].

2.4.3 Antibodies and western blotting

Cells were harvested in 2× Laemmli buffer (0.125M Tris-HCl at pH 6.8, 10% β-mercaptoethanol, 4% sodium dodecyl sulphate, 20% glycerol, 0.05% bromophenol blue) and then centrifuged at 13,000 RPM for 10 minutes. Proteins were resolved on 10% NuPAGE gels (Invitrogen). Gel electrophoeresis was carried out for 1 hour at 160v in MES buffer

(Invitrogen). Protein was then transferred to PVDF membranes using a TransBlot SD semi-dry transfer cell (Biorad, Hercules, CA). Membranes were blocked for nonspecific binding with 5% nonfat milk in TBS-T for one hour at room temperature and then incubated in primary antibody overnight at 4ºC. Primary antibodies cleaved caspase-3 (#9661) and cleaved caspase-9 (#9501) (Cell Signaling, Beverly, MA) were used at a 1:1000. Following primary antibody hybridization, membranes were washed with TBS-T and incubated in appropriate HRP-conjugated secondary for 1 hour at room temperature. Goat-anti-rabbit (31462) (Pierce, Rockford, IL) was used as the secondary antibodies. Presence of antibody binding was detected using Western Lighting - Plus ECL (Perkin Elmer, Waltham, MA) according to manufactures specifications. Membranes were then exposed on blue X-ray film (Phenix Research Products, Candler, NC) [62].

3.0 RESULTS

3.1 COLORECTAL CANCER RELATED TARGETS, PATHWAYS, AND DRUGS

CRCPlatform (www.cbligand.org/CRC) contains 1059 CRC related proteins, corresponding with 15 FDA-approved and 396 agents in clinical trials for CRC treatment, 271 CRC related pathways and CRC targets related bioassay studies. Figure 5 shows the detailed information of CRCPlatform.

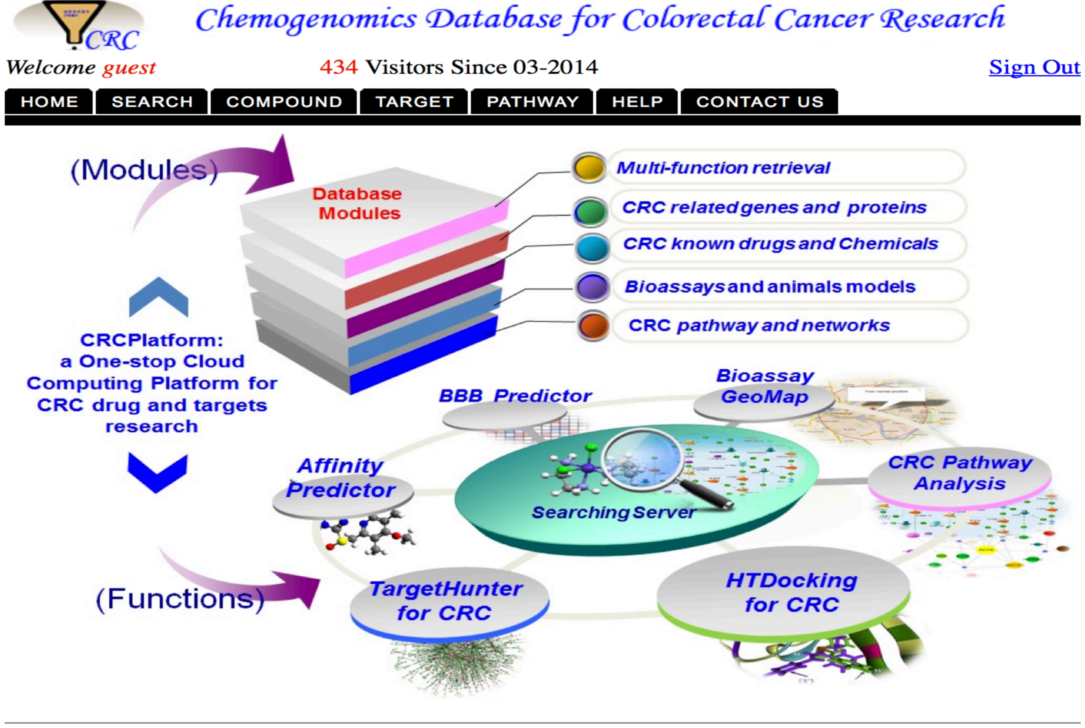


Figure 5. The homepage of CRCPlatform.

The majority of CRC target proteins are enzymes, such as cyclin-dependent kinase 2 (CDK2), Serine/threonine-protein kinase B-Raf (BRAF), CYP450 and Catechol O-methyltransferase (COMT) etc. Additionally, membrane receptors, such as Beta-2 adrenergic receptor (ADBR2), Endothelin-1 receptor (EDNRA), Ephrin type-A receptor 2 (EPHA2); ionic channels, such as Potassium voltage-gated channel subfamily D member 2 (KCND2), transporters like ATP-binding cassette sub-family G member 2 (ABCG2) are also included in the 1051 target proteins. In addition to 15 approved and 396 clinical trial CRC agents and their corresponding target genes and proteins, CRCPlatform also contains some target proteins from drugs for other types of cancers, which might have the potential to be repurposed for treatment of colorectal cancer (**Figure 5**).

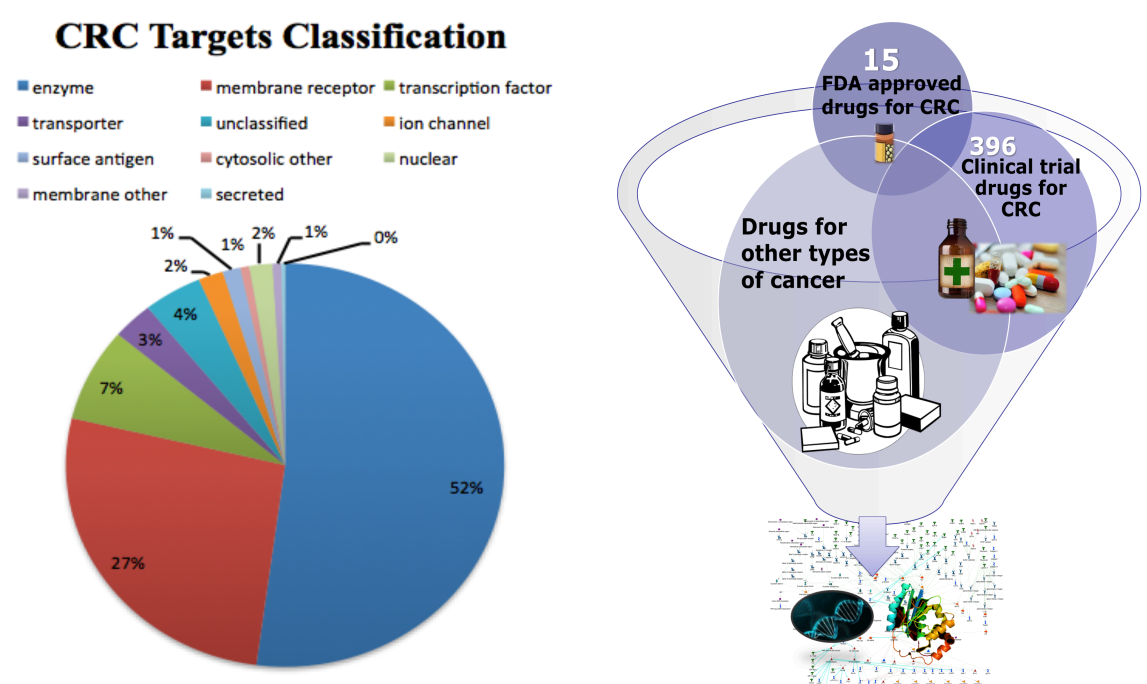
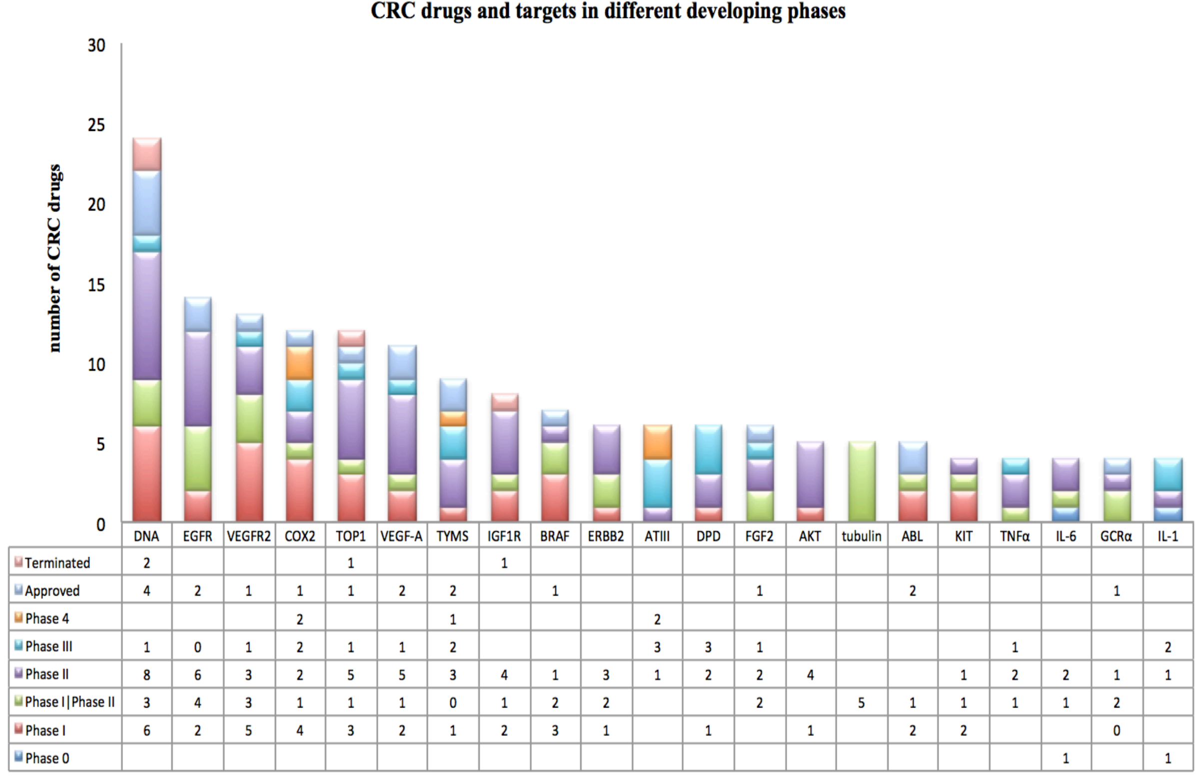


Figure 6. The classification of CRC targets and anti-CRC drugs collected in CRCPlatform. The CRC related targets were classified into enzyme, membrane receptor, transcription factor, transporter, ion channel, and surface antigen etc. The majority of these targets belong to enzymes (accounting for 52%). We mainly focused on the anti-CRC drugs approved by FDA and those in different phases of clinical trials.

By statistical analysis of the CRC drugs, the different development phases for top list 20 drugs were plotted referring to their corresponding targets. As shown in **Figure 6**, the DNA is the major target for CRC drugs on the market, four approved interact directly with DNA molecule to achieve their therapeutic effect, such as Fluorouracil, Capecitabine, Oxaliplatin, and Carboplatin. Other approved CRC drugs act on several other targets incorporated into angiogenesis, signaling and DNA synthesis, including Cetuximab and Panitumumab target EGFR, Aflibercept and Bevacizumab hit VEGF-A, Irinotecan Hydrochloride interacts with TOP1, Leucovorin is related with thymidylate synthase (TYMS), and Regorafenib acts on multiple tyrosine kinases like VEGFR2, BRAF, AKT and ABL.

These traditional targets for cancer chemotherapy relevant with DNA replication (DNA, TYMS, and TOP1 [63]) and angiogenesis (VEGF, VEGFR, EGFR, ERBB2) not only account for the major part of market drugs, but also pose the potential to develop new drugs for CRC, based on the large number of Phase 1, Phase 1|Phase 2, and Phase 2 drugs, and several Phase 3 drugs. Compare the ratio of different clinical phases for each targets, some targets like VEGFR (phase 1+phase1|2:phase 2: phase 3=8:3:1) have more potential in the future than some others like EGFR (phase 1+phase1|2:phase 2: phase 3=6:6:0) [64].



	DNA	EGFR	VEGFR2	COX2	TOP1	VEGF-A	TYMS	IGF1R	BRAF	ERBB2	ATIII	DPD	FGF2	AKT	tubulin	ABL	KIT	TNFα	IL-6	GCRα	IL-1
Terminated	2				1			1													
Approved	4	2	1	1	1	2	2		1				1			2				1	
Phase 4				2			1				2										
Phase III	1	0	1	2	1	1	2				3	3	1					1			2
Phase II	8	6	3	2	5	5	3	4	1	3	1	2	2	4			1	2	2	1	1
Phase I|Phase II	3	4	3	1	1	1	0	1	2	2			2		5	1	1	1	1	2	
Phase I	6	2	5	4	3	2	1	2	3	1		1		1		2	2			0	
Phase 0																			1		1

Figure 7. Statistical analysis of CRC drugs and their targets in different developing clinical trial phases.

Moreover, some relative new targets has raised more attention recently, such as the tyrosine kinases BRAF, KIT, ABL, and AKT [65], enzyme regulate inflammation COX-2 [66] and interleukins, with larger percentage of drugs in phase 1 and phase 1|2 comparing with traditional targets, sometimes with a good ratio and even some drugs in phase 4 (already in the market) like COX-2 (phase 1+phase1|2:phase 2: phase 3:phase 4=5:2:2:2). Furthermore, some targets have drug that has been terminated in early stage of clinical trials, indicating that acting on these targets may have undesired toxicity effects, for example, DNA and TOP 1 [63] both have withdrawn drugs, which means involved in the DNA replication process can kill tumor cells but sometimes poisonous to normal cells.

Interestingly, we also find many agents still in clinical trial for CRC, which have already been in the market for a long time to treat other diseases, especially for other cancers. For instance, Pertuzumab [67] (targeting ERBB2), still in Phase 1 for CRC, is an approved drug for metastatic breast cancer. Vandetanib [68], used in clinic for nonresectable, locally advanced, or metastatic medullary thyroid cancer; Sorafenib Tosylate [69], approved for primary kidney and liver cancer; Gemcitabine [70], indicated for the treatment of advanced ovarian cancer, metastatic ovarian cancer; locally advanced, or metastatic non-small cell lung cancer and adenocarcinoma of the pancreas; Lenalidomide, approved for multiple myeloma; Carmustine, treatment for several types of brain cancer, are also in clinical trial for CRC therapy. Actually, nearly half of trials drugs in Phase 1 and Phase 2 are approved for other diseases, infection, inflammation and other types of cancer, indicating it is a good strategy to study the existing drugs in the chemotherapy for other cancer in order to find a new drug for CRC.

The statistic of CRC targets was also plotted according to the pathways the targets involved. **Figure 7** shows the top list pathways with more than 20 CRC related targets involved. Not surprisingly, we found the pathway of cancer (hsa05200) is the number one pathway with 111 related targets. Also, many pathways involved in carcinogenesis are on the top list pathways, such as PI3K-AKT signaling pathway (hsa04151), MicroRNAs in cancer (hsa05206), MAPK signaling pathway (hsa04010), Ras signaling pathway (hsa04014), Focal adhesion (hsa04510), Rap1 signaling pathway (hsa04015), Viral carcinogenesis (hsa05203), TNF signaling pathway (hsa04668), ErbB signaling pathway (hsa04012), Chemokine signaling pathway (hsa04062), Transcriptional misregulation in cancer (hsa05202), Cell Cycle (hsa04110), Apoptosis (hsa04210), JAK-STAT signaling pathway (hsa04630). Moreover, metabolic pathway (hsa01100) is ranking the second pathway, and many pathways

related to metabolism are also on the top list. Furthermore, there are many pathways relevant to different kinds of infection on the top list, for example, HTLV-I infection (hsa04014), Epstein-Barr virus infection (hsa05169), Viral carcinogenesis (hsa05203), pathogenic Escherichia coli infection (hsa05130), T cell receptor signaling pathway (hsa04660), revealing a close relationship between the carcinogenesis and viral infection. Additionally, pathways for other cancer are easily to be found in the top list pathways, some even involved more targets than colorectal cancer pathway (maybe caused by the smaller total number of targets), such as Prostate cancer (hsa05215) and Small cell lung cancer (hsa05222). It implies the similarity among different cancer pathways, providing foundation for the strategy of drug repurposing to seek for new drugs for a specific cancer among the existing drug pools from the other types of cancer.

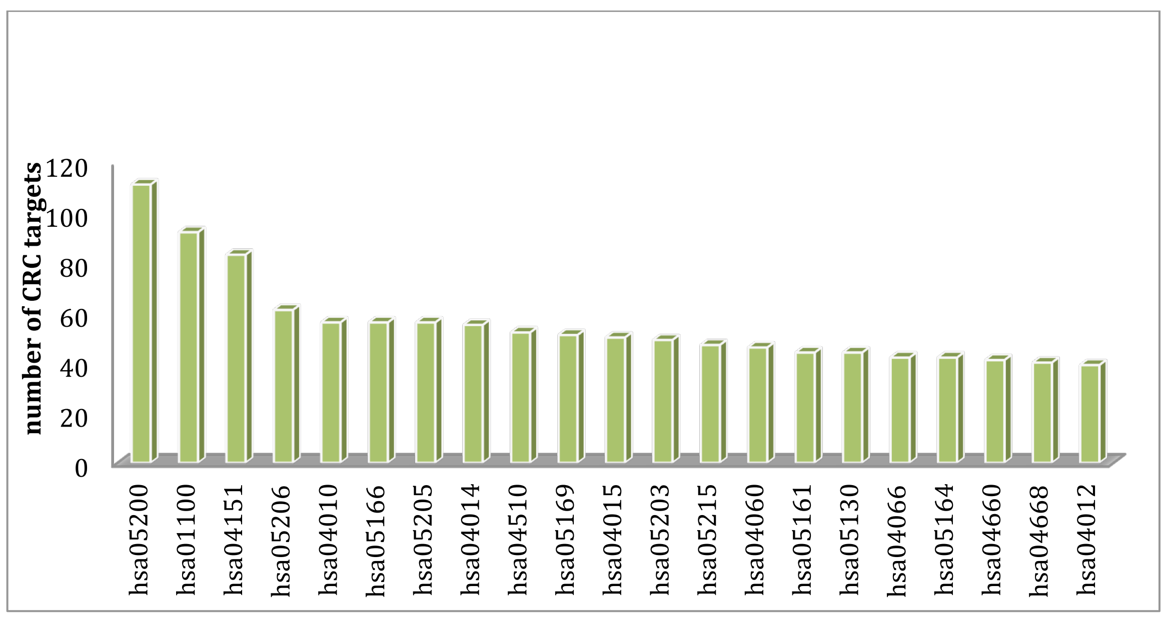


Figure 8. The top list pathways with more than 20 CRC related targets involved.

Table 3. Detailed information of KEGGID and pathway names

KEGGID	Pathway Name
hsa05200	Pathways in cancer - Homo sapiens (human)
hsa01100	Metabolic pathways - Homo sapiens (human)
hsa04151	PI3K-Akt signaling pathway - Homo sapiens (human)
hsa05206	MicroRNAs in cancer - Homo sapiens (human)
hsa04010	MAPK signaling pathway - Homo sapiens (human)
hsa05166	HTLV-I infection - Homo sapiens (human)
hsa05205	Proteoglycans in cancer - Homo sapiens (human)
hsa04014	Ras signaling pathway - Homo sapiens (human)
hsa04510	Focal adhesion - Homo sapiens (human)
hsa05169	Epstein-Barr virus infection - Homo sapiens (human)
hsa04015	Rap1 signaling pathway - Homo sapiens (human)
hsa05203	Viral carcinogenesis - Homo sapiens (human)
hsa05215	Prostate cancer - Homo sapiens (human)
hsa04060	Cytokine-cytokine receptor interaction - Homo sapiens (human)
hsa05161	Hepatitis B - Homo sapiens (human)
hsa05130	Pathogenic Escherichia coli infection - Homo sapiens (human)
hsa04066	HIF-1 signaling pathway - Homo sapiens (human)
hsa05164	Influenza A - Homo sapiens (human)
hsa04660	T cell receptor signaling pathway - Homo sapiens (human)
hsa04668	TNF signaling pathway - Homo sapiens (human)
hsa04012	ErbB signaling pathway - Homo sapiens (human)

3.2 NATURAL PRODUCTS FOR COLORECTAL CANCER

Natural products (NPs) are evolutionarily designed and contain more complex and challenging structures than synthetic compounds [71]. NPs play a significant role in cancer therapeutics, and recently more attentions have been paid to the drug discovery of major lethal malignancies, such as CRC. Many natural products and their analogues have

been identified as potent anti-cancer agents such as Taxol, Vincristin, Campothecin, and more various plants are being identified[72]. We collected natural compounds from the literature with reported anti-CRC activities (**Table 4**). Curvularin, aloe emodin, and kaempferol were selected from Table 4 to further study their potential anti-CRC mechanisms at molecular level via identification of their new targets.

Table 4. Compounds from natural products with anti-CRC activities

Name	Chemical Structure	Molecular Weight	Clog P	Plant Source
6-gingerol		294.39	2.94	*Zingiber officinale Roscoe*
Apigenin		270.24	2.91	*Slaginella tamariscina(Beauv.)Spring*
Baicalin		446.36	0.77	*Scutellaria baicalensis Georgi*
Berberine		322.34	-0.71	*Coptis chinensis Franch.*
Bergenin		328.27	-1.50	*Bergenia purpurascens (Hook. f. et Thoms.) Engl.*
Camptothecin		348.36	0.90	*Camptotheca Acuminata*
Capsaicin		305.42	3.75	*Capsicum annuum*
Cardamonin		270.28	3.49	*Alpinia katsumadai Hayata*

To be continued

Cinobufagin		442.55	3.30	*Bufo bufo gargarizans Cantor*
Curcumin		368.39	2.25	*Curcuma longa L.*
Dihydroartemisinin		284.35	2.45	*Artemisia annua L.*
Diosgenin		414.63	5.91	*Dioscorea opposita Thunb.*
Epicatechin gallate		442.38	2.16	*Camellia sinensis(L.)O.Kuntze 【Thea sinensis L.】*
Epigallocatechin gallate		458.38	2.07	*Camellia sinensis(L.)O.Kuntze 【Thea sinensis L.】*
Evodiamine		303.37	3.66	*Evodia rutaecarpa （Juss.） Benth.*
Gambogic acid		628.76	8.11	*Garcinia hanburyi Hook.f.*
Honokiol		266.34	4.50	*Magnolia officinalis Rehd.et Wils.*
Matrine		248.37	1.5	*Sophora flavescens Ait.*
Oridonin		364.44	-0.13	*Rabdosia rubescens （Hamst.） C. Y. Wu et Hsuan*
Piperidine		85.15	0.94	*Piper longum L.*
Resveratrol		228.25	2.83	*Veratrum nigrum L.*

To be continued

Ursolic acid		456.71	8.63	*Hedyotis diffusa Willd.[Oldenlandia diffusa （Willd.） Roxb.]*
Aloe emodin		270.24	2.70	*Aloe barbadensis Miller*
Andrographo-lide		350.46	2.12	*Andrographis paniculata （Burm. f. ） Nees*
Baicalein		270.24	3.00	*Scutellaria baicalensis Georgi*
Boswellic acid		456.71	9.33	*Boswellia carterii Birdw.*
Caffeic acid phenethyl ester		284.31	3.30	*Apis cerana Fabr*
Cantharidin		196.20	0.18	*Mylabris phalerata Pallas*
Casticin		374.35	2.10	*Vitex trifolia L. var. simplicifolia Cham.*
Coumarin		217.27	2.32	*Psoralea corylifolia L.*
Emodin		270.24	3.62	*Rheum palmatum L*
Ferulic acid		194.19	1.42	*Angelica sinensis （Oliv.） Diels*
Genistein		270.24	1.74	*Kummerowia striata （Thunb.） Schneidl.*
Ginsenoside Rg3		785.03	6.29	*Panax ginseng C. A. Mey.*

To be continued

Protopanaxadiol		460.74	6.78	*Panax ginseng C. A. Mey.*
Ginsenoside Rh2		622.88	6.93	*Panax ginseng C. A. Mey.*
Indirubin-3'-monoxime		279.30	1.18	*Isatis indigotica Fort.*
Kaempferol		286.24	2.10	*Kaempferia galanga L.*
Luteolin		286.24	2.31	*Reseda odorata*
Oxymatrine		264.37	1.63	*Sophora flavescens Ait.*
Pinosylvin		212.25	3.50	*Dioscorea opposita Thunb.*
Plumbagin		188.18	2.78	*Plumbago zeylanica L*
Puerarin		416.38	0.02	*Pueraria lobata （Willd.） Ohwi*
Rocaglamide		505.57	3.73	*Melia toosendan Sieb. et Zucc.*
Shikonin		288.30	3.04	*Arnebia euchroma (Royle) Johnst.*
Silibinin		482.44	1.95	*Silybum marianum (L.) Gaertn.[Carduus marianus L.]*

To be continued

Tanshinone IIA		294.35	5.74	*Slauia miltiorrhiza Bunge*
Thymoquinone		164.20	2.17	*Mosla grosseserrata Maxim.[Orthodon groswsweserratum (Maxim.)Kudo]*
Triptolide		360.41	-0.46	*Tripterygium wilfordii Hook. f.*
Wogonin		284.27	3.33	*Scutellaria baicalensis Georgi*
Zerumbone		218.34	5.28	*Camellia sinensis(L.)O.Kuntze [Thea sinensis L.]*
Quercetin		302.24	1.50	*Sophora japonica L.*
Allicin		162.27	1.9	*Allium sativum L.*
Curvularin		292.33	2.38	*Patrinia scabiosaefolia Fisch. ex Link.*

3.2.1 New targets identification for curvularin

According to the report, one research group isolated the (S)-curvularin, kaempferol, and some other constituents from the Traditional Chinese Herbal, the root of Patrinia Scabra, which is originally used for the gastrointestinal tract diseases in Chinese Medicine [73]. In addition, (S)-curvularin is a kind of fungal metabolites produced by a number of fungi such as Curvularia, *Penicillium* for which antibacterial, cytotoxic, nemeticidial, antitrypanosomal and NF-κB inhibiting activities have been reported [74]. Some researchers also showed that

(S)-curvularin may reduce the expression of the proinflammatory enzyme iNOS in a glucocorticoid-resistant model of rheumatoid arthritis by inhibiting the JAK/STAT signaling pathway [75, 76]. We mainly focused on elucidating the potential mechanism of curvularin by finding new CRC-related targets with the help of our established tool TargetHunter (**Figure 8**).

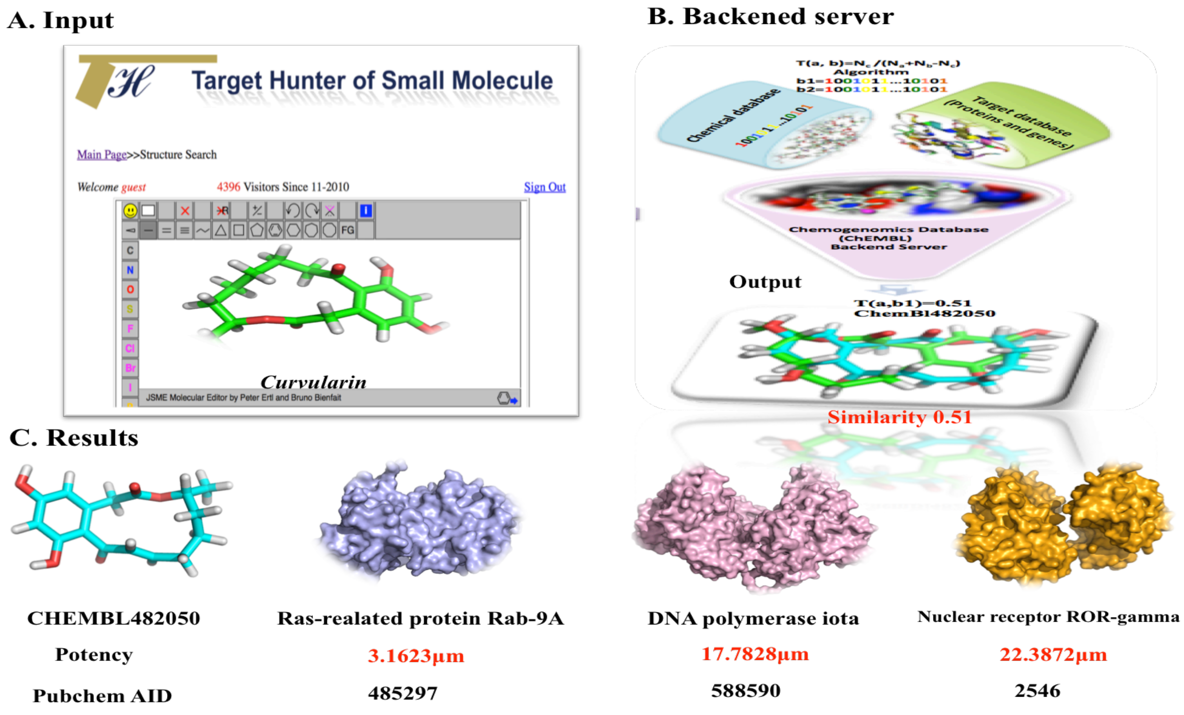


Figure 9. Overview of targets predictions for curvularin by TargetHunter.

In order to validate these predicted targets for curvularin efficiently, we applied the unique and useful function of TargetHunter to find the potential collaborators nearby our area by clicking 'find assay nearby'.

We further analyzed the nuclear receptor ROR-gamma for curvularin by SYBYL docking results comparisons. **Figure 9A** shows the interactions between ChemBL4802050 and nuclear receptor ROR-gamma (PDB: 4NIE). Its docking score is 5.4104. The key residues for this ligand are Tyr B308, Lys B311, Glu B303, and Asp A291. The major

interaction between CHEMBL4802050 and its residues is hydrogen bond. Their distances are listed in the table for the comparison. **Figure 9B** shows the interactions between curvularin and nuclear receptor ROR-gamma (PDB: 4NIE). The docking score is 6.9184. The key residues for the curvularin are Asp A291, Gly B307, Arg B310, Asp A291, and Tyr B308. We compared these two docking results by listing the key residues in. These two compounds share the two residues Asp A291 and Tyr B308. Based on the principle that similar chemical structures have similar bioactivities, curvularin might act on the target.

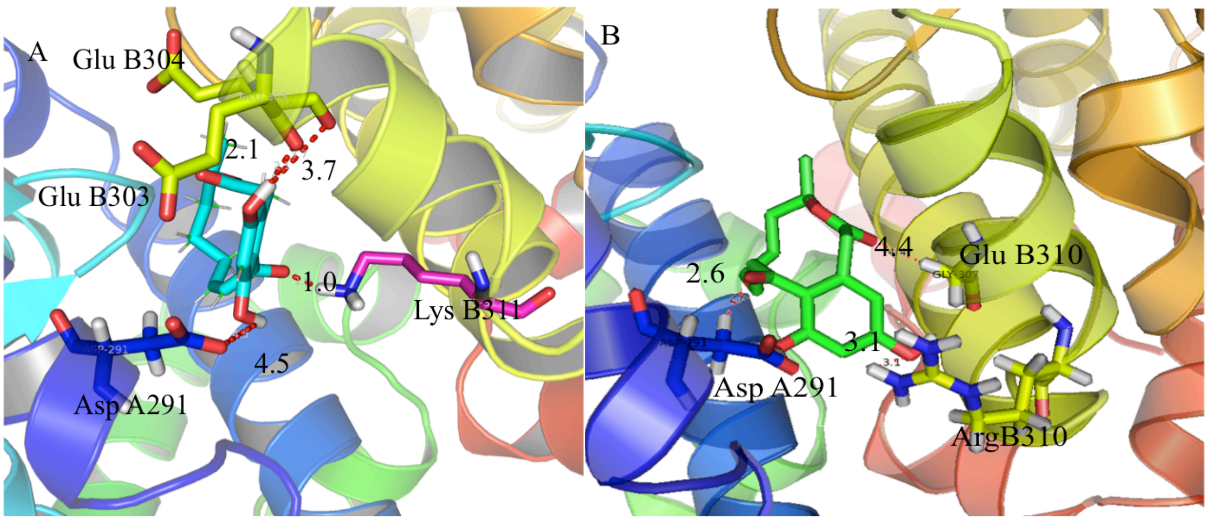


Figure 10. Docking results of CHEMBL4802050 (A) and Curvularin(B). (A) The interactions between ChemBL4802050 and nuclear receptor ROR-gamma (PDB: 4NIE). Its docking score is 5.41. The key residues for this ligand are Tyr B308, Lys B311, Glu B303, and Asp A291. The major interaction between CHEMBL4802050 and its residues is hydrogen bond. Their distances are listed in the table for the comparison. (B) The interactions between curvularin and nuclear receptor ROR-gamma (PDB: 4NIE). The docking score is 6.92. The key residues for the curvularin are Asp A291, Gly B307, Arg B310, Asp A291, and Tyr B308.

3.2.2 New targets identification for aloe emodin

Aloe Emodin (AE) is a natural active compound mainly present in the leaves of *Aloe vera* and the rhizome of *Rheum palmatum*, which are frequently used Traditional Chinese Medicinal herbals. AE has been demonstrated to have various pharmacological activities including antiviral, antimicrobial and hepatoprotective activities [77, 78]. Also, some studies reported that AE exhibited anticancer activity on colon cancer cells, neuroectodermal tumors, lung squamous cell carcinoma, and hepatoma cells [79-81].

It is reported the molecular mechanisms involved in the anti-migratory and anti-angiogenic activity of this hydroxyl anthraquinone in colon cancer cell and AE can down regulate MMP-2/9, RhoB and VEGF via reduced DNA binding activity of NF-κB [82] Another research group did the in vitro inhibition experiments indicating that AE caused the release of apoptosis-inducing factor and cytochrome c from mitochondria and activated caspase-3 leading to DNA fragmentation, nuclear shrinkage and apoptosis, and inhibited casein kinase II activity [79]. However, the molecular mechanisms of anti-CRC effect for AE are still not completely clarified. The online TargetHunter program is an ideal tool to help comprehend the unexplored anticancer targets for understanding anti-CRC mechanism of AE at the molecular level. Therefore, the structure of AE was submitted in the form of SDF as a query to the online TargetHunter program.

There are two related compounds (CHEMBL53418 and CHEMBL418068) were retrieved with similarity scores 0.57 and 0.55 respectively. These two compounds share the same target (tyrosyl-DNA phosphodiesterase 1) but the compound CHEMBL418608 with the potency of 0.5012μm is more potent than CHEMBL53418. The human tyrosyl-DNA phosphodiesterase belong to the phospholipase D (PLD) superfamily. Tdp1 is a monomer

composed of two similar domains that are related by a pseudo-2-fold axis of symmetry. Each domain contributes conserved histidine, lysine and asparagine residues to develop a single active site[83]. Tyrosyl-DNA phosphodiesterase (Tdp1) is a key enzyme involved in the repair of Topoisomerase 1 associated DNA breaks by catalyzing the hydrolysis of 3'-phosphotyrosyl bonds [84]. Colorectal cancer is characterized by the presence of endogenous DNA damage [85]. The targeting of DNA repair enzymes for anticancer therapeutic intervention can be used as strategy to potentiate the cytotoxicity of currently available DNA damaging agents toward cancer cell [86]. Thus, inhibiting Tdp1 might be an underlying anti-CRC mechanism for AE.

Additionally, 15-hydroxyprostaglandin dehydrogenase [NAD+] (15-PGDH) targeted by compound (CHEMBL418068) is shown to have tumor suppressor activity and to be down-regulated in various cancers, including CRC [87]. 15-PDGH controls the rate-limiting step in Prostaglandin E2 (PGE2) catabolism by conversion of PGE2 to 15-keto-PGE2 coupled to the reduction of NAD+ to NADH [88]. PGE2 is the most abundant PG in human colon and its overexpression in colorectal neoplasia compared with normal colorectum [89]. Recently, a research group suggested that PGE2 levels were significantly higher in the center of CRC liver metastasis (CRCLM) compared with peripheral tissue and there were increased levels of 15-PGDH protein in the center of CRCLM associated with reduced 15-PGDH activity and low NAD+/NADH levels. Furthermore, they concluded that based on the intra-tumoral differences in PGE2 metabolism is essential for the development of optimal anti-CRC therapy aimed at the COX-PGE2-15-PGDH [88] (**Figure 10**). Therefore, 15-PGDH might be the potential target of AE for CRC treatment.

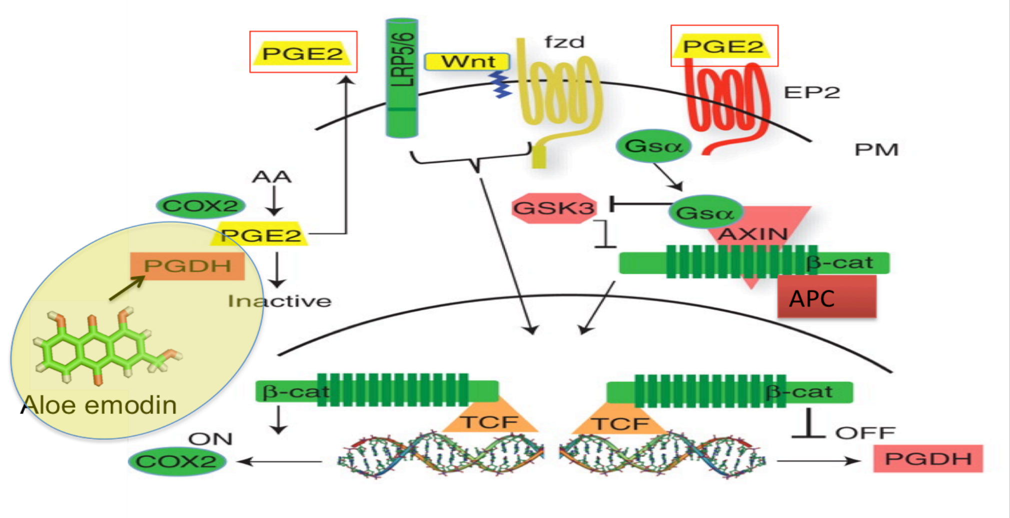


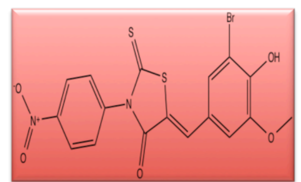
Figure 11. Crosstalk between Wnt and Prostaglandin signalling. Activation of canonical Wnt signalling drives expression of COX2, which catalyzes production of PGE2, and repression of 15-PDGH. PGE2 activates the prostaglandin GPCR receptor EP2 releasing the Gsα subunit that displaces GSK3 form AXIN, resulting in the stabilization of β-catenin.

The unique function of Bioassay Geomap integrated in TargetHunter facilitates further experimental validation for this prediction. Four research laboratories close to Pittsburgh were found by this function (**Table 5**). Hopefully, they could be the potential partners for the further validation of the predicted targets. In a word, our online TargetHunter program provides us with an innovative clue to explore new molecular mechanisms for natural compounds derived from Traditional Chinese Medicine.

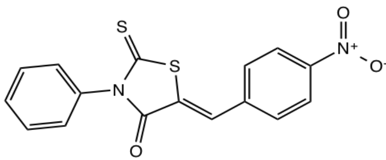
Table 5. Potential collaborators found by Bioassay Geomap function of TargetHunter

Target Name	Address	Reference
Tdp 1	Center for Drug Design, Academic Health Center, University of Minnesota, Minneapolis, Minnesota 55455, USA.	[90]
Tdp 1	Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, and the Center for Cancer Research, Purdue University, West Lafayette, Indiana 47907, USA.	[91]
Tdp 1	Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, DHHS, Frederick, MD 21702, USA	[83]

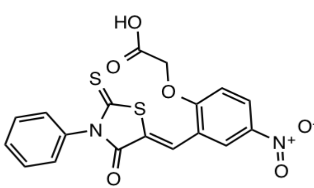
Bioassay Geomap function shows three research institute could potentially validate our predictions not only by TargetHunter but also by our virtual screening compounds. Chemical similarity as a criterion for in silico target identification is based on the well-established medicinal chemistry concept that structurally similar compounds have similar physicochemical properties and possibly similar biological profiles[92, 93]. Leading Tdp1 inhibitors are categorized into four families according to their structure-activity relationship (SAR). Three were amenable for medicinal chemistry follow-up including family 1 members are characterized by analogs derived from Paar-Knorr synthesis of pyrroles (Figure 11,chemical structure in blue box), family 2 members are characterized by a rhodanine substructure (Chemical structure in red box seen in Figure 11), and family 3 members are characterized by an alkylidene barbiturate moiety (Chemical structure in green box) [94]. Based on chemical similarity of reported Tdp1 with a good IC_{50} value, we virtual screening NCI database to find potentially putative Tdp1 inhibitors. NCI database is publicly available from the Enhanced NCI Database Browser: http://cactus.nci.nih.gov/ncidb2.2/(**Figure 11).**



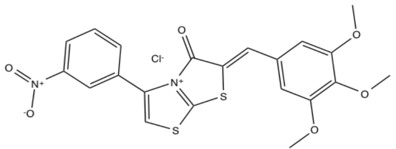
5675956 Tdp1 inhibitors
IC50 1.08μM



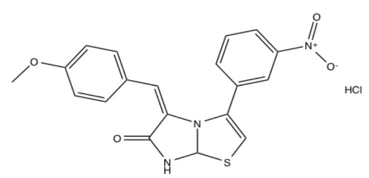
NSC409044
MW 342.39
ClogP 3.79
Similarity 0.75



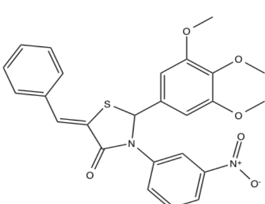
NSC90946
MW 416.42
ClogP 3.30
Similarity 0.75



NSC658290
MW 492.95
ClogP 5.13
Similarity 0.75

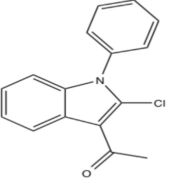


NSC659453
MW417.86
ClogP 3.50
Similarity 0.75

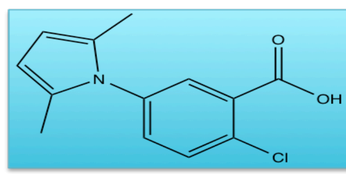


NSC702349
MW 478.52
ClogP 5.71
Similarity 0.75

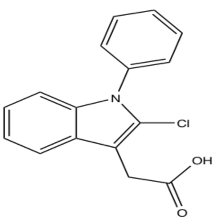
NSC92327
MW 283.76
ClogP 6.38
Similarity 0.75



NSC253525
MW 269.73
ClogP 4.79
Similarity 0.75



Tdp1 inhibitor:5952489
IC50 0.11μM



NSC247069
MW 285.73
ClogP 4.72
Similarity 0.75

NSC89379
MW 265.31
ClogP 5.08
Similarity 0.75

Tdp1 inhibitor compound 50
IC50 0.87μM

NSC31101
MW 381.34
ClogP 1.46
Similarity 70%

NSC122612
MW 311.37
ClogP 1.10
Similarity 70%

NSC731222
MW 345.39
ClogP 2.12
Similarity 70%

NSC142513
MW 311.37
ClogP 1.10
Similarity 70%

NSC122480
MW 297.34
ClogP :0.77
Similarity 70%

NSC98008
MW 292.31
ClogP :0.86
Similarity 65%

NSC409045
MW 364.47
ClogP :1.462
Similarity 65%

NSC660015
MW 226.23
ClogP :2.56
Similarity 65%

NSC694230
MW 380.33
ClogP :-1.53
Similarity 65%

Figure 12. The potential compounds from NCI database are screened for Tdp1 based on chemical similarity.

3.2.3 New target identification for kaempferol

The capsaicin receptor or transient receptor potential vanilloid type 1 (Trpv1) is a heat-activated (at 52°C) cation non-selective channel, which is predominantly expressed in the distal colon and rectum in the gastrointestinal tract [95]. TRPV1 expression is increased in the colon of patients with inflammatory bowel diseases (IBD) known as a risk for the development of CRC and this overexpression is thought to contribute to the ongoing pain and visceral hypersensitivity in these patients. What's more, many epidemiologic and experimental studies have highlighted the relationship between inflammation and cancer [96].

Based on a recent study, the researchers used capsazepine, a TRPV1 antagonist, for its ability to sensitize human colon cancer cells to tumor necrosis factor-related apoptosis-

inducing ligand (TRAIL). They found that capsazepine down regulated cell survival proteins (e.g., survivin) and increased the expression of pro apoptotic proteins (e.g., Bax and p53) [97]. Furthermore, it is reported that TRPV1 has been used to selectively kill cancer cells by activating Ca^{2+} and Na^{+} entry, producing a sustained increase in the cytoplasmic concentration of these ions, and subsequent cell death by apoptosis and necrosis [98]. Immunohistochemical studies have indicated that TRPV1 was expressed in colon adenocarcinoma and the concentration of extracellular polyamines known as TRPV1 agonist, in tissues of the gastrointestinal tract increase in cancer and inflammation [99, 100]. This suggested that TRPV1 may be activated by polyamines in colon cancer, and contribute to the cancer pain. Therefore, researchers hold the particular interests in finding potent compounds target TRPV1.

However, so far there are no reported crystal structures of TRPV1. In addition, to our knowledge, some of the models of TRPV1 are constructed based on the X-ray structure of the voltage-dependent shaker family K+ channel (PDB: 2R9R) [101] from non-TRP family channel. Our group constructed the 3D-homology tetramer models of hTRPV1 based on the cyro-EM derived structure of rTPV1 and selected the best one by using molecular dynamics simulations, energy minimizations, and prescreening. The binding pocket of our model is formed by trans-membrane segments S3, S4, the S4-S5 linker and S5 of one monomer, as well as segments S5 and S6 of the adjacent monomer (subunit) (**Figure 12A)** [102]. We used this model to explore and compare the detailed interactions between hTRPV1 and its antagonist by using molecular docking (**Figure 12B**).

Kaempferol seen in (**Table 4**) was docked with our lab already established TRPV1 protein model. Kaempferol, a flavonol widely found in tea and broccoli is claimed to have an anti-proliferative effect on colon cancer cell lines [103]. Kaempferol with anti-inflammatory,

antiangiogenic properties was previously reported in literature [104]. The docking score of Kaempferol and hTRPV1 protein model is 6.62. Their key residues are Tyr 511, Asn551, Thr 550, Ser 512, Ile 572 [102]. The ligand interacts with residues Tyr 511 and Ile 572 with π-H interaction. Asn551, Thr 550 and Ser 512 bound to Kaempferol with hydrogen bond in the distance of 1.9 Å, 2.1Å and 2.4Å respectively. Manual docking partly helps understand how Kaempferol works on colon cancer cells at the molecular level.

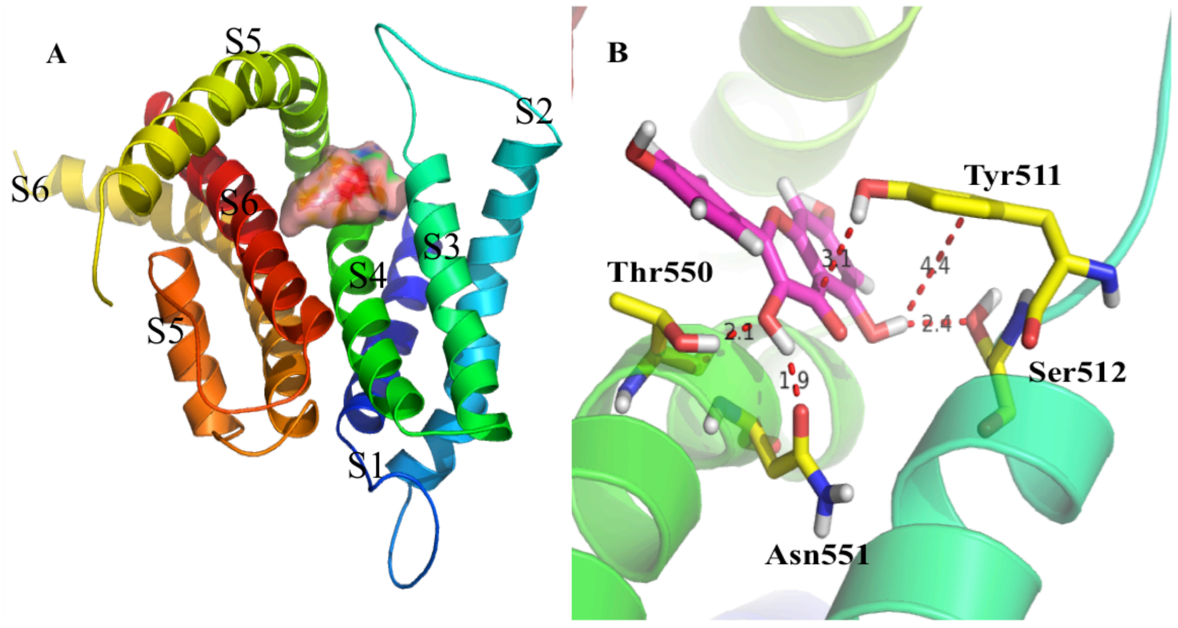


Figure 13. Docking analysis of Kaempferol and hTRPV1 protein model. (A) Overview of TRPV1 and its binding pocket. The binding pocket of our model is formed by trans-membrane segments S3, S4, the S4-S5 linker and S5 of one monomer, as well as segments S5 and S6 of the adjacent monomer (subunit). (B) Docking result of Kaempferol and hTRPV1 with docking score 6.62.

3.3 POLYPHARMACOLOGY ANALYSIS OF ANTI-CRC DRUGS AND THEIR TARGETS

As a validation procedure, we used our established computational chemogenomics method HTDocking to predict the potential targets for several FDA-approved drugs, including Lapatinib (a dual tyrosine kinase inhibitor which interrupts the HER2/neu (ERBB2) and epidermal growth factor receptor (EGFR) pathways), Axitinib [105] (a multi-targeted tyrosine kinase inhibitor), Sorafenib [69] (an inhibitor of several tyrosine protein kinases (VEGFR and PDGFR) and Raf kinases), Sunitinib [106] (a multi-targeted receptor tyrosine kinase inhibitor), OSI-930 (a novel selective inhibitor of the receptor tyrosine kinases Kit (KIT)), and erlotinib [107] (an EGFR receptor inhibitor). The possible interactions between these drugs and CRC proteins were predicted and ranked by docking scores. These associations were plotted as a polypharmacological interaction network (**Figure 13)**.

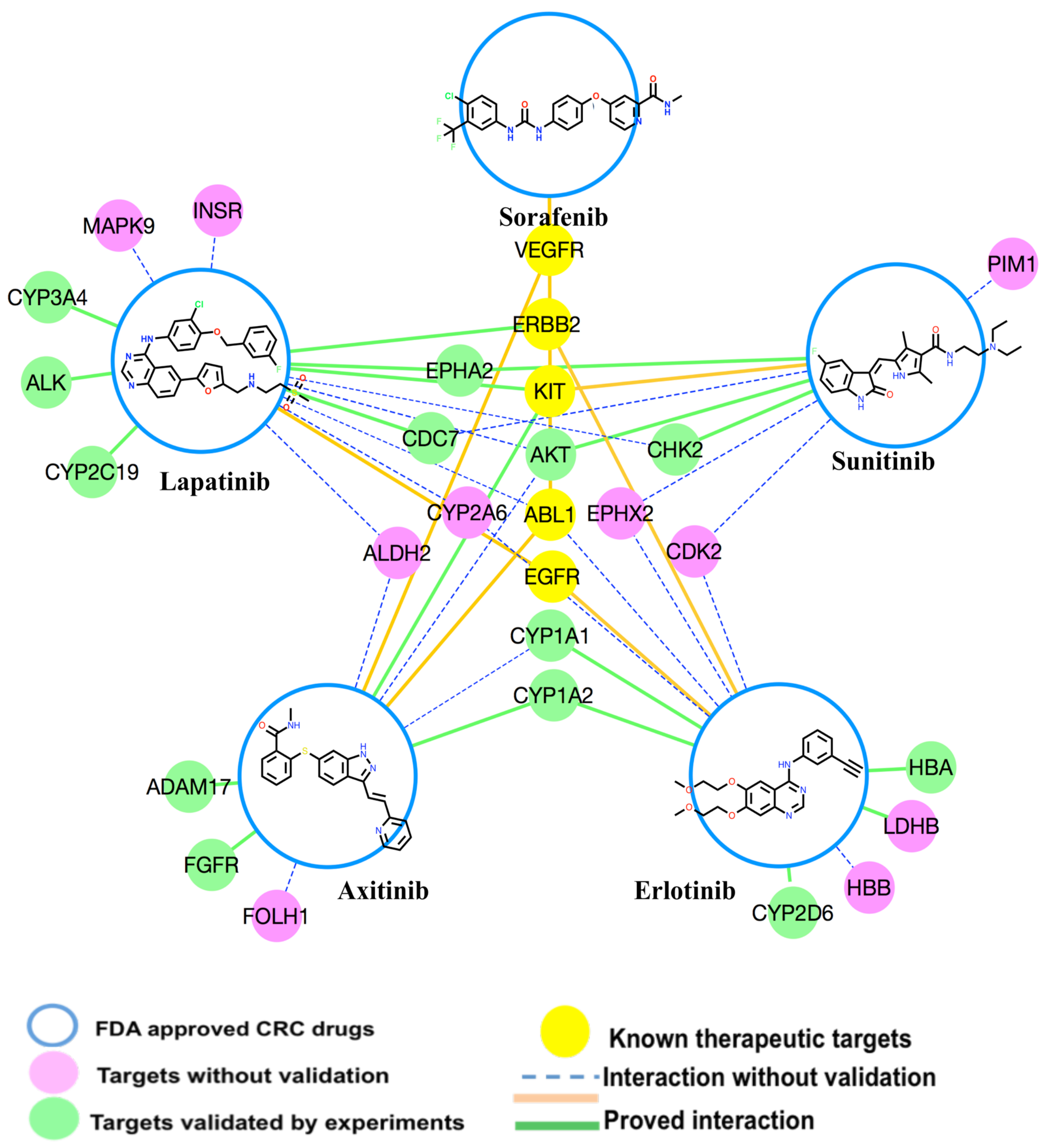


Figure 14. Polypharmacology analysis of five FDA approved anti-CRC drugs.

Primarily, we compared our predicted drug-target association against reported ones in journal articles and databases such as CHEMBL and WOMBAT. Not surprisingly, among our predicted results, the known therapeutic targets or primary targets for these drugs such as EGFR, ERBB2, KIT, and Tyrosine-protein kinase ABL1 were ranked on the top of corresponding predicted target list, such as (Axitinib, Sunitinib, Erlotinib, Lapatinib, and

OSI-930) [108]. Moreover, through comparing the predicted docking scores with the experimental Ki or Kd values, we revealed that our predicted binding affinities against these targets were consistent with the bioactivity data (**Table 6**). Furthermore, some additional predictions of drug-target associations were also validated through data mining of literatures. Our results showed that the predicted interactions between several targets and these drugs were also reported in literature, indicating the reliability of the HT docking program for CRC (**Table 7**). In addition, the other predicted associations, though not been reported in literature yet, could be indicators of novel targets for these drugs, which needs further experimental validations.

Table 6. The comparison of the experimental pKi or pKd data and the predicted pKd values for the FDA-approved CRC drugs

Drug	Target	Docking score	Experimental -Log (Kd or Ki)	Reference
Axitinib	Mast/stem cell growth factor receptor Kit	8.93	8.77	[109]
Axitinib	Disintegrin and metalloproteinase domain-containing protein 17	8.12	9.31	[109]
Axitinib	Fibroblast growth factor receptor 1	7.15	6.42	[109]
Axitinib	Tyrosine-protein kinase ABL1	6.93	8.82	[109]
Lapatinib	Receptor tyrosine-protein kinase erbB-2	9.46	8.80	AID435402
Lapatinib	Tyrosine-protein kinase ABL1	7.03	6.20	AID435146
Erlotinib	Tyrosine-protein kinase ABL1	6.80	6.40	AID624984
Erlotinib	Receptor tyrosine-protein kinase erbB-2	6.57	7.70	AID435907
Sorafenib	Tyrosine-protein kinase ABL1	7.19	6.80	AID256665

Table 7. Verification of other predicted targets of approved CRC drugs by experiment

Approved Drug	Target	Docking (pKd)	Actions	Reference
Erlotinib	Cytochrome P450 1A1	8.59	Substrate	[110]
Erlotinib	Cytochrome P450 1A2	8.08	Substrate	[111]
Erlotinib	Cytochrome P450 2D6	9.11	Substrate	[111]
Lapatinib	Cytochrome P450 2C19	9.29	Substrate	[111]
Lapatinib	Cytochrome P450 3A4	6.84	Substrate inhibit	[112]
Lapatinib	Epidermal growth factor receptor	7.21	Antagonist	[113]
Sunitinib	Mast/stem cell growth factor receptor Kit	6.48	Antagonist	[114]
Lapatinib	Epidermal growth factor receptor	7.21		[115]
Erlotinib	L-lactate dehydrogenase B chain	6.74	Inhibition	[116]
Erlotinib	Hemoglobin subunit alpha	6.51	Inhibition 77.9%	[117]
Erlotinib	Epidermal growth factor receptor	6.34	Inhibition 81%	[118]
Sunitinib	Serine/threonine-protein kinase Chk2	7.94	Inhibition	[119]
Sunitinib	Ephrin type-A receptor 2	7.08	Inhibition	[105]
Sunitinib	RAC-alpha serine/threonine-protein kinase	7.03	Inhibition	[119]
Sunitinib	Mast/stem cell growth factor receptor Kit	6.48	Inhibition 84.8%	[105]

The predicted drug-target association may also be used to find a novel combined medication to achieve improved pharmacological therapeutic effect with reduced side effect [108]. Sorafenib is known to inhibit kinases that regulate angiogenesis (VEGFR2), and proliferation (BRAF), RET liposomal degradation (RET). For instance, BRAF phosphorylates MAP2K1, and thereby contributes to the MAP kinase signal transduction pathway, which affects cell division, differentiation, and secretion. According to our prediction, Sorafenib can also interact with tyrosine kinase AKT and ABL1 (Abelson murine

leukemia viral oncogene homolog 1), which play important roles in the regulation of proliferation [69]. On the other hand, Erlotinib is designed to inhibit EGFR tyrosine kinase, which is highly expressed in cancer cells. Erlotinib may also bind to other potential targets based on our prediction, such as ABL1, CDK2, ERBB2, and several cytochromes P450 [107]. CDK2 is required for the transition from G1 to S phase in cell cycle [65]. Combined use of Erlotinib and Sorafenib, may further inhibit the angiogenesis by blocking EGFR, VEGFR, ERBB2 and ABL1, which are angiogenesis related proteins. Among them, ABL1 regulates the CBL (Castitas B-lineage Lymphoma) family of ubiquitin ligases and phosphorylation of CBL leads to an increased EGFR stability [120]. ABL1 is also involved in the late stage autophagy by regulating lysosomal components; mediates mitochondrial dysfunction and cell death by controlling response to oxidative stress, and involved in DNA-damage response and apoptosis by translocation with DNA binding site.

Together with other predicted targets, they all play different but important roles in the tumor regulation. There are many clinical studies focus on the combination therapy of Sorafenib and erlotinib to treat a variety range of cancers, such as non–small cell lung cancer (NSCLC) [121], hepatocellular carcinoma [122], advanced pancreatic cancer, and other advanced solid tumors [123]. The two drugs provided additional support for studying this combination therapy [124] (**Figure 13**).

Cyclin-dependent kinase 2 (CDK2) is a novel target for erlotinib predicted by HTDocking (**Figure 13**). Based on this new prediction, erlotinib originally was indicated for NSCLC [125] and might be repurposed for other type of cancers. We further analyzed the interactions between CDK2 and erlotinib by manual docking. First, we aligned erlotinib with the original inhibitor indazole [126], and we found that erlotibib matched well with indazole

in original co-crystal (**Figure 14A**) with docking score of 9.89. Then, we analyzed their interactions between erlotinib and CDK2. The key residues are Leu 83, Phe 80, Val 64, Ile10, Leu 134, Asp145, Ala 31, Ala 144 and Phe 146. The major interactions are hydrogen-bond and polar interactions. Erlotinib interacts with Asp 145 through two hydrogen bonds with the distance of 2.5 Å and 2.9Å. Erlotinib also binds to Phe146 through polar interactions with the distance of 2.1 Å (**Figure 14 B and C**).

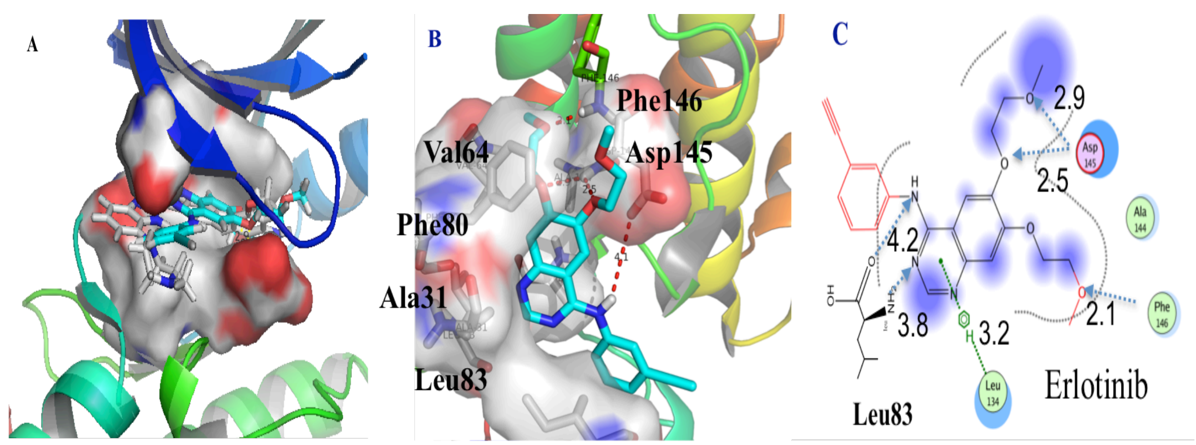


Figure 15. The detailed information of docking analysis between erlotinib and CDK2. (A) The alignment between original CDK2 inhibitor indazole (in white) and erlotinib (in blue); (B) The interactions between erlotinib and CDK2; and (C) The detailed information for the binding interactions and their key residues.

3.4 VIRTUAL SCREENING AND EXPERIMENTAL VALIDATION

3.4.1 Virtual screening of in-house compounds

We applied our in-house compound library to the CRCPlatform. Top 10 targets derived from FDA approved CRC drugs in CRCPlatform were selected to do virtual screening on our in-house compound. We found that insulin growth factor 1 receptor (IGF1R) was predicted as

the potential target for our in-house compounds. IGF1R is frequently expressed in human colon cancers and plays important roles in promoting malignancy. The oncogenic effects are mainly due to the mitogenic and antiapoptotic properties regulated by their IGF1R [127]. We selected those compounds with their docking scores above 6.0 (**Table 8**). Compound XIE62-1032, with the highest docking score of 9.58 on IGF1R was further analyzed (**Figure 15**). Key residues of IGF1R, such as Asp 1056, Arg 1128 and Met 1112 interact with XIE62-1032. Their major interactions are hydrogen bonds, and the hydrogen-bond distances from Asp1056 and Arg 1128 to XIE62-1032 are 2.2Å and 2.1 Å, respectively.

Table 8. **Top seven in-house compounds with docking score above 6 in complex with IGF1R**

Compound-ID	MW	ClogP	Predicted Kd (nM)	Docking Score
XIE35-1107	417.48	4.63	6.20	8.21
XIE95-1170	422.58	6.79	19.96	7.70
XIE62-1032	391.50	5.11	0.26	9.58
XIE18-1014	336.38	2.50	106.49	6.97
XIE18-1025	362.40	1.78	123.08	6.91
XIE62-1044	306.36	2.73	466.77	6.33
XIE95-1156	432.67	5.06	268.29	6.57

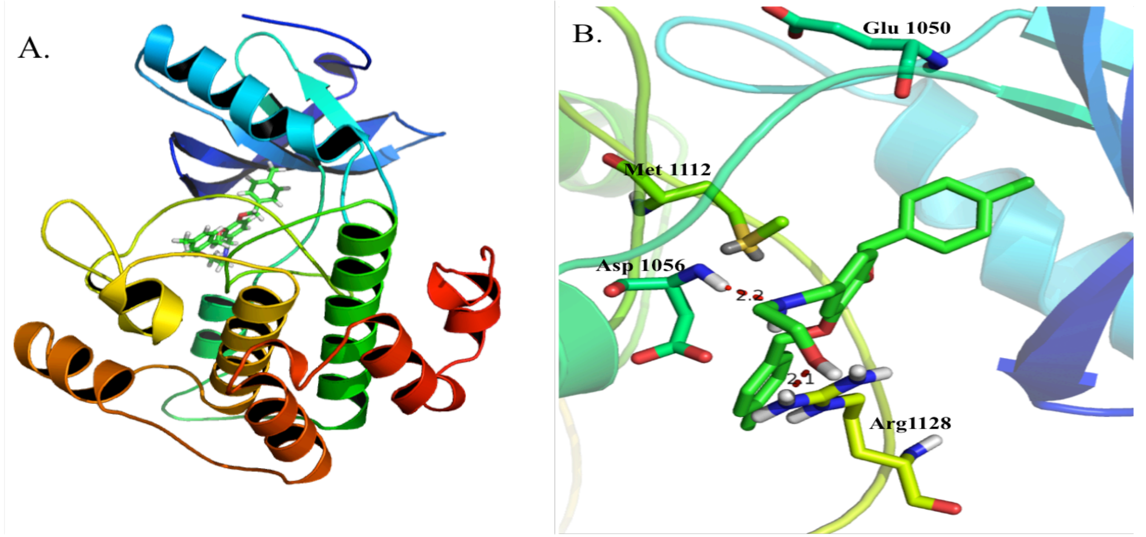


Figure 16. The detailed docking analysis of XIE62-1032 in IGF1R (PDB: 3NW6). (A) The overview of compound XIE62-1032 in IGF1R. (B) The docking results of XIE62-1032 and IGF1R.

3.4.2 Experimental validation of in-house compounds

After the virtual screening and docking analysis, we further validated these top seven compounds. MTS cell proliferation assay and apoptosis analysis were used to test our compounds using the HTC116 colorectal cancer cell line. The MTS assay results showed that XIE62-1032 is the most potent compound to inhibit the proliferation of HCT116 cells (**Figure 16)**, with an IC50 of 3.77 μM (**Figure17**). XIE62-1044 was the second most potent compound with an IC50 around 10 μM. These two compounds were chosen for further testing.

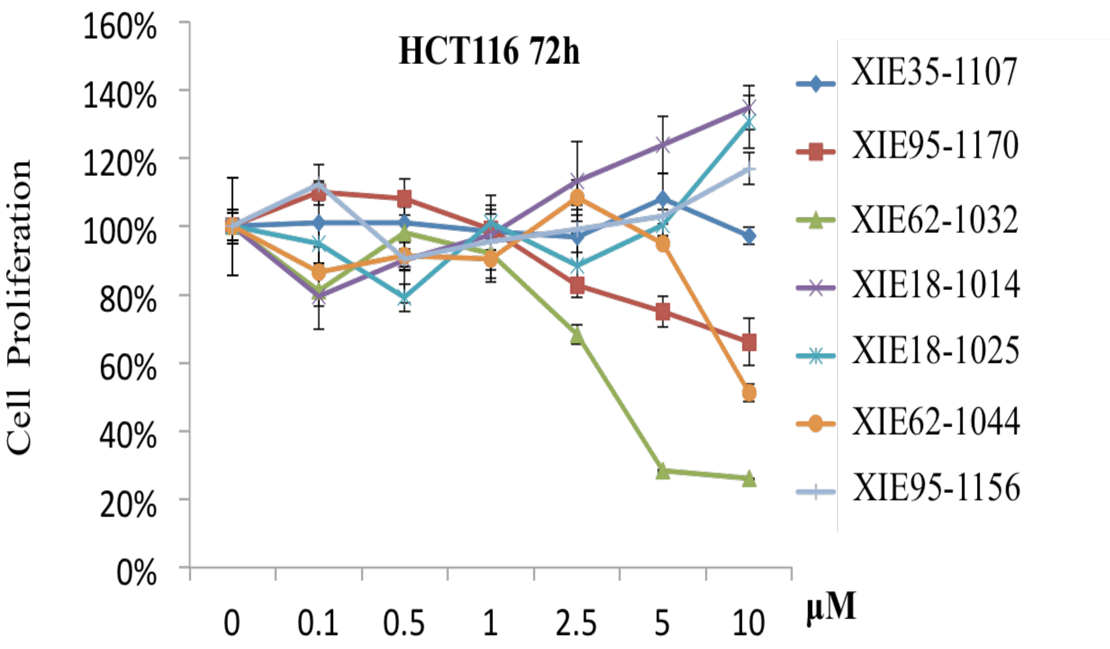


Figure 17. HCT116 cells were treated with indicated compounds for 72 h. Cell viability was determined by MTS assay. Results were expressed as means ± S.D. of three independent experiments.

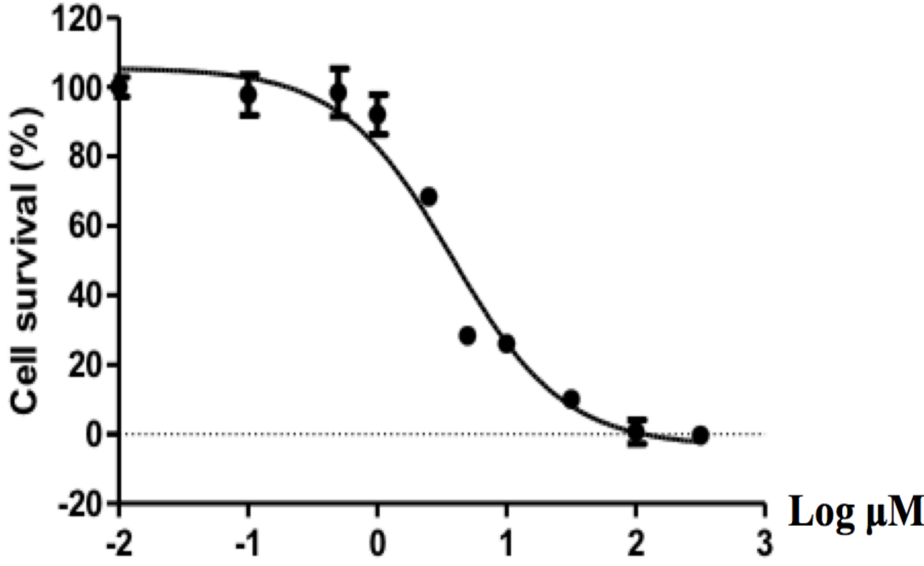


Figure 18. HCT116 cells were treated with XIE62-1032 at different concentrations for 72 h. Cell viability was determined by MTS assay. Results were expressed as means ± S.D. of three independent experiments. IC50 was calculated by Prism.

Induction of apoptosis is an important mechanism of many anticancer agents, and caspase activation is a hallmark for apoptosis and indicated by cleavage [128] Apoptosis is triggered by a series of well-ordered biochemical events and regulated by complex signaling networks[129]. Many important pathways controlling apoptosis and cell survival are commonly altered in cancer[129]. We observed signs of cell killing such as rounding and detachment from plates, and further examined the effects of XIE62-1032 and XIE62-1044 on caspase activation. We observed dose-dependent induction of cleaved caspase-9 and cleaved caspase-3 in (**Figure 18**). Consistent with the MTS data, cells treated with XIE62-1032 showed higher levels of cleaved caspase-9 compared to those treated with XIE62-1044. Interestingly, XIE62-1032 appeared to strongly induce the expression of caspase-9. However, cleavage of caspase-3 appeared more significant upon XIE62-1044 treatment. These results suggest that our in-house compound XIE62-1032 and XIE62-1044 inhibit colon cancer proliferation and promote apoptosis.

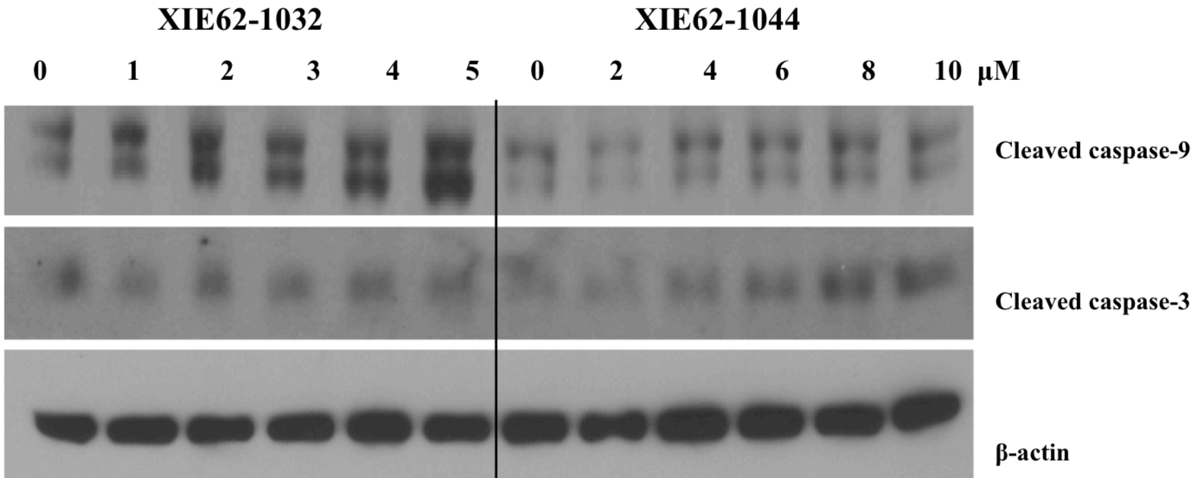


Figure 19. Cells were treated with XIE62-1032 and XIE62-1044 for 24h at indicated concentrations. The indicated proteins were analyzed by western blotting. β-actin is a loading control.

4.0 DISCUSSION

Colorectal carcinogenesis is the complicated adenoma-carcinoma sequence and multiple -step tumor genesis that is induced and influenced by multiple molecular pathways, gene mutations, and their interactions, which may take decades [130]. In order to reduce the impact of CRC, concerted efforts have been made on the prevention, diagnosis, and treatment to control colorectal cancer, leading to the accumulation of a large volume of scientific data related to CRC. Although PubChem and other databases have archived most of the literatures, the published researches are still scattered. So searching, organization, and validation of the reported CRC-related small chemical molecules, and exploitation of the previous bioassay target research results for CRC are not very easy.

Several CRC-related databases have been reported to associate and analyze the results of CRC studies. The CRCgene database provides a comprehensive field synopsis of genetic association studies and meta- analyses for all eligible polymorphisms in Colorectal Cancer [8]. An integrated Oracle database was also constructed contains information in in CRC outcomes, prognosis, and epidemiology, which enables the successful collection of data for CRC prognosis cohort studies [131]. Danish database is mainly for clinical studies [132] and DUBLIN - Maven Semantic can reach individual research in CRC [133]. However, publicly CRC specific chemical genomics (or chemogenomics) database that focuses on small molecules targeting CRC-related proteins is still not available as far as we know. CRCPlatform bridges the knowledge gap between biology and chemistry related to CRC,

improving CRC target studies, drug mechanism explorations, polypharmacology analysis and new drug discovery.

Our case studies implied that the protein structure-based HTDocking program has been successfully predicted the CRC-related targets for small molecules, such as FDA approved drugs, lead compounds and natural products. The reliability of HTDocking program was verified by comparison of the predicted values with the reported pKi values for known CRC drugs. Furthermore, CRCPlatform offers an alternative way to map the drug –target network for the investigation on polypharmacology, clinical combinational therapy and predicting potential adverse drug reaction. So far, "one-compound-one-target" therapeutic paradigm has been challenged despite the fact that enormous efforts have been made to illustrate the pathophysiology, anti-CRC. Recently the multi-target-directed ligand approach has gained increasing attention by many research groups, which have developed plenty of novel compounds act on various biological targets [134]. For example, we identified CDK2 as a new target for erlotinib by polypharmacology analysis, which could be repurposed to treat other types of cancer. However, HTDocking is limited to the availability of co-crystal protein structures, such as TRPV1 receptor is not included in CRCPlatform.

On the other hand, an online service, the ligand-based TargetHunter, is built on the TAMOSIC algorithm to automate the target prediction and it is free accessible to academic and industrial researchers [58]. It is of great use for drug repurposing, and the investigation of potential side effects (off-target) related to CRC drugs. The new targets identification for the compounds of natural products could provide a clue to further investigate the potential mechanism of these natural compounds for CRC. What's more, this powerful cloud-

computing tool embeds the geographical bioassay locator to facilitate the finding nearby collaborators who have published the related bioassay (**Table 5**).

Based on the experimental validation results of in-house compounds, our CRCPlatform acted as a guide to assist finding the proper targets for CRC drug discovery. We virtual screened the potential in-house compounds for CRC and found collaborators to validate the target prediction of the in-house compounds by using HCT-116 colorectal cancer cell line. The MTS assay results suggest that XIE62-1032 is the most potent compound inhibiting the cell proliferation, which is consistent with our prediction that XIE62-1032 with the highest docking score of 9.58. Additional experimental data indicate that XIE62-1032 treatment increases the levels and activation of caspase-9.

The IGF-1R pathway is well known to promote survival and suppress the activation of apoptosis mediated by Caspase-9 and caspase-3 [32]. We hypothesized that our in-house compound XIE62-1032 might bind to the IGF1R and trigger apoptosis signaling pathway by up-regulating capase-9 and capase-3 (see **Figure 19**). Further experimental validation of the binding of our in-house compound XIE62-1032 to IGF1R at protein level are on-going now. The potential collaborators are found by Bioassay Geomap implemented in our TargetHunter and published literature related to IGF1R. The potential collaborators are listed in **Table 9**.

Table 9. The list of potential collaborators for XIE62-1032 and IGF1R binding assay

Institute name	Bioassay	Paper
Department of Biochemistry and Molecular Pharmacology Skirball Institute NYU school of Medicine	Continuous spectrophotometric kinase assay	Small-molecule inhibition and activation-loop trans phosphorylation of the IGF1 receptor
Eppley Institute and Fred& Pamela Buffett Cancer center	In vitro inhibition of IGF1R signaling by the selective IGF1R kinase inhibitor PQIP	Anti-tumor Activity of IGF-1R Kinase Inhibitor PQIP in Colon Cancer
Fox Chase Cancer Center, 333 Cottman Avenue, Room P3165, Philadelphia, Pennsylvania	Kinase selectivity profiling,"HotSpot" assay platform	A highly selective dual insulin receptor (IR)/insulin-like growth factor 1 receptor (IGF-1R) inhibitor derived from an ERK inhibitor*

Table 5. Potential collaborators found by Bioassay Geomap function of TargetHunter

Target Name	Address	Reference
Tyrosyl-DNA Phosphodiesterase 1 (Tdp 1)	Center for Drug Design, Academic Health Center, University of Minnesota, Minneapolis, Minnesota 55455, USA.	[90]
Tyrosyl-DNA Phosphodiesterase 1 (Tdp 1)	Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, and the Center for Cancer Research, Purdue University, West Lafayette, Indiana 47907, USA.	[91]
Tyrosyl-DNA Phosphodiesterase 1 (Tdp 1)	Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, DHHS, Frederick, MD 21702, USA	[83]

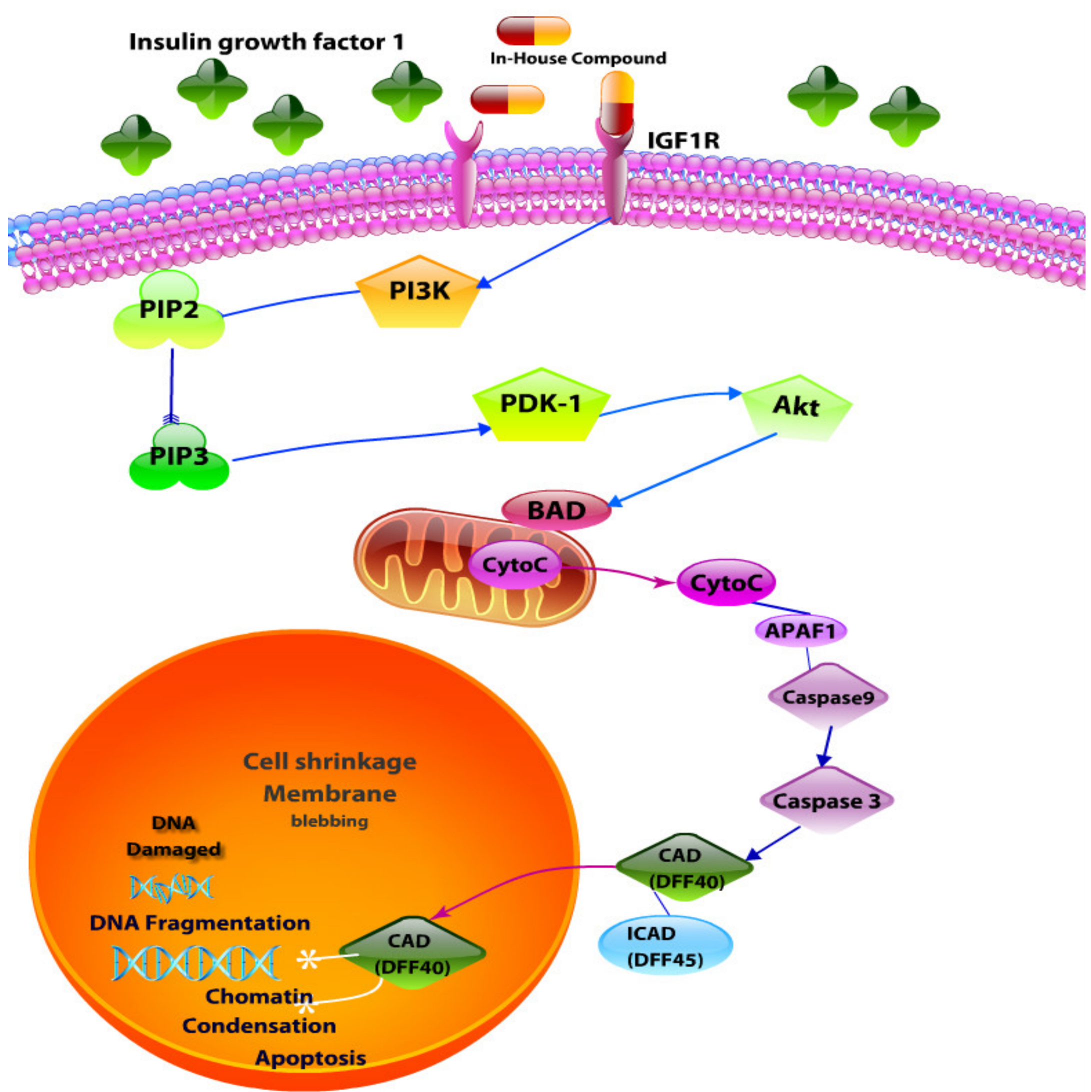


Figure 20. The proposed mechanism of anti-CRC activity of our in-house compound XIE62-1032. Our in-house compound binding IGF1R triggers the apoptosis signaling pathway. Blue arrows, activation; mangenta arrows, translocation from organelles to cytoplasm.

Our chemogenomics tools make this new approach more convenient and efficient for new drug discovery by collecting a variety of CRC related drugs and small molecule with identified targets. Also, our CRC Platform provides CRC related pathways, CRC biomarkers etc. to assist researchers conducting comprehensive CRC studies.

5.0 SUMMARY

CRCPlatform (http://www.cbligand.org/CRC) developed as a publicly accessible cloud-computing server provides comprehensive data and efficient tools for CRC drug and targets research. We employed our FDA-approved or in different phases of clinical trial drugs, natural compounds with anti-CRC activities, and in-house compounds to test and validate our CRCPlatform. The results of FDA approved drugs predicted by HTDocking are validated by literature via comparison of our HTDocking score and literature reported Ki or Kd value. TargetHunter not only can help reveal the mechanisms of natural compounds, but also predict the off-target. The GPU for the acceleration of Tanimoto calculation is imbedded in TargetHunter accelerating the chemical similarity calculations. Furthermore, in-house compounds targeted IGF1R predicted by CRCPlatform were further validated by our collaborators, which indicate that our CRCPlatform is reliable for the CRC research. This online platform will be beneficial for the CRC research groups and bridge the gap between the biology and chemistry, enhance the efficiency of computer-aided drug design research process by finding the potential collaborators and employing existing data as well as computational technology.

6.0 FUTURE SPECULATIONS

This thesis has discussed the CRCPlatform constructions and CRCPlatform tests were carried out in the form of case studies. Different types of chemogenomics tools such as TargetHunter, HTDocking imbedded in CRCPlatform are employed to find potential new targets. What's more, the polypharmacology analysis is used to identify new targets and combination therapy for FDA-approved drugs for CRC. Some of the FDA-approved drugs are already confirmed by the reported literature. In addition, our collaborator conducted experimental validations including MTS assays and apoptosis analysis for our in-house compounds. To further conduct CRC drug discovery research, there is still a long way to go. We will mainly summarize CRC research plan for the next stage as you can see below.

1) Validation of our FDA approved drugs by experiments. We have proposed potentially new targets of FDA approved drugs for combination therapies or drug repurposing. For example, based on the polypharmacology analysis, Erlotinib with newly predicted targets including LDHB, HBB, CDK2, EPHX2; Sunitinib with newly predicted targets including PIM1, CDK2, EPHX2. Both of these two drugs share two similar targets CDK2 and EPHX2, implying that these two drugs may be combined together to improve their therapeutic effects. We still can find more anti-cancer drugs to determine more pairs of combination by applying polypharmcology analysis. Furthermore, high throughput screening assay of anti-cancer drug combination therapy might be

carried out to further confirm their improved effects predicted by our CRCPlatform.

2) The new targets predictions by TargetHunter for natural compounds with anti-CRC activities will be validated by potential collaborators found by Bioassay GeoMap imbedded in TargetHunter (See Table 5). We not only will test the natural compounds but also NCI database compounds by virtual screening.

3) In-house compound XIE62-1034 inhibiting IGF1R related pathway analysis study might be carried out based on our hypothesis described in the discussion part by our potential collaborators. In addition, continuous spectrophotometric kinase assay may also be applied. Finally, if the results were good enough, we can further do research on IGF1R and our in-house compound XIE62-1034 co-crystal.

APPENDIX ABBREVIATIONS

15-PGDH	15-hydroxyprostaglandin dehydrogenase [NAD+]
ABCG2	ATP-binding cassette sub-family G member 2
ADBR2	Beta-2 adrenergic receptor
ADME	Absorption, distribution, metabolism, excretion
AE	Aloe Emodin
AKT1	RAC-alpha serine/threonine-protein kinase
ALB	Serum albumin
APC	Adenomatous polyposis coli
ATIII	Antithrombin-III
BAX	bcl-2-like protein 4
BRAF	Serine/threonine-protein kinase B-raf
CBL	Castitas B-lineage Lymphoma
CDK2	Cyclin-dependent kinase 2
CDK5	Cyclin-dependent kinase 5
CIN	Chromosomal Instability
COMT	Catechol O-methyltransferase
COX-2	Prostaglandin-endoperoxide synthase 2
CRC	Colorectal Cancer
CYP1A1	Cytochrome P450 1A1
CYP1A2	Cytochrome P450 1A2
CYP1B1	Cytochrome P450 1B1
CYP2A6	Cytochrome P450 2A6
CYP2C19	Cytochrome P450 2C19
CYP2D6	Cytochrome P450 2D6
CYP3A4	Cytochrome P450 3A4
DPD	Dihydropyrimidine dehydrogenase [NADP(+)]
EDNRA	Endothelin-1 receptor
EGFR	Epidermal growth factor receptor
EPHA2	Ephrin type-A receptor 2
EPHA2	Ephrin type-A receptor 2
EPHX2	Cytosolic epoxide hydrolase 2
ERBB2	Receptor tyrosine-protein kinase erbB-2
FAP	Familial adenomatous polyposis
FGF2	Fibroblast growth factor 2
GALE	UDP-glucose 4-epimerase
GCR	Glucocorticoid receptor
GI	Gastrointestinal
GPCR	G protein coupled receptor
GPU	Grahphics processing unit
GWAS	Genome-wide association studies
HSD11B1	Corticosteroid 11-beta-dehydrogenase isozyme 1
HSP90	Heat shock protein (HSP) 90-alpha

IBD	Inflammatory bowel diseases
IGF1R	Insulin-like growth factor 1 receptor
IGF1R	Insulin-like growth factor-1 receptor
IL-1	Interleukin-1
IL-6	Interleukin-6
KCND2	Potassium voltage-gated channel subfamily D member 2
KIT	Mast/stem cell growth factor receptor Kit
KPCT	Protein kinase C theta type
KRAS	V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LAP-1	Heat shock cognate 71 kDa protein
MAPK	Mitogen-activated protein kinase
MMR	DNA mismatch repair
MSI	Microsatellite Instability
NCI	National cancer institute
NOS3	Nitric oxide synthase, endothelial
NSCLC	Non–small cell lung cancer
NTRK1	High affinity nerve growth factor receptor
PARP2	Poly [ADP-ribose] polymerase 2
PBS	Phosphate buffer solution
PGE2	Prostaglandin E2
PI3K	Phosphoinositide 3'-kinase
PLD	phospholipase D
PVDF	Polyvinylidene Fluoride
RTK	Receptor tyrosine kinase
SAR	Structure-activity relationship
SMILES	Simplified molecular-input line-entry system
Tdp1	Tyrosyl-DNA phosphodiesterase 1
TNF-alpha	Tumor necrosis factor (Cachectin) (TNF-alpha)
TOP1	DNA topoisomerase 1
TRAIL	Tumor necrosis factor-related apoptosis-inducing ligand
Trpv1	Transient receptor potential vanilloid type 1
TYMS	thymidylate synthetase
VEGFR	Vascular Endothelial Growth Factor Receptor

BIBLIOGRAPHY

1. Brenner, H., M. Kloor, and C.P. Pox, *Colorectal cancer.* The Lancet. 383(9927): p. 1490-1502.

2. Atkin, W.S., et al., *Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial.* Lancet, 2010. **375**(9726): p. 1624-33.

3. Siegel, R., et al., *Cancer statistics, 2014.* CA Cancer J Clin, 2014. 64(1): p. 9-29.

4. Howlader N, N.A., Krapcho M, et al., *SEER Cancer Statistics Review, 1975-2010,* 2013.

5. NCI, *General Information About Colon Cancer.* 2014-05-12.

6. Slattery, M.L., et al., *Diet, physical activity, and body size associations with rectal tumor mutations and epigenetic changes.* Cancer Causes Control, 2010. 21(8): p. 1237-45.

7. Sonnenberg, A. and R.M. Genta, *Helicobacter pylori is a risk factor for colonic neoplasms.* Am J Gastroenterol, 2013. 108(2): p. 208-15.

8. Kostic, A.D., et al., *Genomic analysis identifies association of Fusobacterium with colorectal carcinoma.* Genome Res, 2012. 22(2): p. 292-8.

9. Karsa, L.V., et al., *The dimensions of the CRC problem.* Best Pract Res Clin Gastroenterol, 2010. 24(4): p. 381-96.

10. Armour, B.S., et al., *State-level prevalence of cigarette smoking and treatment advice, by disability status, United States, 2004.* Prev Chronic Dis, 2007. 4(4): p. A86.

11. *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. 487(7407): p. 330-7.

12. Sadanandam, A., et al., *A colorectal cancer classification system that associates cellular phenotype and responses to therapy.* Nat Med, 2013. 19(5): p. 619-25.

13. Kinzler KW, V.B., *Lessons from hereditary colorectal cancer.* Cell, 1996(87): p. 159-70.

14. Perea, J., et al., *Early-onset colorectal cancer is an easy and effective tool to identify retrospectively Lynch syndrome.* Ann Surg Oncol, 2011. 18(12): p. 3285-91.

15. Fearon, E.R., *Molecular genetics of colorectal cancer.* Annu Rev Pathol, 2011. **6**: p. 479-507.

16. Jass, J.R., *Classification of colorectal cancer based on correlation of clinical, morphological and molecular features.* Histopathology, 2007. 50(1): p. 113-30.

17. Silla, I.O., et al., *Early-onset colorectal cancer: a separate subset of colorectal cancer.* World J Gastroenterol, 2014. 20(46): p. 17288-96.

18. Vogelstein, B., et al., *Cancer genome landscapes.* Science, 2013. 340(6127): p. 1546-1558.

19. Risio, M., *The natural history of adenomas.* Best Pract Res Clin Gastroenterol, 2010. 24(3): p. 271-80.

20. Jones, S., et al., *Comparative lesion sequencing provides insights into tumor evolution.* Proc Natl Acad Sci U S A, 2008. 105(11): p. 4283-8.

21. Geiersbach, K.B. and W.S. Samowitz, *Microsatellite instability and colorectal cancer.* Arch Pathol Lab Med, 2011. 135(10): p. 1269-77.

22. Jass, J.R., J. Young, and B.A. Leggett, *Evolution of colorectal cancer: change of pace and change of direction.* J Gastroenterol Hepatol, 2002. 17(1): p. 17-26.

23. Leggett, B. and V. Whitehall, *Role of the serrated pathway in colorectal cancer pathogenesis.* Gastroenterology, 2010. 138(6): p. 2088-100.

24. Samadder, N.J., K. Jasperson, and R.W. Burt, *Hereditary and Common Familial Colorectal Cancer: Evidence for Colorectal Screening.* Dig Dis Sci, 2014.

25. Roy, H.K. and H.T. Lynch, *Diagnosing Lynch syndrome: is the answer in the mouth?* Gut, 2003. 52(12): p. 1665-7.

26. Rustgi, A.K., *The genetics of hereditary colon cancer.* Genes Dev, 2007. 21(20): p. 2525-38.

27. Nielsen, M., et al., *Germline mutations in APC and MUTYH are responsible for the majority of families with attenuated familial adenomatous polyposis.* Clin Genet, 2007. 71(5): p. 427-33.

28. Baserga, R., et al., *The IGF-I receptor in cell growth, transformation and apoptosis.* Biochim Biophys Acta, 1997. 1332(3): p. F105-26.

29. Butler, A.A., et al., *Insulin-like growth factor-I receptor signal transduction: at the interface between physiology and cell biology.* Comp Biochem Physiol B Biochem Mol Biol, 1998. 121(1): p. 19-26.

30. Reinmuth, N., et al., *Blockade of insulin-like growth factor I receptor function inhibits growth and angiogenesis of colon cancer.* Clin Cancer Res, 2002. 8(10): p. 3259-69.

31. Pollak, M., *Insulin and insulin-like growth factor signalling in neoplasia.* Nat Rev Cancer, 2008. 8(12): p. 915-28.

32. Chitnis, M.M., et al., *The type 1 insulin-like growth factor receptor pathway.* Clin Cancer Res, 2008. 14(20): p. 6364-70.

33. Chowdhury S, D.I., Sharratt E, Spernyak J, Brattain MG, et al., *Anti-tumor Activity of IGF-1R Kinase Inhibitor PQIP in Colon Cancer.* Clin Exp Pharmacol 2013. S4: 005.

34. Paul, S.M., et al., *How to improve R&D productivity: the pharmaceutical industry's grand challenge.* Nat Rev Drug Discov, 2010. 9(3): p. 203-14.

35. Rotella, D.P., *Drug discovery 2012 and beyond.* ACS Med Chem Lett, 2012. 3(3): p. 172-3.

36. Lamb, J., et al., *The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.* Science, 2006. 313(5795): p. 1929-35.

37. Hindorff, L.A., et al., *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.* Proc Natl Acad Sci U S A, 2009. 106(23): p. 9362-7.

38. Segal, E., et al., *Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data.* Nat Genet, 2003. 34(2): p. 166-76.

39. Hopkins, A.L., *Network pharmacology: the next paradigm in drug discovery.* Nat Chem Biol, 2008. **4**(11): p. 682-90.

40. Boran, A.D. and R. Iyengar, *Systems approaches to polypharmacology and drug discovery.* Curr Opin Drug Discov Devel, 2010. 13(3): p. 297-309.

41. Peters, J.U., *Polypharmacology - foe or friend?* J Med Chem, 2013. 56(22): p. 8955-71.

42. Rask-Andersen, M., M.S. Almen, and H.B. Schioth, *Trends in the exploitation of novel drug targets.* Nat Rev Drug Discov, 2011. 10(8): p. 579-90.

43. Xie, X.Q., et al., *Chemogenomics knowledgebased polypharmacology analyses of drug abuse related G-protein coupled receptors and their ligands.* Front Pharmacol, 2014. 5: p. 3.

44. Roth, B.L., D.J. Sheffler, and W.K. Kroeze, *Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia.* Nat Rev Drug Discov, 2004. 3(4): p. 353-9.

45. Anighoro, A., J. Bajorath, and G. Rastelli, *Polypharmacology: challenges and opportunities in drug discovery.* J Med Chem, 2014. 57(19): p. 7874-87.

46. Liu, H., et al., *AlzPlatform: an Alzheimer's disease domain-specific chemogenomics knowledgebase for polypharmacology and target identification research.* J Chem Inf Model, 2014. 54(4): p. 1050-60.

47. Wang, L., et al., *TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database.* The AAPS journal, 2013. 15(2): p. 395-406.

48. O'Boyle, N.M., et al., *Open Babel: An open chemical toolbox.* Journal of Cheminformatics, 2011. 3(1): p. 1-14.

49. Ertl, P., *Molecular structure input on the web.* Journal of cheminformatics, 2009. 2(1): p. 1-9.

50. Wishart, D.S., et al., *DrugBank: a knowledgebase for drugs, drug actions and drug targets.* Nucleic acids research, 2008. 36(suppl_1): p. D901-D906.

51. Rédei, G.P., *PubChem*. 2008, Springer: New York. p. 1599.

52. National Cancer, I., *Navigate the National Cancer Institute: cancer.gov*. Vol. no. 03-5008.; no. 03-5008. 2002, Bethesda, eBook: U.S. Dept. of Health and Human Services, National Institutes of Health, National Cancer Institute.

53. Ogata, H., et al., *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic acids research, 1999. 27(1): p. 29-34.

54. Kanehisa, M., et al., *The KEGG databases at GenomeNet.* Nucleic acids research, 2002. 30(1): p. 42-46.

55. Bento, A.P., et al., *The ChEMBL bioactivity database: an update.* Nucleic acids research, 2013. 42(D1): p. D1083-D1090.

56. Buolamwini, J.K. and A.A. Adjei, *Novel anticancer drug protocols*. Vol. 85.; 85. 2003, Totowa, NJ: Humana Press.

57. Sandeep, G., et al., *AUDocker LE: A GUI for virtual screening with AUTODOCK Vina.* BMC research notes, 2011. 4(1): p. 445-445.

58. Wang, L., et al., *TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database.* Aaps j, 2013. 15(2): p. 395-406.

59. Kogej, T., et al., *Multifingerprint based similarity searches for targeted class compound selection.* J Chem Inf Model, 2006. 46(3): p. 1201-13.

60. Sun, Q., et al., *Chemosensitization of head and neck cancer cells by PUMA.* Mol Cancer Ther, 2007. 6(12 Pt 1): p. 3180-8.

61. Sun, Q., et al., *Chemosensitization of head and neck cancer cells by PUMA.* Mol Cancer Ther, 2007. 6(12): p. 3180-8.

62. Yu, J., et al., *PUMA induces the rapid apoptosis of colorectal cancer cells.* Mol Cell, 2001. 7(3): p. 673-82.

63. Giovanella, B.C., et al., *DNA Topoisomerase I-Targeted Chemotherapy of Human Colon Cancer in Xenografts.* Science, 1989. 246(4933): p. 1046-1048.

65. Krause, D.S. and R.A. Van Etten, *Tyrosine kinases as targets for cancer therapy.* The New England journal of medicine, 2005. 353(2): p. 172-187.

66. *Cancer Therapies; COX-2 Inhibitors Suppress Prostate Cancer Spread.* Drug Week, 2000: p. 13.14.

67. Barthélémy, P., et al., *Pertuzumab: development beyond breast cancer.* Anticancer Research, 2014. 34(4): p. 1483.

68. Frampton, J.E., *Vandetanib.* Drugs, 2012. 72(10): p. 1423.

69. Rini, B.I., *Sorafenib.* Expert opinion on pharmacotherapy, 2006. 7(4): p. 453-461.

70. Yu, Z., et al., *Gemcitabine Adjuvant Therapy for Resected Pancreatic Cancer: A Meta-analysis.* American Journal of Clinical Oncology, 2014: p. 1.

71. Miura, K., et al., *The use of natural products in colorectal cancer drug discovery.* Expert Opin Drug Discov, 2015: p. 1-16.

72. Prakash, O., et al., *Anticancer Potential of Plants and Natural Products: A Review.* American Journal of Pharmacological Sciences, 2013. 1(6): p. 104-115.

73. GU Zheng-bing, Y.G.-j., CONG Hai-ying, XU Yi-xin, CHEN Hai-sheng, ZHANG Wei-dong, *Studies on chemical constituents from root of Patrinia scabra.* Chinese Traditional and Herbal Drugs, 2002. 09.

74. Aly, A.H., et al., *NF kappa B inhibitors and antitrypanosomal metabolites from endophytic fungus Penicillium sp. isolated from Limonium tubiflorum.* Bioorg Med Chem, 2011. 19(1): p. 414-21.

75. Elzner, S., et al., *Inhibitors of inducible NO synthase expression: total synthesis of (S)-curvularin and its ring homologues.* ChemMedChem, 2008. 3(6): p. 924-39.

76. Schmidt, N., et al., *Transcriptional and post-transcriptional regulation of iNOS expression in human chondrocytes.* Biochem Pharmacol, 2010. 79(5): p. 722-32.

77. Eshun, K. and Q. He, *Aloe vera: a valuable ingredient for the food, pharmaceutical and cosmetic industries--a review.* Crit Rev Food Sci Nutr, 2004. 44(2): p. 91-6.

78. Lin, M.L., et al., *Aloe-emodin induces apoptosis of human nasopharyngeal carcinoma cells via caspase-8-mediated activation of the mitochondrial death pathway.* Cancer Lett, 2010. 291(1): p. 46-58.

79. Lin, K.Y. and Y.H. Uen, *Aloe-emodin, an anthraquinone, in vitro inhibits proliferation and induces apoptosis in human colon carcinoma cells.* Oncol Lett, 2010. 1(3): p. 541-547.

80. Lee, H.Z., et al., *Effects and mechanisms of aloe-emodin on cell death in human lung squamous cell carcinoma.* Eur J Pharmacol, 2001. 431(3): p. 287-95.

81. Pecere, T., et al., *Aloe-emodin is a new type of anticancer agent with selective activity against neuroectodermal tumors.* Cancer Res, 2000. 60(11): p. 2800-4.

82. Suboj, P., et al., *Aloe emodin inhibits colon cancer cell migration/angiogenesis by downregulating MMP-2/9, RhoB and VEGF via reduced DNA binding activity of NF-kappaB.* Eur J Pharm Sci, 2012. 45(5): p. 581-91.

83. Weidlich, I.E., et al., *Inhibitors of human tyrosyl-DNA phospodiesterase (hTdp1) developed by virtual screening using ligand-based pharmacophores.* Bioorg Med Chem, 2010. 18(1): p. 182-9.

84. Walker, S., et al., *Development of an oligonucleotide-based fluorescence assay for the identification of tyrosyl-DNA phosphodiesterase 1 (TDP1) inhibitors.* Anal Biochem, 2014. 454: p. 17-22.

85. Krell, J., et al., *Growth Arrest-Specific Transcript 5 Associated snoRNA Levels Are Related to p53 Expression and DNA Damage in Colorectal Cancer.* PLoS One, 2014. 9(6): p. e98561.

86. Dexheimer, T.S., et al., *Tyrosyl-DNA phosphodiesterase as a target for anticancer therapy.* Anticancer Agents Med Chem, 2008. 8(4): p. 381-9.

87. Kang, P.S., et al., *Prognostic implication of 15-hydroxyprostaglandin dehydrogenase down-regulation in patients with colorectal cancer.* J Korean Soc Coloproctol, 2012. 28(5): p. 253-8.

88. Tai, H.H., et al., *NAD+-linked 15-hydroxyprostaglandin dehydrogenase: structure and biological functions.* Curr Pharm Des, 2006. 12(8): p. 955-62.

89. Rigas, B., I.S. Goldman, and L. Levine, *Altered eicosanoid levels in human colon cancer.* J Lab Clin Med, 1993. 122(5): p. 518-23.

90. Sirivolu, V.R., et al., *5-Arylidenethioxothiazolidinones as inhibitors of tyrosyl-DNA phosphodiesterase I.* J Med Chem, 2012. 55(20): p. 8671-84.

91. Nguyen, T.X., et al., *Synthesis and biological evaluation of the first dual tyrosyl-DNA phosphodiesterase I (Tdp1)-topoisomerase I (Top1) inhibitors.* J Med Chem, 2012. 55(9): p. 4457-78.

92. Martin, Y.C., J.L. Kofron, and L.M. Traphagen, *Do structurally similar molecules have similar biological activity?* J Med Chem, 2002. 45(19): p. 4350-8.

93. Heikamp, K. and J. Bajorath, *Large-scale similarity search profiling of ChEMBL compound data sets.* J Chem Inf Model, 2011. 51(8): p. 1831-9.

94. Dean, R.A., et al., *Identification of a putative Tdp1 inhibitor (CD00509) by in vitro and cell-based assays.* J Biomol Screen, 2014. 19(10): p. 1372-82.

95. Matsumoto, K., et al., *Localization of TRPV1 and contractile effect of capsaicin in mouse large intestine: high abundance and sensitivity in rectum and distal colon.* Am J Physiol Gastrointest Liver Physiol, 2009. 297(2): p. G348-60.

96. Karin, M. and F.R. Greten, *NF-kappaB: linking inflammation and immunity to cancer development and progression.* Nat Rev Immunol, 2005. **5**(10): p. 749-59.

97. Sung, B., et al., *Capsazepine, a TRPV1 antagonist, sensitizes colorectal cancer cells to apoptosis by TRAIL through ROS-JNK-CHOP-mediated upregulation of death receptors.* Free Radic Biol Med, 2012. 53(10): p. 1977-87.

98. Prevarskaya, N., L. Zhang, and G. Barritt, *TRP channels in cancer.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2007. 1772(8): p. 937-946.

99. Domotor, A., et al., *Immunohistochemical distribution of vanilloid receptor, calcitonin-gene related peptide and substance P in gastrointestinal mucosa of patients with different gastrointestinal disorders.* Inflammopharmacology, 2005. 13(1-3): p. 161-77.

100. Ahern, G.P., X. Wang, and R.L. Miyares, *Polyamines are potent ligands for the capsaicin receptor TRPV1.* J Biol Chem, 2006. 281(13): p. 8991-5.

101. Long, S.B., et al., *Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment.* Nature, 2007. 450(7168): p. 376-82.

102. Feng, Z., et al., *Structural Insight into Tetrameric hTRPV1 from Homology Modeling, Molecular Docking, Molecular Dynamics Simulation, Virtual Screening and Bioassay Validations.* J Chem Inf Model, 2015.

103. Li, W., et al., *Kaempferol induces apoptosis in human HCT116 colon cancer cells via the Ataxia-Telangiectasia Mutated-p53 pathway with the involvement of p53 Upregulated Modulator of Apoptosis.* Chem Biol Interact, 2009. 177(2): p. 121-7.

104. Chahar, M.K., et al., *Flavonoids: A versatile source of anticancer drugs.* Pharmacogn Rev, 2011. **5**(9): p. 1-12.

105. Wang, J., et al., *Discovery and bioactivity of 4-(2-arylpyrido[3',2':3,4]pyrrolo[1,2-f][1,2,4]triazin-4-yl) morpholine derivatives as novel PI3K inhibitors.* Bioorg Med Chem Lett, 2012. 22(1): p. 339-42.

106. Deeks, E.D. and E. Raymond, *Sunitinib.* BioDrugs, 2011. 25(5): p. 307.

107. Lyseng-Williamson, K.A., *Erlotinib.* Molecular Diagnosis & Therapy, 2013. 17(1): p. 57.

108. Levitzki, A., *Tyrosine kinase inhibitors: views of selectivity, sensitivity, and clinical performance.* Annu Rev Pharmacol Toxicol, 2013. 53: p. 161-85.

109. Davis, M.I., et al., *Comprehensive analysis of kinase inhibitor selectivity.* Nat Biotechnol, 2011. 29(11): p. 1046-51.

110. van Erp, N.P., H. Gelderblom, and H.J. Guchelaar, *Clinical pharmacokinetics of tyrosine kinase inhibitors.* Cancer Treat Rev, 2009. 35(8): p. 692-706.

111. Li, J., et al., *Differential metabolism of gefitinib and erlotinib by human cytochrome P450 enzymes.* Clin Cancer Res, 2007. 13(12): p. 3731-7.

112. Teng, W.C., et al., *Mechanism-based inactivation of cytochrome P450 3A4 by lapatinib.* Mol Pharmacol, 2010. 78(4): p. 693-703.

113. Lorenzen, S., et al., *Lapatinib versus lapatinib plus capecitabine as second-line treatment in human epidermal growth factor receptor 2-amplified metastatic gastro-oesophageal cancer: A randomised phase II trial of the Arbeitsgemeinschaft Internistische Onkologie.* Eur J Cancer, 2015.

114. Abrams, T.J., et al., *SU11248 inhibits KIT and platelet-derived growth factor receptor beta in preclinical models of human small cell lung cancer.* Mol Cancer Ther, 2003. 2(5): p. 471-8.

115. Sotelo, M.J., et al., *Lapatinib plus trastuzumab in pretreated human epidermal growth factor receptor 2-positive metastatic breast cancer.* J Cancer Res Ther, 2014. 10(4): p. 967-72.

116. Cai, J., et al., *Design and synthesis of novel 4-benzothiazole amino quinazolines Dasatinib derivatives as potential anti-tumor agents.* Eur J Med Chem, 2013. 63: p. 702-12.

117. Wu, X., et al., *Design and synthesis of novel Gefitinib analogues with improved anti-tumor activity.* Bioorg Med Chem, 2010. 18(11): p. 3812-22.

118. Usui, T., et al., *Discovery of indenopyrazoles as EGFR and VEGFR-2 tyrosine kinase inhibitors by in silico high-throughput screening.* Bioorg Med Chem Lett, 2008. 18(1): p. 285-8.

119. Hottecke, N., et al., *Inhibition of gamma-secretase by the CK1 inhibitor IC261 does not depend on CK1delta.* Bioorg Med Chem Lett, 2010. 20(9): p. 2958-63.

120. Levkowitz, G., et al., *Ubiquitin Ligase Activity and Tyrosine Phosphorylation Underlie Suppression of Growth Factor Signaling by c-Cbl/Sli-1.* Molecular Cell, 1999. 4(6): p. 1029-1040.

121. Lind, J.S.W., et al., *A multicenter phase II study of erlotinib and sorafenib in chemotherapy-naive patients with advanced non-small cell lung cancer.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2010. 16(11): p. 3078-3087.

122. Yang, B.L., et al., *[Effects of Baicalin on an orthotopic transplantation mouse model of mismatch repair gene deficient colorectal cancer].* Zhonghua Wai Ke Za Zhi, 2012. 50(9): p. 843-7.

123. Duran, I., et al., *Phase I targeted combination trial of sorafenib and erlotinib in patients with advanced solid tumors.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2007. 13(16): p. 4849-4857.

124. Lane, D., *Designer combination therapy for cancer.* Nature biotechnology, 2006. 24(2): p. 163-164.

125. Abera, M.B. and M.G. Kazanietz, *Protein Kinase C Alpha Mediates Erlotinib Resistance in Lung Cancer Cells.* Mol Pharmacol, 2015.

126. Trujillo, J.I., et al., *2-(6-Phenyl-1H-indazol-3-yl)-1H-benzo[d]imidazoles: design and synthesis of a potent and isoform selective PKC-zeta inhibitor.* Bioorg Med Chem Lett, 2009. 19(3): p. 908-11.

127. LeRoith, D. and C.T. Roberts, Jr., *The insulin-like growth factor system and cancer.* Cancer Lett, 2003. 195(2): p. 127-37.

128. Yu, J., et al., *PUMA mediates the apoptotic response to p53 in colorectal cancer cells.* Proc Natl Acad Sci U S A, 2003. 100(4): p. 1931-6.

129. Zhang, L. and J. Yu, *Role of apoptosis in colon cancer biology, therapy, and prevention.* Curr Colorectal Cancer Rep, 2013. 9(4).

130. Cunningham, D., et al., *Colorectal cancer.* The Lancet, 2010. 375(9719): p. 1030-1047.

131. Liu, R.Z., et al., *Abstract B94: An integrated database in colorectal cancer outcomes, prognosis, and epidemiology (The COPE database).* Cancer Prevention Research, 2010. 3(12 Supplement): p. B94-B94.

132. Harling, H., *146. Danish colorectal cancer database – Items in the audit.* European Journal of Surgical Oncology (EJSO), 2012. 38(9): p. 781.

133. Anonymous, *Maven Semantic: Colorectal Cancer Research Database Update11-01-25&rft.pub=Business+Wire&rft.externalDocID=2247843681¶mdict=en-US U7 - Newspaper Article*. 2011, Business Wire: New York.

134. Dias, K.S. and C. Viegas, Jr., *Multi-Target Directed Drugs: A Modern Approach for Design of New Drugs for the treatment of Alzheimer's Disease.* Curr Neuropharmacol, 2014. 12(3): p. 239-55.