JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Chemogenomics Approaches to Rationalizing the Mode-of-Action of Traditional Chinese and Ayurvedic Medicines

Fazlin Mohd Fauzi,[‖,†] Alexios Koutsoukas,[‖] Robert Lowe,[‡] Kalpana Joshi,[§] Tai-Ping Fan,[⊥] Robert C. Glen,[‖] and Andreas Bender*[,‖]

[‖]Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[†]Universiti Teknologi MARA (UiTM) Malaysia, 40 450 Shah Alam, Selangor, Malaysia

[‡]Blizard Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, The Blizard Building, 4 Newark Street, London E1 2AT, United Kingdom

[§]Symbiosis School of Biomedical Sciences, Symbiosis International University, Pune, India

[⊥]Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1PD, United Kingdom

Supporting Information

ABSTRACT: Traditional Chinese medicine (TCM) and Ayurveda have been used in humans for thousands of years. While the link to a particular indication has been established in man, the mode-of-action (MOA) of the formulations often remains unknown. In this study, we aim to understand the MOA of formulations used in traditional medicine using an *in silico* target prediction algorithm, which aims to predict protein targets (and hence MOAs), given the chemical structure of a compound. Following this approach we were able to establish several links between suggested MOAs and experimental evidence. In particular, compounds from the 'tonifying and replenishing medicinal' class from TCM exhibit a hypoglycemic effect which can be related to activity of the ingredients against the Sodium-Glucose Transporters (SGLT) 1 and 2 as well as Protein Tyrosine Phosphatase (PTP). Similar results were obtained for Ayurvedic anticancer drugs. Here, both primary anticancer targets (those directly involved in cancer pathogenesis) such as steroid-5-alpha-reductase 1 and 2 were predicted as well as targets which act synergistically with the primary target, such as the efflux pump P-glycoprotein (P-gp). In addition, we were able to elucidate some targets which may point us to novel MOAs as well as explain side effects. Most notably, GPBAR1, which was predicted as a target for both 'tonifying and replenishing medicinal' and anticancer classes, suggests an influence of the compounds on metabolism. Understanding the MOA of these compounds is beneficial as it provides a resource for NMEs with possibly higher efficacy in the clinic than those identified by single-target biochemical assays.

INTRODUCTION

Traditional medicines, in particular traditional Chinese medicine (TCM) and traditional Indian medicine (or Ayurveda, which will be used from here on), have been used by humans for thousands of years.[1] (See Figure 1 on how balance defines health in both TCM and Ayurveda.) Furthermore, these two traditional medicines have provided us with important drugs e.g. artemisinin (an antimalarial drug)[2] and reserpine (an antihypertensive drug).[3] Most recently, in the world's largest international clinical trial, it was concluded that *Artesunate*, a derivative of the Chinese herb *qinghao*, should replace quinine as the treatment of severe malaria in both adult and children worldwide.[4] Similarly, the TCM formulation, *Danshen* Dripping Pill, is currently in a phase III clinical trial for angina pectoris.[5] Thus, although both traditional medicines are considered as complementary and alternative medicines, the medicinal compounds contained therein can potentially be developed into new molecular entities (NMEs).

Traditional medicines have been connected to efficacy in man for thousands of years (though admittedly often not in controlled clinical trial settings),[8] and hence they represent an alternative source of NMEs with the hope of exhibiting better efficacy in the clinical setting. Given that currently efficacy is one of the main reasons for failure in phase II and phase III clinical trials,[9] this is of high practical importance. This is also true when paying attention to the particular chemistry present in traditional medicines: from the year 1981 to 2007, it was observed that 67% of the NMEs introduced into the market were either natural product-based or natural product-inspired compounds, whereas the rest originated from synthetic compounds.[10] This is despite the fact that the majority of the pharmaceutical companies were focusing more on synthetic compounds than natural products as potential NMEs due to the ease of synthesis and modification.[11] Nature has evolved a

Received: November 16, 2012

dx.doi.org/10.1021/ci3005513 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

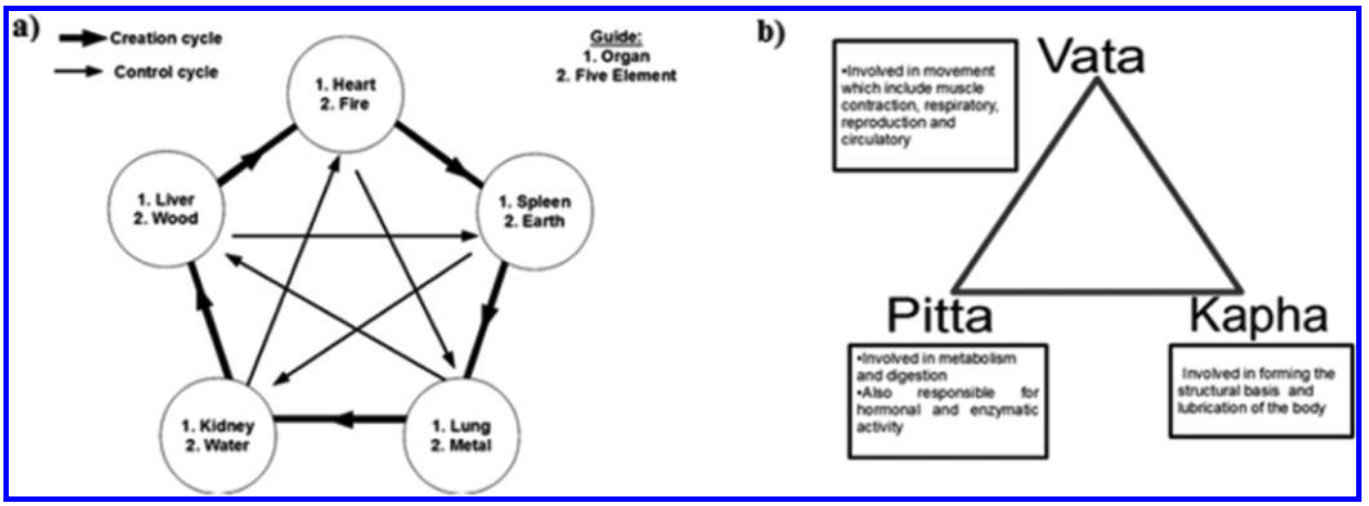


Figure 1. The concept of balance that defines health in both TCM and Ayurveda. TCM seeks to understand the internal workings of the body and also serves as a means to judge and consequently adjust the balance of the human body through The Five phase Theory[6] seen in part (a). Five internal organs are classified according to the five elements in Chinese philosophy that it is most similar to.[6] These organs are interconnected, and an organ can either create or control another organ. For example, the heart controls the lungs hence overexcited lungs can be controlled by reducing the function of the heart.[6] In Ayurveda, the balance between the three *dosha* or *tridosha* determines the health of an individual[7] as can be seen in part (b). These *doshas* are forces that govern the physical and physiological processes in the body.[7] For instance, *Pitta* governs the metabolism and digestion process, *Kapha* is responsible for the formation of the structural basis and lubrication of the body, and *Vata* is responsible for the various movements in the body including muscle contraction and respiration.[7] The interplay of the *tridosha* plays an integral part in regulating different functions in the human body so when one *dosha* is affected, so will the others.[7] Therefore, the balance state between the *tridosha* equals health, whereas an imbalance state signifies a diseased state.[7]

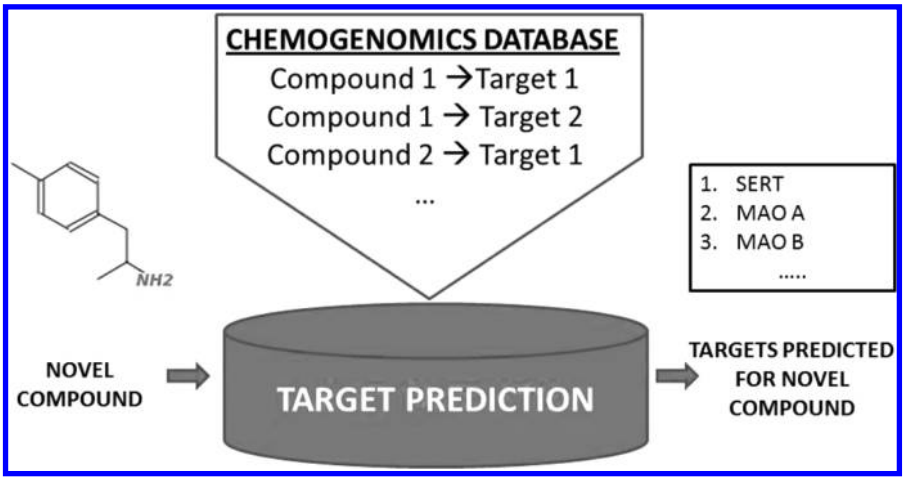


Figure 2. Workflow of the *in silico* target prediction employed in this work for MOA analysis of traditional medicines. First, bioactivity data obtained from the ChEMBL database[26] is compiled and converted into circular fingerprints[27] which are then used to build a Bayes Classifier.[28] Then, when a new compound is encountered, its target(s) can be predicted by comparing the similarity of the unknown compound ('orphaned compound', left) to the ligands where the bioactivity profile is known ('chemogenomics database' and 'target prediction tool', middle). The algorithm will return a list of potential targets, ranked by likelihood of binding, which provide mode of action hypotheses for the compound under consideration.

multitude of chemical compounds with desirable properties that can lead to successful NMEs (while admittedly, often representing challenges to synthetic chemists).[11] First, natural products on average are more soluble than synthetic compounds.[12] It was found that half of the 24 natural products that were successfully developed into drugs but violated Lipinski's Rule-of-Five were still highly bioavailable.[12] This is due to both physicochemical properties in more favorable ranges as well as active transport playing a more important role for endogenous metabolites and biosynthetic intermediates than for synthetic compounds.[12] Second, natural products are often found to embody 'privileged structures'[13] or chemical structures that are more frequently found to bind to a variety of proteins in different living organisms.[14,15] Lastly, the safety and tolerance factors of natural products used especially in traditional medicines are more comprehensively known than their synthetic counterparts.[8] All of these factors lead to the conclusion that natural products as well as traditional medicines have been an undervalued resource of lead structures in the current practice of drug discovery, despite the challenges discussed below.

Corson and Crews[16] outlined four major challenges preventing the transformation of medicinal compounds used in TCM and Ayurveda into possible NMEs, viz. (i) isolation of active constituents, (ii) synthesis of active constituents, (iii) elucidation of the mode of action (MOA), and (iv)

dx.doi.org/10.1021/ci3005513 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

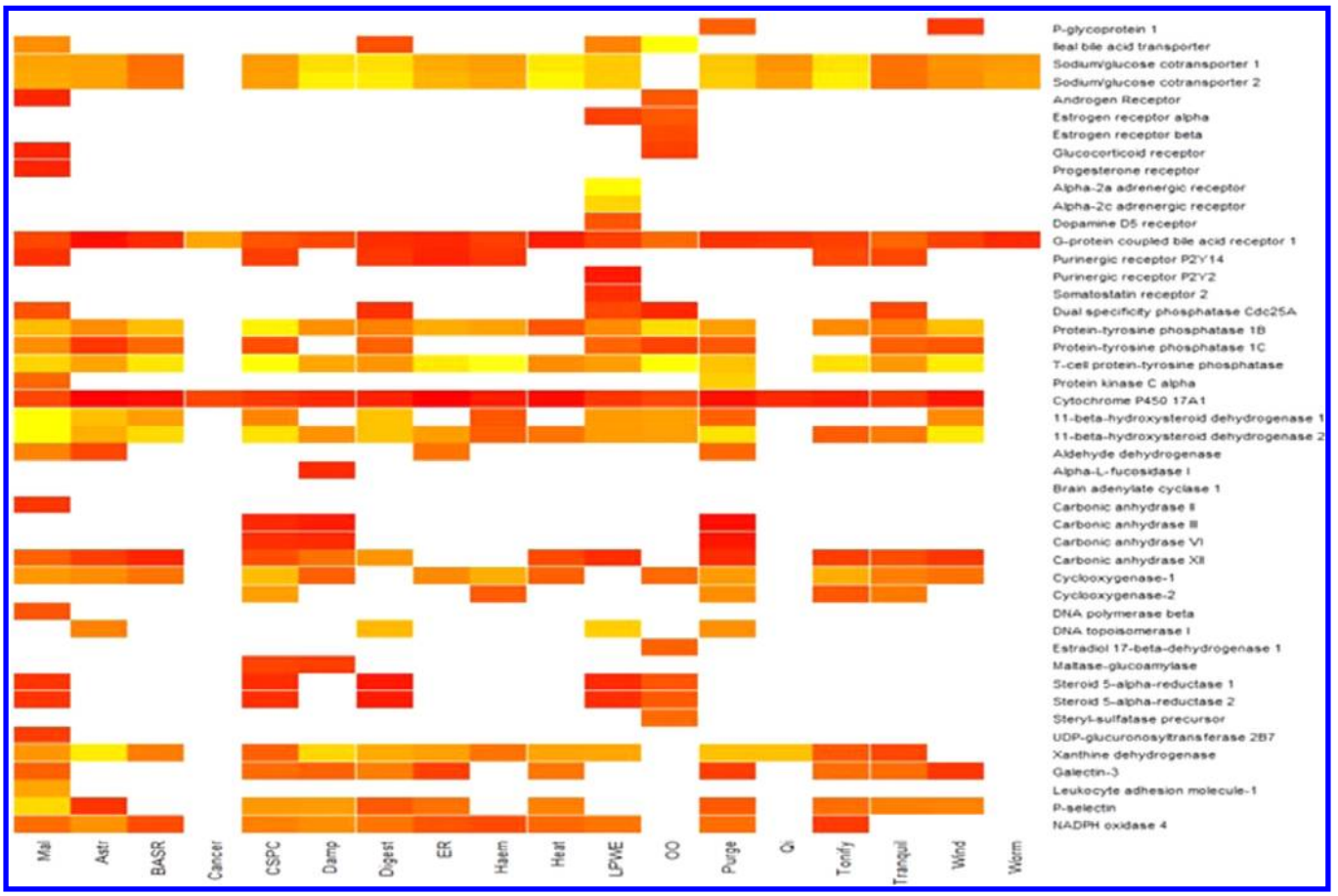


Figure 3. Correlations between classes of Traditional Medicine (bottom) and protein targets representing MOA hypotheses (right). The correlation between the different classes and enriched protein target can then rationalize the action of formulations used in traditional medicine. Highly enriched class-target associations are marked in yellow, less enriched associations in red, and not enriched associations in white. For full information on the heat map, please see the main text and Supporting Information S1. Abbreviations: Mal: antimalarial, Astr: astringent, BASR: blood-activating stasis-resolving, CSPC: cough suppressing and panting calming, Damp: dampness-resolving, Digest: digestant, ER: exterior releasing, Haem: hemostatic, Heat: heat clearing, LPWE: liver-pacifying and wind-extinguishing, OO: orifice-opening, Purge: purgative, Qi: qi regulating, Tonify: tonifying and replenishing, Tranquil: tranquilizing, Wind: wind-dampness dispelling, Worm: worm expelling medicine.

development as a drug. Out of these challenges, in particular, the elucidation of MOA[16] is of relevance to the current work, hence making it easier to accept traditional medicines as a source of practically relevant bioactive compounds. Despite not knowing the MOA, the phenotypes or biological activities of compounds used in both TCM and Ayurveda are well-known, and the current work aims to deconvolute the protein targets of traditional medicine formulations. Given the recent availability of databases that provide chemical structures of the compounds used in both TCM[17] and Ayurveda[18] and its phenotypic readouts, we can now perform an *in silico* analysis of these medicinal compounds in order to establish their MOA.

Phenotypic information is vital as it indicates the therapeutic impact and efficacy of a compound against a certain disease,[19] and we aim to tackle the problem of yet unknown MOA through the use of an *in silico* target prediction recently established in our group. (For a more comprehensive review on *in silico* target prediction please see a recent review article.[20]) Given a chemical structure of a compound, this algorithm predicts potential targets modulated by the compound (Figure 2). Subsequently, the targets predicted enable the rationalization of the MOA of compounds. Previous studies where a similar approach was successful include other phenotypic readouts such as from high-content screening[21] and adverse drug reactions.[22] Also target prediction approaches have been applied to traditional medicines such as the recent effort by Zhang et al. where the inverse docking method was used to identify potential targets for 19 natural products derived from two medicinal plants used as anti-inflammatory (*Daphne odora* Thunb. var. *marginata)* and antidiabetic (*Bacopa monnieri* (L.) Wettst) agents.[23] It was found that the dipeptidyl peptidase (DPP-IV), a known diabetic target, was the target most frequently predicted.[23] This was then confirmed using an *in vitro* enzyme assay where five natural products showed inhibitory activities with IC_{50} values between 14.13 μM to 113.76 μM against DPP-IV.[23] Ehrman et al.[24] used pharmacophore-based docking, LigandScout, to look for possible TCM herbs from the Chem-TCM database[25] that may bind against targets that includes p38 MAP kinase (p38), cyclooxygenases 1 and 2 (COX), type 4 cAMP-specific phosphodiesterase (PDE4), and c-Jun terminal-NH2 kinase (JNK). The aim of this study was to find potential multitargeted anti-inflammatory drugs.[24] It was concluded from this study that a ligand that is multitargeted is commonly found in TCM herbs, as 48% of the herbs analyzed were found to bind to at least two targets.[24]

dx.doi.org/10.1021/ci3005513 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

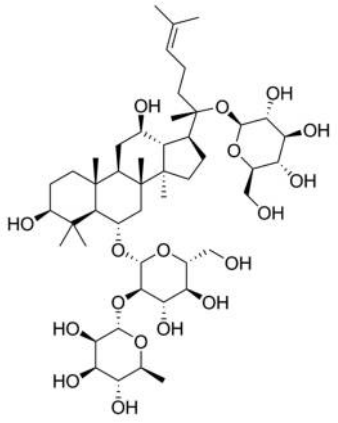
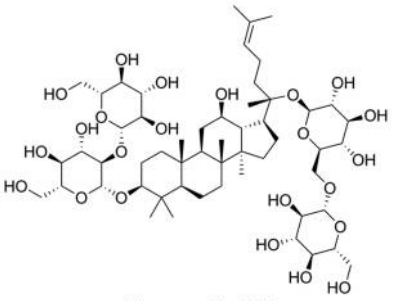
Table 1. Protein Targets Predicted To Rationalize the Activity of Formulations Used As 'Tonifying and Replenishing Medicines', Sorted in Decreasing Order of Normalization Rate (NR)[a]

target	TF	NR	description of target	scientific literature linking target to phenotypes of the tonifying and replenishing medicinal class
sodium/glucose co-transporter 2 (SGLT2)	534	35.17	SGLT2 transport sodium and glucose in the kidney.[34] Inhibition of SGLT2 has been implicated as a possible treatment for type 1 and 2 diabetes and/or obesity as this increases the loss of glucose and energy in the urine.[34]	• Found to lower fasting plasma levels of glucose and insulin, improve oral glucose tolerance, and increase glucose-insulin index.[31]
T-cell protein-tyrosine phosphatase (TC-PTP)	262	33.56	TC-PTP is a protein tyrosine phosphatase (PTP) that phosphorylates of protein tyrosine kinase (PTK) such as p42/44 MAPK (ERK).[35] Dysregulation of tyrosine phosphorylation of even a single PTK can lead to oncogenic transformation.[35]	• Antimetastatic, antiproliferative, and antitumor have been suggested as one of the therapeutic actions of this class.[32]
sodium/glucose co-transporter 1 (SGLT1)	538	33.52	SGLT1 is primarily located in the small intenstine cells.[34] Here, SGLT1 acts as a sodium-glucose/galactose transporter.[34]	• Found to lower fasting plasma levels of glucose and insulin, improve oral glucose tolerance, and increase the glucose-insulin index.[31]
cyclooxygenase-1 (COX-1)	216	24.90	COX-1 converts arachidonic acid into pro-inflammatory prostaglandins.[36] COX-1 is implicated in inflammation-related diseases and cancer.[36]	• Found to have an anti-inflammatory effect.[30]
protein-tyrosine phosphatase 1B (PTP1B)	170	18.66	PTP1B is a negative regulator of insulin signaling by dephosphorylating the phosphotryosine residues of insulin receptor kinase. PTP1B also dephosphorylate PTK such as EGFR.[35]	• Found to lower fasting plasma levels of glucose and insulin, improve oral glucose tolerance, and increase glucose-insulin index.[31] • Antimetastatic, antiproliferative, and antitumor have been suggested as one of the therapeutic actions of this class.[32]
P-selectin	125	14.41	P-selectin has been found to play a role in artherogenesis, atrial fibrillation, hypertension, and coronary artery disease.[37]	• Found to produce a hypotensive effect in anaesthetized rats.[33]
galectin-3	180	14.31	This protein plays a role in several cellular functions including innate immunity, apoptosis, T-cell regulation, and cell adhesion.[38] Galectin-3 has been implicated in the regulation of inflammation and immune response[38] and as a possible cancer target.[39]	• Found to have an anti-inflammatory effect.[30] • Immunomodulatory has been suggested as one of the therapeutic actions of this class.[29]
11-beta-hydroxysteroid dehydrogenase 2	70	11.52	Mutations of this gene can cause the syndromes associated with hypertension and apparent mineralocorticoid excess.[40]	• Found to produce a hypotensive effect in anaesthetized rats.[33]
xanthine dehydrogenase	78	10.57	Xanthine dehydrogenase converts xanthine into uric acid.[41] High levels of uric acid can lead to gout.[41]	
cyclooxygenase 2 (COX-2)	78	10.57	COX-2 converts arachidonic acid into pro-inflammatory prostaglandins.[36] COX-2 is implicated in inflammation-related diseases and cancer.[36]	• Found to have an anti-inflammatory effect.[30]
purinergic receptor P2Y14	172	8.44	P2Y receptors participates in the stem cell compartment regulation and neuroimmune function and have been associated with MAPK upon activation.[42]	• Found to prevent death in neuronal cells.[43] • MAPK has been identified as an oncogene,[44] and antimetastatic, antiproliferative, and antitumor have been suggested as one of the possible therapeutic actions.[32]
G-protein coupled bile acid receptor 1 (GPBAR1)	357	6.85	GPBAR1 binds to bile acid.[45] Upon binding, the active hormone triiodothyronine (T_3) is produced.[46] In turn, T_3 activates thyroid hormone receptors which increases metabolic rate.[46]	
carbonic anhydrase XII (CA12)	132	6.62	CA12 regulates the acidity of the microenvironment in cancer cells which modulates tumor malignant phenotype.[47] CA12 was found to be overexpressed in renal cancer cells[48] and breast cancer cells.[47]	• Antimetastatic, antiproliferative, and antitumor have been suggested as one of the therapeutic actions of this class.[32]
NADPH oxidase 4	66	6.09	This protein regulates signaling cascade of insulin, bone resorption, apoptosis, and lipopolysaccharide-mediated activation of NF-κB[49] and may produce superoxide in the nucleus.[50] NADPH oxidase is also involved in artherosclerosis.[49]	• Found to suppress nuclear transcription factor NF-κB activation in lipopolysaccharide-stimulated mouse macrophage cell line.[30] • Found to act as a superoxide scavenger.[51]
cytochrome (CYP) P450 17A1	96	1.77	A key enzyme in the steroidogenic pathway which produces mineralcorticoids, glucocorticoids, estrogen, progestins, and androgens.[52] CYP450 17A1 is a cancer target for castration-resistant prostate cancer.[52]	• Antimetastatic, antiproliferative. and antitumor have been suggested as one of the therapeutic actions of this class.[32] • Found that *Cordycep sinensis*, one of the Cordyceps species, reduces the production of human chorionic gonadotropin-stimulated testosterone in mice through the inhibition of cytochrome P450 scc.[53]

[a]As can be seen, predicted targets are both supported by published literature and cover a wide range of different aspects of bioactivities in this class. (TF = Target Frequency).

dx.doi.org/10.1021/ci300551t | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. Target(s) Predicted for Some of the Active Ingredients of *Panax ginseng*[a]

Structure	Predicted Target(s)	Literature Support
 Ginsenoside Re	• SGLT1 • SGLT2 • TC-PTP	• This ginsenoside was found to have a hypoglycemic effect.[55] All three targets have been implicated in the pathogenesis of diabetes although for TC-PTP, the evidence is not sufficient yet.[35]
 Ginsenoside Rb1	• PTP1B • TC-PTP • COX-2	• Ginsenoside-Rb1 was found to completely block PTK activation in human umbilical vein endothelial cells at a wide range of concentration.[56] Both PTP1B and TC-PTP are involved in the phosphorylation of PTK.[35] • Ginsenoside-Rb1 have been found to suppress COX-2 expression.[57]

[a]Predicted targets were corroborated by published literature (see table), and, in addition, the targets predicted were rather different for the two ginsenosides despite their structural similarity (Tanimoto coefficient of 0.6702 using ECFP_4 as descriptor). This is in some sense surprising, given the focus of our method on chemical features shared between structures (which is rather large in the cases shown here) where only additional features, the sugar rings, of the same sort are added or removed, respectively.

RESULTS

To view the relationship between the targets and the 20 different activity classes from both traditional medicines, we first generated a heat map between both domains. Figure 3 shows the heat map of the enriched targets predicted for each activity class when a target prediction score of 30 was used. (In effect only considering targets with a more that ca. 80% likelihood of being true target proteins of the compounds considered; see the Methods section for more details.) The color spectrum of the heat map represents normalized target enrichments, where red bands indicate low target enrichments, while yellow bands indicate high enrichments. As can be seen in Figure 3, not all of the 20 phenotypic classes were included. Here, the class "parasitic elimination, dampness reduction and itchiness relief medicine" and "interior warming medicine" were excluded as the second cut off for target selection (of Target Frequency (TF) $\geq$ 5% of total compounds; see the Methods section for details) limits the number of protein target predicted for these classes. No targets were retained for the class "interior warming medicine", while only a single target was retained for the class "parasitic elimination, dampness reduction and itchiness relief medicine". Hence, those classes were excluded from Figure 3.

The rationalization of the MOA can then be done by analyzing the link between the different classes of the Traditional Medicine (bottom of Figure 3) and the protein targets predicted for the classes (right side of Figure 3). Out of the classes presented we will discuss 'tonifying and replenishing medicine' (abbreviated as 'Tonify') and Ayurvedic anticancer drugs (abbreviated as 'Cancer') in more detail.

Tonifying and Replenishing Medicine (TCM). The therapeutic class of 'tonifying and replenishing medicine' is responsible for the maintenance of health and to delay the onset of senescence.[29] In TCM, senescence are the changes that the body goes through with time such as that the body is more vulnerable to diseases.[29] The suggested therapeutic actions (or phenotypes and will be used from here on) for this class are as follows:

- Anti-inflammatory[30]
- Antioxidant, neuroprotective, and antiaging activity[29]
- Hypoglycemic activity and effect on the secretion of insulin[31]
- Immunomodulatory[29]
- Antimetastatic, antiproliferative, and antitumor[32]
- Hypotensive[33]

The description of the targets predicted for this class can be seen in Table 1.

One important observation can be made from Table 1: the targets predicted for this class can be associated with the phenotypes reported. To illustrate targets predicted to be

dx.doi.org/10.1021/ci3005513 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 3. Top 10 Predictions for Anticancer Drugs of Ayurvedic Origin When a Cut off of 10 Was Used[a]

target name	TF	NR	notes
ryanodine receptor 1 (RyR1)	87	26.79	RyR1 mediates the release of calcium in the sarcoplasmic reticulum, and the release of calcium can influence cancer pathogenesis through cell cycle progression, angiogenesis, apoptosis, and tumor metastasis.[58]
TC-PTP	115	23.98	This target is a negative regulator of p42/44 MAPK (ERK), Janus kinases (JAKs), signal transducer and activator of transcription (STATs), epidermal growth factor receptor (EGFR), and insulin receptor ß.[35]
CYP P450 17A1	322	18.86	CYP450 17A1 is a cancer target for castration-resistant prostate cancer.[52]
GPBAR1 1	271	18.74	GPBAR1 binds to bile acid.[45] Upon binding, the active hormone triiodothyronine (T_3) is produced.[46] In turn, T_3 activates thyroid hormone receptors which increases metabolic rate.[46]
steroid 5-alpha-reductase 1	202	15.50	This enzyme has been found to be involved in the progression of prostate cancer and is overly expressed in the malignant state of prostate cancer.[59]
progesterone receptor (PR)	223	14.73	This receptor is an important therapeutic target and has been targeted in the treatment of breast cancer and endometrial hyperplasia.[60]
steroid 5-alpha-reductase 2	159	14.62	This enzyme has been found to be involved in the progression of prostate cancer and thought to be more involved in the benign state rather than the malignant state.[59]
P-glycoprotein 1	86	14.60	P-glycoprotein is overly expressed in cancer cells causing drugs to be pumped out of the cells faster than they can enter, preventing therapeutic doses sufficient to kill the cancerous cells.[61]
estradiol 17-beta-dehydrogenase 1	185	14.24	This enzyme is overly expressed and shown to have significance in the prognosis of hormone-dependent breast cancer, leiomyoma, and endometriosis.[62]
PTP1B	76	12.66	PTP1B dephosphorylate kinases such as receptor PTK (such as EGFR), intracellular PTKs (such as Jak2, Tyk2), transcription factor (STAT5a, STAT5b), and adapter proteins (such as Crk)[35]

[a]Targets are sorted in decreasing order of normalization rate (NR). (TF = Target Frequency.) As can be seen in this table, primary cancer targets (that is directly involved in cancer pathogenesis) and synergistic targets (such as P-gp that increases the intracellular concentration of anticancer drugs) were predicted. In addition, it was also noted that the majority of the targets predicted were hormone-dependent cancer targets such as steroid-5-alpha-reductase 1 and 2, PR, CYP450 17A1, and estradiol 17-beta-dehydrogenase 1.

involved in the MOA of formulations from this class we highlight frequently predicted targets as follows:

i. *Cytochrome P450 17A1 (CYP450 17A1).* CYP450 17A1 (also known as 17α-hydroxylase/$C_{17,20}$-lyase) is involved in the pathways that catalyze testosterone from pregnenoline and progesterone.[52] Testosterone has been found to participate in the pathogenesis of prostate cancer.[52] The drug Abiraterone inhibits CYP450 17A1 and was approved for the treatment of castration-resistant prostate cancer.[52] This target can be linked to the antimetastatic, antiproliferative, and antitumor property reported for this class.[32]

ii. *Protein tyrosine phosphatase 1B (PTP1B).* PTP1B is involved in the dephosphorylation of activated insulin receptor kinase.[35] Studies have shown that PTP1B negatively regulates insulin receptor signaling, and inhibition of PTP1B may improve insulin sensitivity in people with type II diabetes.[54] This shows that the hypoglycemia activity and the effect of insulin secretion of this class can be linked to PTP1B.

In some cases, the targets predicted may also point us to novel MOA and potential side effects for this class, most notably the following:

i. Antigout action. Xanthine dehydrogenase is blocked by the drug Febuxostat, which is used to treat gout.[41] By blocking xanthine dehydrogenase, xanthine cannot be converted to uric acid.[41]

ii. Effect on metabolism as a potential side effect by binding to GPCR bile acid receptor (GPBAR1). GPBAR1 is a G-protein coupled receptor (GPCR) that binds to bile acid.[45] Upon binding Triiodothyronine (T_3) is produced which activates thyroid hormone receptors and increases metabolic rate.[46] This effect might contribute to the 'vitalizing' effects of this compound class.

Hence, overall we can see that the *in silico* target prediction can predict protein targets involved in the MOA of the formulations which are supported by empirical literature as well as covering a wide variety of biological activities.

We then looked at the targets predicted for two active constituents of *Panax ginseng* explicitly, the result of which is displayed in Table 2.

What can be observed is that the targets predicted were relevant to the phenotypes of this class as well as being supported by literature evidence. For example, Ginsenoside-Re was predicted to bind to SGLT1 and 2 and TC-PTP, which have been found to partake in glucose reuptake and activate the insulin receptor-β respectively. Ginsenoside-Re was found to show hypoglycemic activity,[55] and inhibition of these three targets has been found to cause this phenotype. However, the evidence supporting the involvement of TC-PTP in diabetes is not as sufficient[35] as SGLTs at this stage.

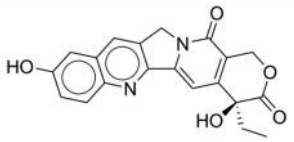
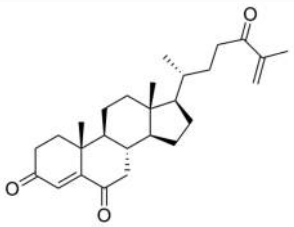
Anticancer Drugs (Ayurveda). While protein targets related to TCMs were in many cases related to the indications where they were used, there was also the aim of investigating other areas of traditional medicine, namely Ayurvedic drugs by analysis of formulations used against cancer. The ten most enriched targets predicted for this class are displayed in Table 3.

Two key observations can be made from Table 3. First, in a similar manner to the 'tonifying and replenishing medicinal' class, the targets enriched in this class can be connected to cancer progression such as PTP1B and T-cell Protein Tyrosine Phosphatase (TC-PTP) which share 70% of their amino acid sequence.[35] Both of these targets belong to the protein tyrosine phosphatase (PTP) family.[35] PTP is involved in the phosphorylation of protein tyrosine kinase (PTK).[35] Dysregulation of the phosphorylation of PTK can lead to oncogenic transformation.[35] It was found that inhibition of PTP1B significantly delays tumorigenesis in ErbB2-induced breast cancer in mice, suggesting inhibitors of PTP1B can potentially treat breast cancer.[63] TC-PTP, which is the most highly enriched among known cancer targets in Table 3, negatively regulates proinflammatory mediators (e.g., IFN-γ and TNF-α) that promote metastasis, tumor invasion, and tumorigenesis.[64] In addition, TC-PTP negatively regulates STAT1, STAT3, and STAT5a/b and Jak1 and Jak3.[64] Hyperactivity of STATs has been observed in both acute and chronic leukemia.[65] Hematopoietic malignancies often show impaired regulation of Jak.[66]

Second, the target prediction algorithm predicts not only primary targets involved in cancer progression but also

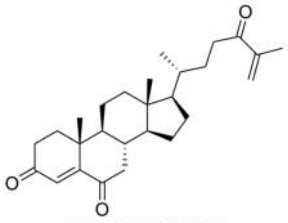
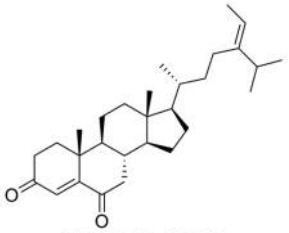
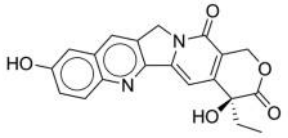
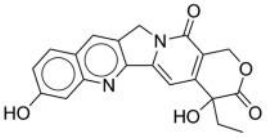
dx.doi.org/10.1021/ci3005513 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 4. Target(s) Predicted for Two of the Compounds Known To Have Anticancer Activity[a]

Structure	Predicted Target(s)	Literature Support
 CHEMBL273862	• DNA Topoisomerase 1 • Estrogen Receptor-β (ER-β)	• DNA Topoisomerase facilitates the separation of the DNA strand for the purpose of replication and transcription.[78] Camptothecin, an anti-cancer drug discovered from the Chinese tree, *Camptotheca acuminata* blocks this enzyme.[79] • ER-β was found to be expressed in breast cancer with ER-α and PR.[80]
 CHEMBL463810	• Steroid 5-alpha-reductase 1 • Steroid 5-alpha-reductase 2	• Both enzymes have been found to be involved in the progression of prostate cancer.[59] The concentration of Steroid 5-alpha-reductase 1 have been found to be higher in malignant prostate cancer whereas Steroid 5-alpha-reductase 2 was more pronounced in the benign state.[59]

[a]It can be seen that the targets predicted for two of the compounds in this class are consistent with the information in the literature which suggests they are involved in cancer pathogenesis.

Table 5. Compounds from the ChEMBL Database[26] That Are Most Similar to Compounds in Table 4 and Their Activity Profiles[a]

Reference Compound	Closest_Similarity	Note
 CHEMBL463810	 CHEMBL 527206 $T_c = 0.9538$	This compound has been found to exhibit cytotoxic effects on human KB squamous cell carcinoma lines (ED50=0.4 μg ml^{-1}), human A549 lung carcinoma cell lines (ED50=1 μg ml^{-1}) and human HT-29 colon adenocarcinoma cell lines (ED50=0.63 μg ml^{-1}) [82]
 CHEMBL273862	 CHEMBL 301951 $T_c = 0.9945$	This compound showed anti-tumor activity against P-338 leukemic cell in mice.[83]

[a]As can be seen, the activity profile shows that the most similar compounds exhibit anticancer properties, which is consistent with the indicated class of anticancer compounds analyzed. (Tc is Tanimoto coefficient, a measure of structural similarity.)

synergistic targets such as the efflux pump, P-glycoprotein (P-gp). P-gp is overly expressed in cancer cells causing drugs to be pumped out of the cells quicker than they can enter[61] and coadministering efflux pump inhibitors together with the compound showing the primary activity will increase intracellular compound concentration and hence lead to cytotoxicity of cancer cells, often more efficiently than without administration of the inhibitor. This finding underlines the hypothesis that traditional medicines rely on synergistic compound action of their ingredients.[6] For example, the antimicrobial action of berberine is potentiated by 5′-methoxyhydnocarpin, a multidrug pump resistant (MDR) inhibitor,[67] while the synergistic interactions among tetraarsenic tetrasulfide, indirubin, and tanshinone IIA contribute to the effective treatment of promyelocytic leukemia by Realgar-*Indigo naturalis* formula.[68] It is expected *via* follow-up work on pathway annotations that more synergistic compound effects will be identified in the future.

For targets like RyR1 and GPBAR1, which do not represent well-established oncoproteins, there has been some literature supporting their involvement in cancer pathogenesis. Inhibition of RyR1 has been found to be useful in treating prostate cancer,[69−71] and expression of RyR has been found to be a reliable prognostic marker for breast cancer.[72] In addition, Dantrolene, a muscle relaxant that blocks RyR1, has been found to be a potent inducer of apoptosis in B-cell lymphoma[73] and was also found to suppress TNF-induced apoptosis in hepatoma cells.[74] In regards to GPBAR1, bile acids have been

dx.doi.org/10.1021/ci3005513 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 6. 20 Phenotypic Classes from Traditional Medicines Used in This Study

phenotypic class	no. of compds	origin	definition according to traditional medicine (approximate)	Western equivalent (approximate)
antimalarial medicine	31	TCM		antimalarial
astringent medicine	239	TCM	Stops abnormal discharge of fluids and other substances from body i.e. blood, sweat, sputum, urine, stool, and vaginal discharges.[a]	astringent, endocrine agent[25]
blood activating and stasis resolving medicine	949	TCM	Treats disorders related to blood flow and involve relieving blood stasis and arresting bleeding.[a]	
cough suppressing and panting calming medicine	634	TCM	Disharmony of bodily fluids can produce either external visible phlegm i.e. sputum secreted by the respiratory tract or internal invisible phlegm formed by dysfunction of the spleen and lungs. This medication resolves phlegm and prevents it from deteriorating so as not to produce internal phlegm.[a]	expectorant, antitussive, anti-inflammatory, sedative, decongestant[25]
dampness resolving medicine	680	TCM	Dampness reduces the efficiency of the internal organs. It causes tiredness, sluggishness, pain in joints, and heavy limbs. Dampness can be due to environmental factors or due to an impaired spleen attributed to poor diet or emotional distress. This medicine promotes drying, excretion of fluids, purging, and increase in urination.[a]	diuretic[25]
digestant medicine	140	TCM	Promote digestion and eliminates food retention.[a]	
exterior releasing medicine	782	TCM	Treats disharmony of the superficial part of the body through inducing releasing muscles, sweating, or promoting eruption. In TCM, the body's surface forms the first barrier against invading pathogens and is dominated by the lung and bladder meridians. When pathogens attack the body, a so-called "exterior syndrome" manifests which is characterized by headache, fever, chills, generalized aching, and a floating pulse. The function of this medicine is to dispel the pathogens from the exterior and from going further into the body.[a]	diaphoretic, antiviral, antibacterial[25]
heat clearing medicine	1690	TCM	In nature, heat causes expansion and overactivity. When there is excessive heat, this can cause inflammation, fever and irritability, ulcers, excessive sweating, and excessive thirst. The main ingredients of this medicine are herbal with cooling properties.[a]	detoxicant, anti-inflammatory, antimicrobial, antiviral, diuretic[25]
hemostatic medicine	575	TCM	Treat disorders related to blood flow and involve relieving blood stasis and arresting bleeding.[b]	
interior warming medicine	404	TCM	When cold attacks the internal organs, this will decrease the activity of the organs. Symptoms include chills, increase urine output, fatigue, and slow and deep pulse. To counter this, the medicine warms the body and eliminates the cold.[a]	
liver pacifying and wind extinguishing medicine	87	TCM	Wind is the cause of the majority of sickness according to TCM. In nature, wind is constantly moving and changing. Because of that, wind can carry pathogens around the body hence causing sickness. Internal wind is associated with an imbalance liver and blood deficiency in the liver. The deficiency causes malnutrition of muscles and tendons consequently producing wind that can manifests as numbness, tremors, and spasms of muscles and tendons.[b]	
orifice opening medicine	37	TCM	Closing of the sensory orifices can lead to epilepsy, delirium, convulsions, and unconsciousness.[a]	
parasites elimination, dampness reduction and itchiness relief medicine	62	TCM	In TCM, dampness is the main cause of pathogen-related diseases. Itchiness in turn is a manifestation of the invasion of pathogens.[b]	antimicrobial, antipyretic, anti-inflammatory[25]
purgative medicine	254	TCM	Breaks down internal accumulations through vigorous bowel evacuation to not only act as a laxative but also purging out toxins and heat.[a]	laxative
qi-regulating medicine	660	TCM	In TCM, qi is an important entity to ensure the health and vitality of the body, just like blood. Qi has constantly been associated with energy.[65] Any interruption of the flow of qi in the body can cause sickness, including stagnation and overflowing. The lung and stomach are associated with qi.[a]	digestive stimulant, circulatory stimulant, analgesic[25]
tonifying and replenishing medicine	1335	TCM	Enriches and replenish yin, yang, blood, and qi of the body when deficient and weak. Deficiency in these results from the body experiencing a depletion of basic substances or insufficient life force.[a]	endocrine agent, antidiuretic, antihypertensive, anticholesterolaemic, immunostimulant[25]
tranquilizing medicine	113	TCM	To tranquilize the mind and relieve palpitation.[a]	tranquilizer, sedative, nerve tonic[25]
wind-dampness dispelling medicinal	408	TCM	As mentioned dampness causes a reduction in the efficiency of organs. When deposited in joints it causes pain. Coupled with the wind, the pain is transferred to different parts of the joints. The symptoms mimic the symptoms of arthritis.[b]	antirheumatic, analgesic, antipyretic, anti-inflammatory, anticoagulant[25]
worm expelling medicine	81	TCM		anthelmintics
Indian Cancer Database	560	Ayurveda		anticancer[18]

[a]Traditional Chinese Medicine (http://www.shen-nong.com/eng/front/index.html) (accessed 24 October 2011). [b]Traditional Chinese Medicine Basics (http://www.tcmbasics.com) (accessed 24 October 2011).

dx.doi.org/10.1021/ci300513t | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

implicated in affecting the signaling pathways involved in apoptosis and cell proliferation by binding to GPBAR1. It was found that inhibition of GPBAR1 may be a new strategy against preventing upper gastrointestinal carcinogenesis[75] and may also prevent the transformation of Barrett's esophagus to esophageal adenocarcinoma.[76] As mentioned in the previous section, the prediction of GPBAR1 also suggest that effects upon metabolism may be a potential side effect as upon activation through the binding of bile acid, T_3 is produced, activating thyroid hormone receptors and increasing the metabolic rate.[46]

In a similar vein to the previous section, we explicitly investigated targets predicted for individual compounds from this class (Table 4). We noted from both Table 3 and Table 4, the anticancer drugs seem to have a preference toward hormone-dependent cancer targets such as steroid-5-alpha-reductase 1 and 2,[59] the progesterone receptor,[60] CYP450 17A1,[52] and estradiol 17-beta dehydrogenase 1.[62] To illustrate further, from Table 4, it can be seen that CHEMBL463810 was predicted to bind to both steroid-5-alpha-reductase 1 and 2. The reason for this is that CHEMBL463810 resembles the structure of the natural ligand of these targets, which is testosterone (both contain a steroid core with a similarity value, using the Tanimoto coefficient, of 0.38). Steroid-5-alpha-reductase metabolizes testosterone to dihydrotestosterone which then binds to androgen receptors, the receptor involved in prostate cancer pathogenesis.[77]

We then looked at the most similar compounds from the ChEMBL database[26] and their bioactivity profiles in the examples from Table 4 (Table 5). This is performed to validate the results generated, as according to the "chemogenomics principle", similar targets share similar ligands.[81] To analyze this, we looked at CHEMBL463810 in Table 5 where the compound most similar to CHEMBL463810 is CHEMBL 516316 and has been shown to exhibit cytotoxic effect in human KB squamous cell carcinoma lines (ED_{50} = 0.4 μg mL^{-1}), human A549 lung carcinoma cell lines (ED_{50} = 1 μg mL^{-1}), and human HT-29 colon adenocarcinoma cell lines (ED_{50} = 0.63 μg mL^{-1}).[82] Hence, the chemogenomics principle appears to hold for this study (at least for these two compounds) as both compounds showed anticancer property and are structurally similar.

DISCUSSION

From the two case studies that were performed in this study, we conclude that the targets predicted were relevant to the known phenotypes of both classes; and hence, in the wider context, *in silico* target predictions indeed provide a useful way to deconvolute MOAs of traditional medicines. In the cases presented here, for the 'tonifying and replenishing medicinal' class, the hypoglycemic phenotype observed can be connected to SGLT1, SGLT2, and PTP1B. In terms of anticancer drugs used in Ayurveda, the algorithm was able to predict both primary targets (those targets directly involved in cancer progression and modulation), such as steroid-5-alpha-reductase, as well as synergistic targets, such as the efflux pump P-gp. The targets predicted may identify novel MOAs (such as in the case of xanthine dehydrogenase which suggests involvement in gout), as well as possible side effects (such as the case of GPBAR1 that may show the possibility of the compounds involvement with metabolism).

While these results successfully connect components of formulations used in traditional medicines to MOAs, some shortcomings of the current method exist. First, our target prediction tool delivers predictions for over 800 protein targets. (See recent methodological work[84] for the complete breakdown of protein target classes covered by the target prediction tool.) While this is a large number of protein targets, it does not cover the entire human proteome. Enzymes make up the majority of the targets in the training set (62.86%) as there are more data available for this type of targets. Consequently, prediction on enzymes can be accepted with higher confidence in comparison to other targets, such as ion channels, which include only 3.4% of the target space.

Second, in the current version, an empirical cutoff was used to decide which target predictions to trust and which connections to judge as not sufficiently significant. However, the applicability domain of this model to natural products-based chemistry needs to be quantified in a more detailed manner. This is important as natural products occupy a different chemical space compared to synthetic compounds, such as a larger molecular weight and a higher number of oxygen-containing functional groups compared to synthetic compounds.[11] Apart from extending chemical space and target space contained in the model, this is the most important future methodological step to take forward in this work.

Despite the limitations, this study has shown that the target prediction algorithm can be applied to elucidate the MOA of medicinal compounds used in TCM and Ayurveda. By establishing the MOA of these compounds, the gap between Western medicine and traditional medicine can be reduced. Both the chemical structures themselves − with their link to efficacy in man − as well as the targets elucidated are then possible starting points for developing bioactive chemical agents that can be used in the drug discovery context, with a large amount of empirical evidence behind them.

MATERIALS AND METHODS

Data Sets for TCM and Ayurvedic Compounds. Compounds were obtained from the TCM Database@Taiwan[17] and the Indian Plant Anticancer Database.[18] The phenotypic classes utilized in this work are listed in Table 6.

Molecular Descriptors: Extended Connectivity Fingerprint. To represent the compounds, we used the Extended Connectivity Fingerprints with a diameter of four bonds (ECFP_4).[27] The ECFP is derived from the Morgan Algorithm.[85] The atom identifier used contains topological information on the atom that includes the number of immediate heavy atoms, the atom's mass, the atom's charge, the number of hydrogens attached to the atom, the valence minus number of hydrogen, and whether it is part of at least one ring.[27]

Target Prediction Algorithm. The target prediction tool uses a Naïve Bayes (NB) classifier as a method for classification that is based on probabilistic approaches.[28] The NB classifier can be explained using the equation below[28]

$$P(C = \omega | D = f) = \frac{P(D = f | C = \omega) P(C = \omega)}{P(D = f)}$$

Here, the probability of a compound given descriptor or atomic features f belonging to the class ω was calculated where $P(C = \omega)$ is the *a priori* probability of class ω and $P(D = f)$ is the *a priori* probability of the features, f. Both of these were ignored as the probability is the same for each class (also known as uniform prior). The key value in this equation is $P(D = f \mid C = \omega)$ which is the likelihood or conditional probability of

dx.doi.org/10.1021/ci3005513 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

the feature f given the class ω. The NB classifier estimates this probability from the training set (see later) where it is assumed that the features are independent of each other in a given class. Although this is not strictly true on the given data set, it has been shown before that the NB classifier is still an effective classifier also in cases where features are correlated.[28] For a more detailed description on the calculation of the conditional probability used here, refer to Supporting Information S2.

Training Set. In machine learning, a training set is used to e.g. deduce a predictive model or classification method from examples in the data set,[86] and this is generally distinct from a test set, held out to evaluate the performance of the model. Compounds used as the training data set were obtained from the ChEMBL (v10.0) database, which is a drug discovery database that consists of small bioactive molecules abstracted from scientific journals.[26] 155,208 bioactive molecules with experimentally defined bioactivities that cover a broad biological space (894 proteins in total) were used to train and test the model. These criteria were used in the formation of the training set: (i) the compounds in the training set had to have a K_i, K_d, IC_{50}, or EC_{50} value of at least 10 μM against human protein targets; (ii) the compound-target association had to have an assay-to-target confidence score of 8 or 9; and (iii) each target class was required to comprise at least 20 compounds associated with the target to be included in the training set. The compounds were stored as SMILES format and were converted to ECFP_4 fingerprints using Scitegic's Pipeline Pilot Student Edition v6.1.5 (Accelrys, Inc.: San Diego, CA)

Validation of the Target Prediction Tool. The performance of the target prediction tool was previously evaluated using 5-fold cross-validation.[84]

Computational Procedure. Compounds were obtained from the traditional medicine databases in SD format. To convert these to ECFP_4 fingerprints, Scitegic's Pipeline Pilot Student Edition v6.1.5 (Accelrys, Inc.: San Diego, CA) was used. Duplicates were removed. The target prediction tool was run using a Python script and for each phenotypic class:

1. Targets with a NB score of thirty or more were kept. The value of thirty was picked for the cut off score as this provides a reasonable trade-off between recall and precision. A NB at a score cutoff of thirty applied to the ChEMBL data set achieved a recall of 77% and a precision of 68%,[84] suggesting that the model is able to capture the majority of active targets without predicting too large of a large number of false positives.
2. The frequency of each target (TF) was calculated.
3. Target enrichment was calculated for each target to obtain a normalization rate (See later).
4. Another cut off was applied where TF $\geq$ 5% of the total number of compounds in that phenotypic class (See later).
5. The result was ranked according to the normalization rate.

For the anticancer drug class, a lower cut off of ten was used. By using a lower cut off, it is expected that there would be more targets to analyze for this class. The resulting targets that were identified can then be used to compare these predicted targets to the targets that, from the perspective and definition of Western medicine, the anticancer drug is known to interact.

Normalization of Enrichments. For this target prediction tool, enrichment was performed to normalize the resulting classification as it was found that the target prediction algorithm had a classification bias, i.e. some classes were predicted more often than others also on the diverse background distribution (such as classes that are larger than others). To tackle this problem, we compared the TF of compounds from a given phenotypic class against a set of randomly chosen compounds. The random compound set consists of 10,000 compounds in total obtained randomly from three different databases: ChEMBL,[26] PubChem,[88] and gdb-13.[89] Unlike the set of compounds used in the training set, no specific criteria such as the target associated with the compounds and its bioactivity concentration of the compounds were considered. The normalized frequency of a given target, t_n in each phenotypic class was calculated as follows

$$\text{Normalized Target Frequency} = \frac{(\text{TF}_{t_n}/\text{TF}_T)\text{actual}}{(\text{TF}_{t_n}/\text{TF}_T)\text{random}}$$

where TF stands for target frequency, TF_{t_n} is hence the target frequency of a particular target in the activity class considered, and TF_T is the total number of targets predicted for this class. The numerator calculates the ratio of those two numbers for the class under consideration, while the denominator calculates the equivalent ratio for the background distribution.

The second cut off used (in addition to the score cutoff mentioned above) was used to determine which targets were considered to be enriched in a particular data subset. Here we used a TF $\geq$ 5% of the total number of compounds as it was found that when predictions were ranked according to the normalization rate, in some targets (where no random compounds were predicted to bind), this target would be pushed to the top of the list. Analysis of this effect showed that when this occurred, the $\text{TF}_{\text{actual}}$ was quite low, in the range of 1 to 3. Hence, this was considered to be noise, which is inevitably present in life science data sets often due to e.g. variability in assay materials. To reduce the influence of random fluctuations in the data, a cutoff target frequency (TF) of TF $\geq$ 5% of the total number of compounds in the data set was applied.

For the anticancer drug class, a TF $\geq$ 10% of the total number of compounds was used. As there will be more targets with a lower cutoff, a more stringent normalization for this phenotypic class was used.

■ ASSOCIATED CONTENT

Supporting Information

Complete information on the heat map (S1) and detailed description of the calculation of the conditional probability used as a scoring function in the target prediction tool (S2). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (44)1223 762983. E-mail: ab454@cam.ac.uk.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Unilever, Universiti Teknologi MARA (UiTM), Malaysia, and the Ministry of Higher Education of Malaysia for funding. The authors would also like to thank Florian Roessler for his help in data extraction.

■ REFERENCES

(1) Patwardhan, B.; Warude, D.; Pushpangadan, P.; Bhatt, N. Ayurveda and traditional Chinese medicine: a comparative overview. *Evidence-Based Complementary Altern. Med.* **2005**, *2*, 465−473.

(2) Miller, L. H.; Su, X. Artemisinin: Discovery from the Chinese Herbal Garden. *Cell* **2011**, *146* (6), 855−858.
(3) Vakil, R. J. A clinical trial of *Rauwolfia serpentina* in essential hypertension. *Br. Heart J.* **1949**, *11*, 350−355.
(4) Dondorp, A. M.; Fanello, C. I.; Hendriksen, I. C. E.; Gomes, E.; Seni, A.; Chhaganlal, K. D.; Bojang, K.; Olaosebikan, R.; Anunobi, N.; Maitland, K.; Kivaya, E.; Agbenyega, T.; Nguah, S. B.; Evans, J.; Gesase, S.; Kahabuka, C.; Mtove, G.; Nadjm, B.; Deen, J.; Mwanga-Amumpaire, J.; Nansumba, M.; Karema, C.; Umulisa, N.; Uwimana, A.; Mokuolu, O. A.; Adedoyin, O. T.; Johnson, W. B. R.; Tshefu, A. K.; Onyamboko, M. A.; Sakulthaew, T.; Ngum, W. P.; Silamut, K.; Stepniewska, K.; Woodrow, C. J.; Bethell, D.; Wills, B.; Oneko, M.; Peto, T. E.; von Seidlein, L.; Day, N. P. J.; White, N. J. Artesunate versus quinine in the treatment of severe falciparum malaria in African children (AQUAMAT): an open-label, randomised trial. *Lancet* **2010**, *376* (9753), 1647−1657.
(5) Zhou, L.; Zuo, Z.; Chow, M. S. S. Danshen: An overview of its chemistry, pharmacology, pharmacokinetics, and clinical use. *J. Clin. Pharmacol.* **2005**, *45* (12), 1345−1359.
(6) Xutian, S.; Zhang, J.; Louise, W. New exploration and understanding of traditional Chinese medicine. *Am. J. Chin. Med.* **2009**, *37*, 411−426.
(7) Jayasundar, R. Ayurveda: a distinctive approach to health and disease. *Curr. Sci.* **2010**, *98*, 908−914.
(8) Patwardhan, B.; Mashelkar, R. A. Traditional medicine-inspired approached to drug discovery: can Ayurveda show the way forward. *Drug Discovery Today* **2009**, *14* (15−16), 804−811.
(9) Arrowsmith, J. Trial watch: Phase II failures: 2008−2010. *Nat. Rev. Drug Discovery* **2011**, *10* (5), 328−329.
(10) Schmidt, B. M.; Ribnicky, D. M.; Lipsky, P. E.; Raskin, I. Revisiting the ancient concept of botanical therapeutics. *Nat. Chem. Biol.* **2007**, *3* (7), 360−366.
(11) Li, J. W.-H.; Vederas, J. C. Drug discovery and natural products: End of an era or an endless frontier? *Science* **2009**, *325* (5937), 161−165.
(12) Ganesan, A. The impact of natural products upon modern drug discovery. *Curr. Opin. Chem. Biol.* **2008**, *12*, 306−317.
(13) Evans, B. E.; Rittle, K. E.; Bock, M. G.; DiPardo, R. M.; Freidinger, R. M.; Whitter, W. L.; Lundell, G. F.; Veber, D. F.; Anderson, P. S. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* **1988**, *31* (12), 2235−2246.
(14) Breinbauer, R.; Vetter, I. R.; Waldmann, H. From protein domains to drug candidates—natural products as guiding principles in the design and synthesis of compound libraries. *Angew. Chem., Int. Ed.* **2002**, *41* (16), 2878−2890.
(15) Kruger, F. A.; Overington, J. P. Global analysis of small molecule binding to related protein targets. *PLoS Comput. Biol.* **2012**, *8* (1), e1002333.
(16) Corson, T. W.; Crews, C. M. Molecular understanding and modern application of traditional medicines: Triumphs and trials. *Cell* **2007**, *130* (5), 769−774.
(17) Chen, C. Y.-C. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. *PLoS One* **2011**, *6*, e15939.
(18) Vetrivel, U.; Subramanian, N.; Pilla, K.; Campus, R. Bioinformation InPACdb - Indian plant anticancer compounds database bioinformation. *Cancer* **2009**, *2063*, 71−74.
(19) Swinney, D. C.; Anthony, J. How were new medicines discovered? *Nat. Rev. Drug Discovery* **2011**, *10*, 507−519.
(20) Koutsoukas, A.; Simms, B.; Kirchmair, J.; Bond, P. J.; Whitmore, A. V.; Zimmer, S.; Young, M. P.; Jenkins, J. L.; Glick, M.; Glen, R. C.; Bender, A. From in silico target prediction to multi-target drug design: Current databases, methods and applications. *J. Proteomics* **2011**, DOI: 10.1016/j.jprot.2011.05.011.
(21) Young, D. W.; Bender, A.; Hoyt, J.; McWhinnie, E.; Chirn, G.-W.; Tao, C. Y.; Tallarico, J. A.; Labow, M.; Jenkins, J. L.; Mitchison, T. J.; Feng, Y. Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat. Chem. Biol.* **2008**, *4*, 59−68.
(22) Scheiber, J.; Chen, B.; Milik, M.; Sukuru, S. C. K.; Bender, A.; Mikhailov, D.; Whitebread, S.; Hamon, J.; Azzaoui, K.; Urban, L.; Glick, M.; Davies, J. W.; Jenkins, J. L. Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis. *J. Chem. Inf. Model.* **2009**, *49* (2), 308−317.
(23) Zhang, S.; Lu, W.; Liu, X.; Diao, Y.; Bai, F.; Wang, L.; Shan, L.; Huang, J.; Li, H.; Zhang, W. Fast and effective identification of the bioactive compounds and their targets from medicinal plants via computational chemical biology approach. *MedChemComm* **2011**, *2* (6), 471−477.
(24) Ehrman, T. M.; Barlow, D. J.; Hylands, P. J. In silico search for multi-target anti-inflammatories in Chinese herbs and formulas. *Bioorg. Med. Chem.* **2010**, *18*, 2204−2218.
(25) Ehrman, T.; Barlow, D.; Hylands, P. Phytochemical informatics of traditional Chinese medicine and therapeutic relevance. *J. Chem. Inf. Model.* **2007**, *47*, 2316−2334.
(26) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40* (D1), D1100−D1107.
(27) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742−754.
(28) Flach, P. A.; Lachiche, N. Naive Bayesian classification of structured data. *Machine Learning* **2004**, *57*, 233−269.
(29) Ko, K. M.; Leung, H. Y. Enhancement of ATP generation capacity, antioxidant activity and immunomodulatory activities by Chinese yang and yin tonifying herbs. *Chin. Med.* **2007**, *2*, 3.
(30) Kim, K.-M.; Kwon, Y.-G.; Chung, H.-T.; Yun, Y.-G.; Pae, H.-O.; Han, J.-A.; Ha, K.-S.; Kim, T.-W.; Kim, Y.-M. Methanol extract of Cordyceps pruinosa inhibits in vitro and in vivo inflammatory mediators by suppressing NF-κB activation. *Toxicol. Appl. Pharmacol.* **2003**, *190* (1), 1−8.
(31) Zhao, C.-S.; Yin, W.-T.; Wang, J.-Y.; Zhang, Y.; Yu, H.; Cooper, R.; Smidt, C.; Zhu, J.-S. CordyMax Cs-4 improves glucose metabolism and increases insulin sensitivity in normal rats. *J. Altern. Complementary Med.* **2005**, *8* (3), 309−314.
(32) Chen, Y.-J.; Shiao, M.-S.; Lee, S.-S.; Wang, S.-Y. Effect of Cordyceps sinensis on the proliferation and differentiation of human leukemic U937 cells. *Life Sci.* **1997**, *60* (25), 2349−2359.
(33) Chiou, W.-F.; Chang, P.-C.; Chou, C.-J.; Chen, C.-F. Protein constituent contributes to the hypotensive and vasorelaxant acttvtties of cordyceps sinensis. *Life Sci.* **2000**, *66* (14), 1369−1376.
(34) Idris, I.; Donnelly, R. Sodium−glucose co-transporter-2 inhibitors: an emerging new class of oral antidiabetic drug. *Diabetes, Obes. Metab.* **2009**, *11* (2), 79−88.
(35) Stuible, M.; Doody, K.; Tremblay, M. PTP1B and TC-PTP: regulators of transformation and tumorigenesis. *Cancer Metastasis Rev.* **2008**, *27* (2), 215−230.
(36) Dubois, R. N.; Abramson, S. B.; Crofford, L.; Gupta, R. A.; Simon, L. S.; A. Van De Putte, L. B.; Lipsky, P. E. Cyclooxygenase in biology and disease. *FASEB J.* **1998**, *12* (12), 1063−1073.
(37) Blann, A. D.; Nadar, S. K.; Lip, G. Y. H. The adhesion molecule P-selectin and cardiovascular disease. *Eur. Heart J.* **2003**, *24* (24), 2166−2179.
(38) Dumic, J.; Dabelic, S.; Flögel, M. Galectin-3: An open-ended story. *Biochim. Biophys. Acta, Gen. Subj.* **2006**, *1760* (4), 616−635.
(39) Danguy, A.; Camby, I.; Kiss, R. Galectins and cancer. *Biochim. Biophys. Acta, Gen. Subj.* **2002**, *1572* (2−3), 285−293.
(40) Ferrari, P. The role of 11β-hydroxysteroid dehydrogenase type 2 in human hypertension. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2010**, *1802* (12), 1178−1187.
(41) Ernst, M. E.; Fravel, M. A. Febuxostat: A selective xanthine-oxidase/xanthine-dehydrogenase inhibitor for the management of hyperuricemia in adults with gout. *Clin. Ther.* **2009**, *31* (11), 2503−2518.
(42) Burnstock, G. Purine and pyrimidine receptors. *Cell. Mol. Life Sci.* **2007**, *64* (12), 1471−1483.

(43) Jin, D.-Q.; Park, B.-C.; Lee, J.-S.; Choi, H.-D.; Lee, Y.-S.; Yang, J.-H.; Kim, J.-A. Mycelial extract of Cordyceps ophioglossoides prevents neuronal cell death and ameliorates b-amyloid peptide-induced memory deficits in rats. *Biol. Pharm. Bull.* **2004**, *27* (7), 1126−1129.
(44) Novak, K. Conference Report—Protein Kinase Inhibitors in Cancer Treatment: Mixing and Matching? In *Proceedings of the Keystone Symposium on Protein Kinases and Cancer*, Medscape General Medicine: Lake Tahoe, CA, USA, 2004.
(45) Kawamata, Y.; Fujii, R.; Hosoya, M.; Harada, M.; Yoshida, H.; Miwa, M.; Fukusumi, S.; Habata, Y.; Itoh, T.; Shintani, Y.; Hinuma, S.; Fujisawa, Y.; Fujino, M. A G protein-coupled receptor responsive to bile acids. *J. Biol. Chem.* **2003**, *278* (11), 9435−9440.
(46) Watanabe, M.; Houten, S. M.; Mataki, C.; Christoffolete, M. A.; Kim, B. W.; Sato, H.; Messaddeq, N.; Harney, J. W.; Ezaki, O.; Kodama, T.; Schoonjans, K.; Bianco, A. C.; Auwerx, J. Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* **2006**, *439* (7075), 484−489.
(47) Barnett, D. H.; Sheng, S.; Howe Charn, T.; Waheed, A.; Sly, W. S.; Lin, C.-Y.; Liu, E. T.; Katzenellenbogen, B. S. Estrogen receptor regulation of carbonic anhydrase XII through a distal enhancer in breast cancer. *Cancer Res.* **2008**, *68* (9), 3505−3515.
(48) Türeci, Ö.; Sahin, U.; Vollmar, E.; Siemer, S.; Göttert, E.; Seitz, G.; Parkkila, A.-K.; Shah, G. N.; Grubb, J. H.; Pfreundschuh, M.; Sly, W. S. Human carbonic anhydrase XII: cDNA cloning, expression, and chromosomal localization of a carbonic anhydrase gene that is overexpressed in some renal cell cancers. *Proc. Natl. Acad. Sci.* **1998**, *95* (13), 7608−7613.
(49) Park, H. S.; Chun, J. N.; Jung, H. Y.; Choi, C.; Bae, Y. S. Role of NADPH oxidase 4 in lipopolysaccharide-induced proinflammatory responses by human aortic endothelial cells. *Cardiovasc. Res.* **2006**, *72* (3), 447−455.
(50) Kuroda, J.; Ago, T.; Matsushima, S.; Zhai, P.; Schneider, M. D.; Sadoshima, J. NADPH oxidase 4 (Nox4) is a major source of oxidative stress in the failing heart. *Proc. Natl. Acad. Sci.* **2010**, *107* (35), 15565−15570.
(51) Ko, K.-M.; Mak, D. H. F.; Chiu, P.-Y.; Poon, M. K. T. Pharmacological basis of 'yang-invigoration' in Chinese medicine. *Trends Pharmacol. Sci.* **2004**, *25* (1), 3−6.
(52) O'Donnell, A.; Judson, I.; Dowsett, M.; Raynaud, F.; Dearnaley, D.; Mason, M.; Harland, S.; Robbins, A.; Halbert, G.; Nutley, B.; Jarman, M. Hormonal impact of the 17[alpha]-hydroxylase//C17,20-lyase inhibitor abiraterone acetate (CB7630) in patients with prostate cancer. *Br. J. Cancer* **2004**, *90* (12), 2317−2325.
(53) Hsu, C.-C.; Tsai, S.-J.; Huang, Y.-L.; Huang, B.-M. Regulatory mechanism of Cordyceps sinensis mycelium on mouse Leydig cell steroidogenesis. *FEBS Lett.* **2003**, *543* (1), 140−143.
(54) Elchebly, M.; Payette, P.; Michaliszyn, E.; Cromlish, W.; Collins, S.; Loy, A. L.; Normandin, D.; Cheng, A.; Himms-Hagen, J.; Chan, C.-C.; Ramachandran, C.; Gresser, M. J.; Tremblay, M. L.; Kennedy, B. P. Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* **1999**, *283* (5407), 1544−1548.
(55) Xie, J.-t.; Wang, C.-z.; Wang, A.-b.; Wu, J.; Basila, D.; Yuan, C.-s. Antihyperglycemic effects of total ginsenosides from leaves and stem of Panax ginseng. *Acta Pharmacol. Sin.* **2005**, *26* (9), 1104−1110.
(56) Dou, D.-Q.; Zhang, Y.-W.; Zhang, L.; Chen, Y.-J.; Yao, X.-S. The inhibitory effects of ginsenosides on protein tyrosine kinase activated by hypoxia/reoxygenation in cultured human umbilical vein endothelial cells. *Planta Med.* **2001**, *67* (1), 19−23.
(57) Surh, Y.-J.; Lee, J.-Y.; Choi, K.-J.; Ko, S.-R. Effects of selected ginsenosides on phorbol ester-induced expression of cyclooxygenase-2 and activation of NF-κB and ERK1/2 in mouse skin. *Ann. N.Y. Acad. Sci.* **2002**, *973* (1), 396−401.
(58) Roderick, H. L.; Cook, S. J. Ca2+ signalling checkpoints in cancer: remodelling Ca2+ for cancer cell proliferation and survival. *Nat. Rev. Cancer* **2008**, *8* (5), 361−375.
(59) Thomas, L. N.; Douglas, R. C.; Lazier, C. B.; Too, C. K. L.; Rittmaster, R. S.; Tindall, D. J. Type 1 and type 2 5α-reductase expression in the development and progression of prostate cancer. *Eur. Urol.* **2008**, *53* (2), 244−252.
(60) Richer, J. K.; Jacobsen, B. M.; Manning, N. G.; Abel, M. G.; Wolf, D. M.; Horwitz, K. B. Differential gene regulation by the two progesterone receptor isoforms in human breast cancer cells. *J. Biol. Chem.* **2002**, *277* (7), 5209−5218.
(61) Patel, K. J.; Tannock, I. F. The influence of P-glycoprotein expression and its inhibitors on the distribution of doxorubicin in breast tumors. *BMC Cancer* **2009**, *9*, (356), doi:10.1186/1471-2407-1189-1356.
(62) Martel, C.; Rhéaume, E.; Takahashi, M.; Trudel, C.; Couët, J.; Luu-The, V.; Simard, J.; Labrie, F. Distribution of 17 beta-hydroxysteroid dehydrogenase gene expression and activity in rat and human tissues. *J. Steroid Biochem. Mol. Biol.* **1992**, *41*, 563−603.
(63) Julien, S. G.; Dube, N.; Read, M.; Penney, J.; Paquet, M.; Han, Y.; Kennedy, B. P.; Muller, W. J.; Tremblay, M. L. Protein tyrosine phosphatase 1B deficiency or inhibition delays ErbB2-induced mammary tumorigenesis and protects from lung metastasis. *Nat. Genet.* **2007**, *39* (3), 338−346.
(64) Doody, K. M.; Bourdeau, A.; Tremblay, M. L. T-cell protein tyrosine phosphatase is a key regulator in immune cell signaling: lessons from the knockout mouse model and implications in human disease. *Immunol. Rev.* **2009**, *228* (1), 325−341.
(65) Lin, T. S.; Mahajan, S.; Frank, D. A. STAT signaling in the pathogenesis and treatment of leukemias. *Oncogene* **2000**, *19* (21), 2496−2504.
(66) Khwaja, A. The role of Janus kinases in haemopoiesis and haematological malignancy. *Br. J. Hamaetol.* **2006**, *134* (4), 366−384.
(67) Stermitz, F. R.; Lorenz, P.; Tawara, J. N.; Zenewicz, L. A.; Lewis, K. Synergy in a medicinal plant: Antimicrobial action of berberine potentiated by 5′-methoxyhydnocarpin, a multidrug pump inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (4), 1433−1437.
(68) Wang, L.; Zhou, G.-B.; Liu, P.; Song, J.-H.; Liang, Y.; Yan, X.-J.; Xu, F.; Wang, B.-S.; Mao, J.-H.; Shen, Z.-X.; Chen, S.-J.; Chen, Z. Dissection of mechanisms of Chinese medicinal formula Realgar-Indigo naturalis as an effective treatment for promyelocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (12), 4826−4831.
(69) Henderson, K.; Stella, S. L., Jr.; Kobylewski, S.; Eckhert, C. D. Receptor activated Ca^{2+} release is inhibited by boric acid in prostate cancer cells. *PLoS One* **2009**, *4*, (6), doi:10.1371/journal.-pone.0006009.
(70) Mariot, P.; Prevarskaya, N.; Roudbaraki, M. M.; Le Bourhis, X.; Van Coppenolle, F.; Vanoverberghe, K.; Skryma, R. Evidence of functional ryanodine receptor involved in apoptosis of prostate cancer (LNCaP) cells. *Prostate* **2000**, *43* (3), 205−214.
(71) Flourakis, M.; Prevarskaya, N. Insights into Ca2+ homeostasis of advanced prostate cancer cells. *Biochim. Biophys. Acta, Mol. Cell Res.* **2009**, *1793* (6), 1105−1109.
(72) Abdul, M.; Ramlal, S.; Hoosein, N. Ryanodine receptor expression correlates with tumor grade in breast cancer. *Pathol. Oncol. Res.* **2008**, *14* (2), 157−160.
(73) Shamash, J.; Salam, A. H.; Davies, D. C.; Williams, A.; Joel, S.; Lister, T. A. In vitro testing of calcium channel blockers and cytotoxic chemotherapy in B-cell low-grade non-Hodgkin's lymphoma. *Br. J. Cancer* **1998**, *77* (10), 1598−1603.
(74) Kim, B.-C.; Kim, H.-T.; Mamura, M.; Ambudkar, I. S.; Choi, K.-S.; Kim, S.-J. Tumor necrosis factor induces apoptosis in hepatoma cells by increasing Ca2+ release from the endoplasmic reticulum and suppressing Bcl-2 expression. *J. Biol. Chem.* **2002**, *277* (35), 31381−31389.
(75) Yasuda, H.; Hirata, S.; Inoue, K.; Mashima, H.; Ohnishi, H.; Yoshiba, M. Involvement of membrane-type bile acid receptor M-BAR/TGR5 in bile acid-induced activation of epidermal growth factor receptor and mitogen-activated protein kinases in gastric carcinoma cells. *Biochem. Biophys. Res. Commun.* **2007**, *354* (1), 154−159.
(76) Hong, J.; Behar, J.; Wands, J.; Resnick, M.; Wang, L. J.; DeLellis, R. A.; Lambeth, D.; Souza, R. F.; Spechler, S. J.; Cao, W. Role of a novel bile acid receptor TGR5 in the development of oesophageal adenocarcinoma. *Gut* **2010**, *59* (2), 170−180.

(77) Brunton, L. L.; Blumenthal, D. K.; Buxton, I. L. O.; Parker, K. L. *Goodman & Gilman's Manual of Pharmacology and Therapeutics*; McGraw-Hill: United State of America, 2008.
(78) Pommier, Y.; Leo, E.; Zhang, H.; Marchand, C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem. Biol.* **2010**, *17* (5), 421−433.
(79) Wall, M. E.; Wani, M. C.; Cook, C. E.; Palmer, K. H.; McPhail, A. T.; Sim, G. A. Plant antitumor agents. I. The isolation and structure of camptothecin, a novel alkaloidal leukemia and tumor inhibitor from Camptotheca acuminata. *J. Am. Chem. Soc.* **1966**, *88* (16), 3888−3890.
(80) Jarvinen, T., A. H.; Pelto-Huikko, M.; Holli, K.; Isola, J. Estrogen receptor b is coexpressed with ERa and PR and associated with nodal status, grade and proliferation rate in breast cancer. *Am. J. Pathol.* **2000**, *156* (1), 29−35.
(81) Garland, S. L.; Gloriam, D. E. A ligand's view of target similarity: Chemogenomic binding-site directed techniques for drug discovery. *Curr. Top. Med. Chem.* **2011**, *11*, 1872−1881.
(82) Sheu, J.-H.; Wang, G.-H.; Sung, P.-J.; Duh, C.-Y. New cytotoxic oxygenated fucosterols from the brown Alga Turbinaria conoides. *J. Nat. Prod.* **1998**, *62* (2), 224−227.
(83) Wani, M. C.; Nicholas, A. W.; Wall, M. E. Plant antitumor agents. 23. Synthesis and antileukemic activity of camptothecin analogs. *J. Med. Chem.* **1986**, *29* (11), 2358−2363.
(84) Koutsoukas, A.; Lowe, R.; Kalantar-Motamedi, Y.; Mussa, H. Y.; Mitchell, J. B. O.; Glen, R. C.; Bender, A. In silico target predictions: comparing multiclass Naïve Bayes and Parzen-Rosenblatt Window and the definition of a benchmarking dataset for target prediction. *J. Chem. Inf. Model.* **2013**, Submitted.
(85) Morgan, H. L. The generation of a unique machine description for chemical structures. *J. Chem. Doc.* **1965**, *5*, 107−113.
(86) Gasteiger, J.; Engel, T. *Chemoinformatics: A Textbook*; Wiley-VCH: Weinheim, 2003.
(87) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection–what can we learn from earlier mistakes? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 213−228.
(88) Bolton, E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. PubChem: Integrated Platform for Small Molecules and Biological Activities. In *Annual Reports in Computational Chemistry*; American Chemical Society: Washington, DC, 2008; Vol. *4*.
(89) Blum, L. C.; Reymond, J.-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.* **2009**, *131* (25), 8732−8733.