Natural Product Reports

REVIEW

View Article Online
View Journal | View Issue

Cite this: *Nat. Prod. Rep.*, 2015, **32**, 1249

Network pharmacology applications to map the unexplored target space and therapeutic potential of natural products

Milla Kibble,*[a] Niina Saarinen,[b] Jing Tang,[a] Krister Wennerberg,[a] Sari Mäkelä[b] and Tero Aittokallio[a]

Covering: 2011 to 2014

It is widely accepted that drug discovery often requires a systems-level polypharmacology approach to tackle problems such as lack of efficacy and emerging resistance of single-targeted compounds. Network pharmacology approaches are increasingly being developed and applied to find new therapeutic opportunities and to re-purpose approved drugs. However, these recent advances have been relatively slow to be translated into the field of natural products. Here, we argue that a network pharmacology approach would enable an effective mapping of the yet unexplored target space of natural products, hence providing a systematic means to extend the druggable space of proteins implicated in various complex diseases. We give an overview of the key network pharmacology concepts and recent experimental–computational approaches that have been successfully applied to natural product research, including unbiased elucidation of mechanisms of action as well as systematic prediction of effective therapeutic combinations. We focus specifically on anticancer applications that use *in vivo* and *in vitro* functional phenotypic measurements, such as genome-wide transcriptomic response profiles, which enable a global modelling of the multi-target activity at the level of the biological pathways and interaction networks. We also provide representative examples of other disease applications, databases and tools as well as existing and emerging resources, which may prove useful for future natural product research. Finally, we offer our personal view of the current limitations, prospective developments and open questions in this exciting field.

Received 14th January 2015

DOI: 10.1039/c5np00005j

www.rsc.org/npr

1 Introduction

Natural products have been used for centuries as medicinal treatments, for disease prevention and as a source of leads for the development of drugs, with pharmacognosy dominating rational drug development until the gradual emergence of target-based drug discovery over the last fifty years.[1–3] Since the turn of the century, the current paradigm of developing highly selective ligands as therapeutics has been challenged mainly due to high late-stage clinical attrition rates which can be largely attributed to lack of efficacy and clinical safety and toxicity.[4,5] Over this same period, the advent of experimental 'omic' technologies and computational modelling of biological pathways and molecular interactions (network biology) combined with the observation that many approved drugs appear to work by modulating multiple nodes of these networks (polypharmacology) has resulted in an alternative systems-level approach to finding new drug candidates. Instead of looking for a single disease-causing gene and drugs which act solely on the appropriate individual target, the whole disease network is considered with the aim of finding multiple nodes which can be modulated *via* multi-target drugs or drug combinations in order to perturb robust disease phenotypes whilst exerting fewer side effects[6] (thus replacing the concept of a 'magic bullet' with one of 'magic shrapnel'[7]). This so-called network pharmacology paradigm[8] invokes the idea that in certain cases, to have an

[a]*Institute for Molecular Medicine Finland (FIMM), Biomedicum Helsinki 2U, 00014 University of Helsinki, Finland. E-mail: milla.kibble@helsinki.fi*

[b]*Institute of Biomedicine, Turku Center for Disease Modeling & Functional Foods Forum, 20014 University of Turku, Finland*

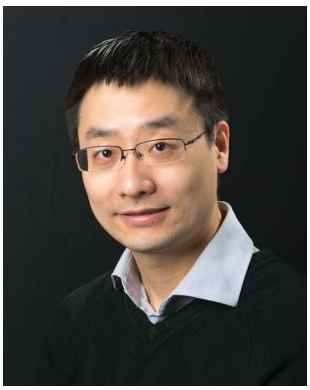
effect, drugs must target multiple pathways and/or work synergistically with other drugs, potentially reducing side-effects in the process through, for example, reduced dose of individual agents.[9] Through systematic mapping of the target interactions behind the globally-measured treatment responses, network modelling provides also system-level insights into the molecular targets of the lead candidates, which is an essential prerequisite for the phenotypic-based drug discovery process.[3]

Interestingly, the network pharmacology concept naturally links back to pharmacognosy. The concept of using mixtures of plants ('botanical drugs') to cure is very popular in traditional phytotherapy.[1] For example, formulae of Traditional Chinese Medicine (TCM) often consist of combinations of herbs where multiple active phytochemical components may theoretically target multiple targets/pathways. For example, the botanical drug PHY906 has been used for more than 1500 years and comprises four herbs each with a distinct pharmacological profile. Experiments in a mouse model of colon cancer have suggested that PHY906 reduces the gastrointestinal toxicity of irinotecan-based chemotherapy *via* multiple synergistic anti-inflammatory mechanisms.[10] Although utilized in many cases for centuries, the mechanism of action of botanical drugs, sometimes containing hundreds of potentially bioactive and bioavailable constituents, are mostly unknown and often cannot be elucidated *via* conventional biochemical methodologies.[11] We argue that a network pharmacology approach would naturally lend itself as an unbiased strategy for uncovering the overall mode of action of multi-targeted natural products and their mixtures, as well as of their combinations with approved drugs, such as irinotecan mentioned above. Further, many nutraceuticals or other diet-derived agents relevant to cancer research are multi-targeted and have been shown to synergize with chemotherapeutic drugs and so would likewise benefit

Milla Kibble received her PhD in Pure Mathematics from University College London in 2000. She subsequently worked at Numbercraft Ltd, employing analytics techniques to gain information from large data sets. Since then she has combined teaching of mathematics and statistics with research into probability and graph theory applications to biological data in the Biomathematics Group at Turku University and the Stochastics Group at Aalto University. Milla joined FIMM as a senior researcher in 2014 and her research focuses on network pharmacology approaches to elucidate mechanisms of action and efficacy of drugs and their combinations, with a particular interest in natural products.

Niina Saarinen obtained her MSc in Food Chemistry in 1998 and, in 2002 received her PhD in Cell Biology from the University of Turku, Finland. She performed her post-doctoral studies in Professor Lilian Thompson's group at the University of Toronto, Canada, and in Professor Charlotta Dabrosin's group at the University of Linköping, Sweden. Currently, she is a senior researcher and adjunct professor at the University of Turku. Her research interests focus on hormone responsive breast cancer, steroid hormone production in peripheral tissues, and on the role of dietary polyphenols in breast cancer development.

Tero Aittokallio received his PhD in Applied Mathematics from the University of Turku in 2001. He did his post-doctoral training in the Systems Biology Lab at the Institut Pasteur (2006–2007), where he focused on network biology applications using high-throughput experimental assays. Dr Aittokallio then launched his independent career as a principal investigator in the Turku Biomathematics Research Group in 2007, and received a five-year appointment as an Academy of Finland Research Fellow (2007–2012). In 2011 he started as EMBL Group Leader at FIMM, where his research group is developing network pharmacology approaches to personalized medicine applications.



Jing Tang received his PhD in Statistics from the University of Helsinki in 2009 and became a research scientist in bioinformatics at the VTT Technical Research Centre of Finland. In 2012 he started as a senior researcher at FIMM focusing on computational network medicine approaches to integrate molecular biology and pharmacology data with mathematical modelling and machine learning techniques. In close collaboration with biologists and clinicians, Dr Tang looks forward to the translation of biological and medicinal knowledge to the understanding of disease mechanisms and treatment, and its future developments into innovative healthcare business models.

from a network pharmacology approach.[12] Mapping of the spectrum of potential interactions between the agents and their cellular targets enables one not only to explore the therapeutic potential of natural products but also to better understand their potential adverse effects prior to the actual clinical trials, thereby de-risking and speeding up the drug development process.

In this review, we survey the recent network pharmacology developments and computational tools that have been applied to natural product research, with many studies unsurprisingly concentrating on examples from TCM. The aim is not to present an exhaustive list of examples, but rather to give an overview of the representative concepts and strategies applied so far, with the ultimate goal of generating hypotheses on the efficacy and mechanisms of action of combinations of compounds, where at least one compound is natural, in order to promote good health, prevent disease and act as effective therapeutics with minimal side effects at a personalized level. We also point to references of other successful applications and methods that could be used in the future for natural product research. The current review focuses mainly on *in silico* approaches using *in vivo* and *in vitro* transcriptomic drug response profiles.[13] However, the computational approaches described are also widely applicable to other types of functional phenotypic response signatures. For a broader perspective of 'omic' techniques used in systems biology approaches to TCM research, the reader is referred to the review article of Buriani *et al.*[14] In other studies, Lagunin *et al.*[15] and Barlow *et al.*[16] reviewed *in silico* research into traditional Indian Medicine and TCM, respectively, though concentrating mainly on virtual screening (VS) and cheminformatics techniques, including pharmacophore search, molecular docking, inverse docking and QSAR modelling by means of chemical descriptors and fingerprints. However, most of these techniques can be given a network pharmacology twist by combining them with pathway information as described below in the context of the work by Zhao *et al.*[17] Finally, we give pointers to useful database resources for network-based modelling of natural product responses.

2 Currently known target classes of natural products

Properties of natural products and how these differ from or are similar to properties of synthetic drugs have been characterized in earlier works; see for example Ramallo *et al.*[18] and Clardy *et al.*[19] and the references therein. Of particular interest have been those properties beneficial in drug design. For example, the class of natural products termed secondary metabolites that are involved in defence and signalling are considered as a good starting point for designing new drugs, thus prompting the development of a natural product-likeness score for screening compound libraries and to assist in designing new lead compounds.[20] On the other hand, the more complex natural compounds can exhibit a polypharmacology that can be both a blessing and a curse, as polypharmacology can lead to adverse effects. In general, poor "drug likeness" properties of natural products, such as difficulty to synthesize and make analogues, and also pharmacological challenges, such as bioavailability, have limited the use of natural products in drug discovery.

Despite the known limitations, natural compounds harbour a tremendous potential in terms of chemical diversity and therefore likely bioactivity diversity. Considering that there are around 200 000 known natural compounds, actually only a small fraction of them have been tested for biological activity, partly due to the fact that many of these chemicals do not exist in reasonable amounts as pure substances and are not commercially available for researchers to acquire and test in a biological assay.[21] For example, of the 150 000 structures in the CRC Dictionary of Natural Products (http://dnp.chemnetbase.com/), only about 1% of them have any biological test results in the MDL Drug Data Report database.[2] Similarly, only 1.8% of the 197 201 natural products in the

Krister Wennerberg is a FIMM-EMBL Group Leader. Prior to coming to FIMM, he received his PhD in biochemistry from Uppsala University, Sweden, postdoctoral training at the University of North Carolina, USA, and worked as an assay development group leader at Southern Research Institute in Birmingham, AL, USA. His research group employs chemical systems medicine approaches to gain an understanding of individualized driver signals and vulnerabilities of cancers with a focus on myeloid leukaemia and KRAS-driven adenocarcinomas, and ultimately how these discoveries can be translated into new effective individualized therapies.

Sari Mäkelä received her MD and PhD degrees from the University of Turku, Finland, followed by researcher appointments at the University of Texas Houston Medical School, USA, and Karolinska Institutet, Sweden. In 1997 she took up a lectureship at the University of Turku, where she is currently Professor of Biomedicine. Her research focuses on hormone-dependent cancers and sex steroid action, with a special interest in plant-derived polyphenols as modulators of hormone action and their role in the development and growth of breast and prostate cancer.

Universal Natural Products Database (UNPD) have target binding information in BindingDB[22] or ChEMBL databases[23] (see Table 2 for database references). The critical challenge therefore is how to make use of the chemical diversity to reach the numerous proteins implicated in complex diseases.

We argue that network pharmacology methods are convenient for unbiased investigation of the potential target space of natural products, given the fact that many natural products have multiple targets.[23,24] The currently known target space of natural products is already relatively broad, including *e.g.* enrichment of protein kinases compared to FDA-approved drugs (Fig. 1), opening up unrealized potential for finding new therapeutics acting on as yet untargeted pathways. Taking cancer as an example, where only a small fraction of cancer drivers can currently be targeted by approved cancer drugs,[26] Luo *et al.*[25] predicted strikingly that most of the natural products target at least one out of 104 cancer-associated protein targets, with several natural products having many cancer-associated targets. Systematic approaches are needed to map the yet unexplored target space of natural products, with the aim to extend the druggable cancer genome and to provide the ingredients for combination treatments to overcome the emerging drug resistance.

3 Discovering novel mechanisms of action

Network pharmacology methods have been introduced to suggest hypotheses on mechanisms of action (MoA) of drugs which can then subsequently be tested *in vitro* and *in vivo*. Here we examine some key representative methods that have been applied to single natural products, with potentially multiple targets, or compound combinations, where at least one component is a natural product. In this section, we deal with the case when the object under consideration is a single entity (*i.e.* a known compound or mixture of compounds) and in Section 4 we will consider the case of multiple combinations out of which we would like to find the most effective ones. All of these methods offer the additional opportunities for drug repositioning, *i.e.* to find new, unexpected uses for natural products, which have a well-known safety profile and pharmacokinetic profile.[27] There are many more methods available which have yet to be applied to natural products and some of these are reviewed by Wu *et al.*[27]

Connectivity Map-based phenotypic approaches

One popular class of methods uses the Connectivity Map (CMap) data. The publicly available CMap data resource,[28] produced by the Broad Institute of MIT and Harvard, comprises genome-wide gene expression profiles of 1309 small molecules, produced upon treatment of the molecule to different human cell lines, mainly the prostate cancer cell line PC3 and the breast cancer cell line MCF7 (see Table 2 of database resources). The database is accompanied by the so-called CMap tool, which allows users to input a query gene expression signature representative of a given phenotype and to compare this signature with the reference catalogue of small molecule gene expression profiles. If the query signature is similar to the transcriptional

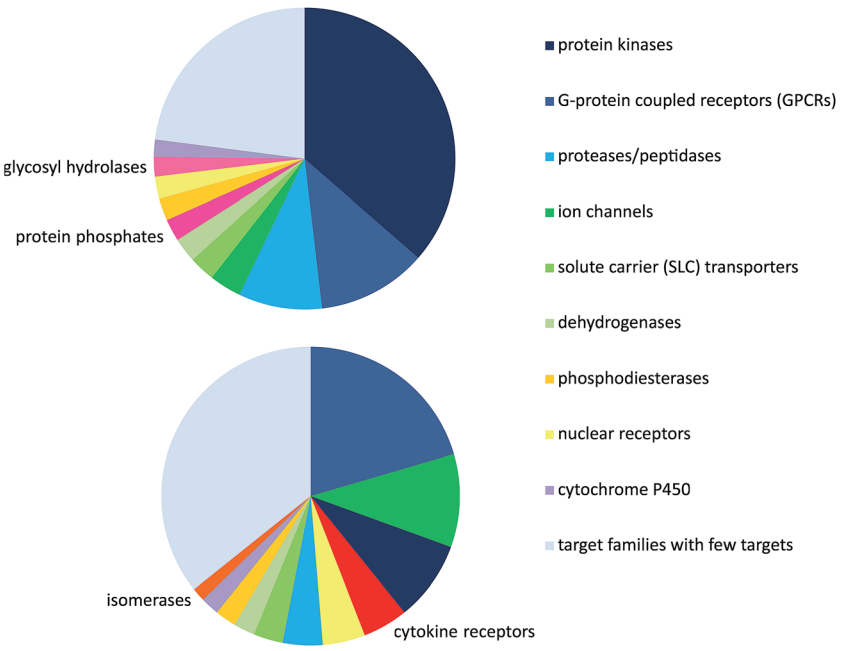


Fig. 1 Top target families of natural products (top) and FDA-approved drugs (bottom). Targets for natural products were obtained from Gu *et al.*[23] who extracted all known interactions of the 197 201 natural products in Universal Natural Products Database (UNPD) for which there is target binding information in BindingDB and/or ChEMBL (1.8% of natural products). The target annotations for the FDA-approved drugs were obtained from the Therapeutic Targets Database (TTD) and the target analysis is of the same form as in Overington *et al.*[100] For both pie charts, target family allocations were adapted manually from those given in the canSAR database.[88] Unique top target families are labeled for natural products (protein phosphatases and glycosyl hydrolases) and for FDA-approved drugs (cytokine receptors and isomerases).

Table 1 Examples of methods used to elucidate MoA of natural products or their combinations[a]

Compound(s)	Disease/phenotype and/or MoA	Method	Input data	Reference	Limitations/comments
Pristimerin	Myeloma, inhibition of proteasome chymotrypsin-like activity	CMap tool[28] to find the mechanism of action of a given compound	CMap data, transcriptional response of cells treated with pristimerin	Tiedemann *et al.*[44]	Potential limitations of the CMap tool are discussed in the original paper of Lamb *et al.*[28] Many of these limitations have been addressed in the more recent CMap-based methods listed in Section 3
Ursolic acid	Skeletal muscle atrophy	CMap tool[28] to find a compound which would act as a therapeutic for a given disease	CMap data, mRNA expression signatures of skeletal muscle atrophy	Kunkel *et al.*[45]	A representative expression signature of the disease of interest is required. A discussion on deriving gene expression signatures for heterogeneous diseases is given by Chen *et al.*[112]
TCM formula Si-Wu-Tang (SWT)	Phytoestrogenic and cancer chemopreventive effect	CMap tool[28] to find the mechanism of action of a mixture of compounds	CMap data, gene expression profiles of SWT at different doses on MCF7 cells	Wen *et al.*[47]	This work demonstrated the potential of CMap methods in deriving MoA for combinations of compounds
15 different TCM formulae	Multiple common properties amongst the formulae, for example anti-cancer, anti-inflammatory, and antioxidative effects	CMap tool to discover properties common to multiple formulae	CMap data and gene expression profiles produced from mouse kidney and liver tissue following oral administration of the different TCM formulae to the mice	Cheng *et al.*[104]	To our knowledge, there are no studies addressing the question of whether MoA predictions are similar across cell line and *in vivo* samples, which would enable the extension of the methods to *in vivo* samples
Celastrol	Acute myelogenous leukemia (AML)	CMap tool and correlation-based metric	CMap and GEO data and gene expression signature resulting from *ex vivo* treatment of human primary acute myelogenous leukemia (AML) with parthenolide	Hassane *et al.*[106]	As above
Curcumin and the TCM formula Si-Wu-Tang	Anti-inflammatory, anti-infective, and neurological regulation for curcumin and anti-neoplastic, antibacterial, vasodilatation and sedative effects of Si-Wu-Tang	Bi-clustering algorithm on the CMap data matrix	CMap data and transcriptional response of cells treated with either curcumin or Si-Wu-Tang	Quan *et al.*[107]	One of the few CMap-based methods applied to natural products which does not use the CMap tool
Natural product library composed of extracts from 92 marine-derived bacterial strains and 20 marine invertebrates	Autophagy, chemotaxis mediated by discoidin domain receptor 2, activation of the kinase AKT	Functional signature ontology (FUSION) to measure the similarity between gene expression signatures produced by chemical and genetic perturbations	Gene expression signatures produced by addition of the natural product fraction to the HCT116 cell line, gene expression signatures produced by siRNAs and miRNA mimics	Potts *et al.*[49]	Potential off-target effects of siRNAs[113] may influence this method. However, here the results were experimentally validated, proving the usefulness of this approach
TCM formula PHY906	Protection against chemotherapy-induced intestinal toxicity and enhancement of anti-tumour activity produced by chemotherapy	Statistical analysis of differences in gene expression between different treatments and analysis of the key pathways involved using IPA	Transcriptional changes produced upon administration of compounds both individually and in combination in a preclinical mouse model of colon cancer	Wang *et al.*[50]	Here the transcriptional profiles produced from the *in vivo* experiment were not compared to the CMap cell line profiles

Table 1 (*Contd.*)

Compound(s)	Disease/phenotype and/or MoA	Method	Input data	Reference	Limitations/comments
Astragaloside IV	Cardiovascular disease (CVD)	Inverse docking of compound to targets in CVD associated pathways. Network analysis *via* a simulated annealing algorithm	Targets of FDA-approved drugs for CVD and their pathways, protein–protein interaction network of the human	Zhao *et al.*[17]	This method relies on the accuracy and completeness of the disease pathway, constructed using targets of FDA-approved CVD drugs, and thus might miss the new pre-clinical CVD related targets
Bu-Shen-Huo-Xue formula (BSHX)	Chronic kidney diseases (CKD)	Docking of compounds to CKD associated targets. Analysis of compound–target network characteristics	Known targets of the compounds in BSHX and drug targets related to CKD	Shi *et al.*[48]	Requires knowledge of the hundreds of components of the mixture formula and targets related to the disease of interest

[a] We have listed some potential limitations of these methods. However, these concerns are not meant to undermine or question the specific methods and results presented, rather to open up discussion on how to further improve these methodologies and the reliability and translationability of the results in natural product research.

response profile of a particular small molecule, then prior knowledge about the small molecule could infer mechanisms of the phenotype. For example, one may discover the unknown MoA of a compound *via* the comparison of its gene expression profile to those of small molecules of known MoA (see Fig. 2 and the accompanying illustrative example in the text). Further, if the phenotype of interest is a disease, then the inverse similarity of its query signature with the gene expression profile of a particular small molecule implies that the molecule could act as a therapeutic for the disease.

An overall review of the applications of the CMap resource for drug repositioning, lead discovery, MoA elucidation and systems biology is given by Qu *et al.*[29] It should be noted that even just for MoA discovery there is an abundance of variations on the original method in the CMap tool, which is based on Gene Set Enrichment Analysis (GSEA)[30] as described in the landmark CMap paper of Lamb *et al.*,[28] and several new methods have been published subsequent to the aforementioned review. Notably, the method of Jahchan *et al.*[31] incorporates information on both the known targets and enriched pathways of the top-scoring small molecules; Laenen *et al.*[32] incorporate a functional protein association network into their model and Wu *et al.*[33] incorporate a side effect score based on differential expression of essential genes. Further, some novel probabilistic methodologies are beginning to emerge[34–36] and a new combined unsupervised and supervised approach[37] has also been shown to perform well on CMap data. Below, we detail a representative set of CMap-related methods applied to natural products that are in order of most traditional methods to newer variants (see Table 1 for a summary).

As an illustrative example (see Fig. 2), we consider the major isoflavone present in soybeans, genistein, which is known to inhibit human cancer cell growth.[38] We selected genistein as it is a good example of a natural compound with multiple targets and with a complicated dose–response profile highly dependent on the test system. Especially in the case of oestrogen-responsive breast cancer, the preclinical settings have poorly predicted the outcomes in women. This suggests that genistein has targets that are poorly modelled in currently available preclinical models of breast cancer. We extracted the transcriptional response profiles for genistein in MCF7 and PC3 cell lines from the CMap database, and a consensus transcriptional response across the different experimental settings was compared to the transcriptional profiles of the other small molecules in CMap using the method of Iorio *et al.*[39] *via* the freely available online tool MANTRA.[40] This network-level method produces a distance measure to quantify the similarity in transcriptional response between small molecules, thus mapping a drug–drug network. The known targets of the small molecules closest to genistein in the network were extracted from ChEMBL and DrugBank using the KIBA method[41] (see Table 2 for database resources). Among these targets were PTGS2 (COX-2) and several HDACs. Importantly, we validated the PTGS2 target experimentally *via* measurement of downstream PGE2 on both PC3 and MDA-MB-231 cells (Fig. 2). Subsequently, we found literature evidence that genistein is an HDAC inhibitor,[42] as well as further literature evidence to support PTGS2 as a genistein target.[43] This

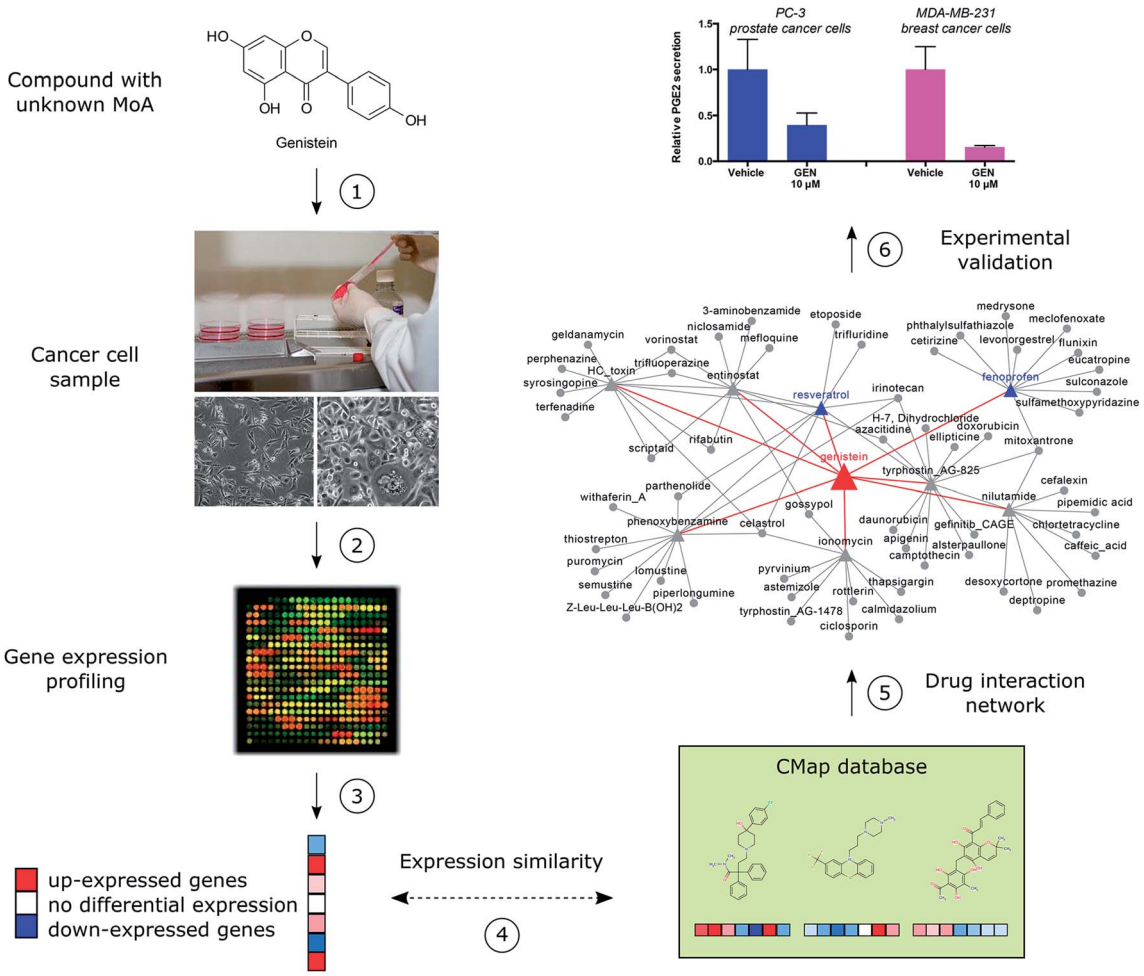


Fig. 2 Pipeline for Connectivity Map methods for discovering unknown mechanisms of action. (1) We start with a compound (here, genistein), or mixture of compounds, with unknown MoA. The compound is added at a given dose (typically 10 μM) to a particular cell line (*e.g.* MCF7 or PC3). (2) The genome-wide gene expression response after the perturbation is measured. (3) Genes are ranked according to their differential expression upon treatment (*i.e. versus* the vehicle control) to produce a transcriptional profile for the compound (and given dose and cell line). Some methods then aggregate all profiles for a compound to give a consensus gene expression signature for the compound. (4) This signature is then compared to signatures of the 1309 drugs in the CMap database *via* a predefined measure of similarity (see Fig. 3 for more details of the CMap data). (5) The compounds can then be mapped into a drug network using this measure of similarity of transcriptional response as distance, *i.e.* drugs close together in the network are similar in terms of their transcriptional response. Only edges below a certain threshold are shown here. (6) We take the drugs which are closest to our compound of interest (triangle shape nodes, connected through red edges) and find out their established MoA from databases such as ChEMBL (so-called 'guilt-by-association'). For example, a particular target may be enriched in the targets of neighbouring drugs (here PTGS2 is a target of fenoprofen and resveratrol, both coloured blue) or perhaps a certain target pathway is over-represented. From this information, we can gain an understanding of possible mechanisms for our compound which can be put forward as hypotheses for experimental validation (here, PTGS2 was validated as a target for genistein *via* the measurement of downstream PGE2). For the CMap-based approaches, most studies use a similar pipeline, for example, with genistein replaced by pristimerin or SWT. In the work by Kunkel *et al.*,[45] in effect one replaces the genistein profile with the disease profile of muscle atrophy and considers *inverse* similarity to the CMap compound profiles in step (4); the compounds having the closest inverse similarity are then put forward as possible therapeutics for the disease.

example demonstrates the power of unbiased network-level analyses in predicting the potential cellular targets of natural products.

In a similar manner, but using the original CMap tool rather than the method of Iorio *et al.* to measure similarity in transcriptional responses, Tiedemann *et al.*[44] elucidated the mechanism of action of the natural product pristimerin, which was of interest having been identified from Spectrum library screening as the top-ranked suppressor of cyclin D2 promoter transactivation. It was found that the transcriptional response of cells treated with pristimerin closely resembles cellular responses elicited by proteasome inhibitors, with rapid induction of heat shock proteins, activating transcription factor 3 (ATF3) and C/EBP homologous protein (CHOP). These inferred mechanisms for pristimerin were validated *in vivo* and *in vitro*. In an earlier work, Hieronymus *et al.*[101] used a similar methodology to generate the hypothesis that celastrol and gedunin function as HSP90 inhibitors, and indeed the two compounds were shown *in vitro* to inhibit HSP90 activity and HSP90 clients, including the androgen receptor (AR).

Table 2 Examples of database resources and their network pharmacology applications[a]

Database	URL/availability/developers	Description	Example applications
CMap[28]	http://www.broadinstitute.org/cmap/, publicly available, produced by the Broad Institute of MIT and Harvard	A database of genome-wide gene expression profiles produced upon treatment of 1309 compounds on cancer cell lines. No natural product labelling of compounds but includes at least 301 natural products (identified *via* the CRC dictionary of natural products)	Elucidation of drug mechanisms of action and drug repurposing[44,45,47,84]
GEO[115]	http://ncbi.nlm.nih.gov/geo/, publically available, run by the National Center for Biotechnology Information	A public functional genomics data repository. Tool to mine partial and genome-wide signatures of drug responses and disease states	Prediction of lead compounds for a given disease[40,106]
ChEMBL[85]	http://www.ebi.ac.uk/chembl, publicly available, Run by the European Bioinformatics Institute in Cambridge, UK	Database with bioactivity measurements for almost 1.5 million distinct compounds and over 10 000 protein targets. No natural product label (see comment in legend)	Prediction of drug–target interactions,[86] prediction of structure–activity relationships[87]
canSAR[88]	http://cansar.icr.ac.uk/, publicly available, developed at the Cancer Research UK Cancer Therapeutics Unit	Biological, pharmacological, chemical, structural biology, molecular activity and protein network data for over 1 million compounds. No natural product label	Identifying novel druggable cancer targets[26]
Therapeutic Target Database (TTD)[89]	http://bidd.nus.edu.sg/group/TTD/ttd.asp, publicly available, developed by the bioinformatics & drug design group at the National University of Singapore	Database with information on 2360 known and explored therapeutic targets, the targeted disease, pathway information and the corresponding 20 667 drugs. The database also contains biomarkers, drug scaffolds, 1008 nature-derived agents, 20 818 multi-target agents against 385 target-pairs, and the activity data of 1436 agents against 297 cell-lines. List of FDA-approved, clinical trial and preclinical nature-derived drugs with specific annotation as to which are natural products	Identification of patterns in species origin of nature-derived drugs;[90] identification of novel cancer drug targets[91]
Universal Natural Products Database (UNPD)	http://pkuxxj.pku.edu.cn/UNPD, publicly available, Peking University, China	Database containing 229 358 natural products	To analyse molecular descriptors, distribution in chemical space and biological activities of natural products and compare them to FDA-approved drugs[23]
Traditional Chinese Medicine (TCM) database@Taiwan[92]	http://tcm.cmu.edu.tw/, publicly available, developed at China Medical University in Taiwan	Chemical information including 2D and 3D molecular structures for 37 170 natural compounds from 352 TCM ingredients. No target information	Virtual screening for compounds of specific activity[93]
Chem-TCM[94]	http://chemtcm.com/, not freely available. However earlier version of the database (CHCD) may be available from the authors upon request, developed at King's College London in collaboration with the Shanghai Institute of Materia Medica	Chemical and botanical information on 12 070 compounds found in approximately 350 Chinese herbs. Includes predicted activity against 41 common Western therapeutic targets[24] and estimated molecular activity according to traditional Chinese herbal medicine categories[95]	Virtual screening for compounds of specific activity[96,97]

Table 2 (*Contd.*)

Database	URL/availability/developers	Description	Example applications
Traditional Chinese Medicine Systems Pharmacology database and analysis platform (TCMSP)[98]	http://sm.nwsuaf.edu.cn/lsp/tcmsp.php, publicly available, Center for Bioinformatics, Northwest A&F University, Yangling, Shaanxi, China	Information on more than 30 000 compounds from more than 500 medicinal herbs, including twelve ADME-related properties, known and predicted drug targets and diseases. Compound–target and target–disease networks and tools for network visualization and analysis	Drug discovery, mechanisms of action and disease associations[68]
Dictionary of Natural Products (DNP)	http://dnp.chemnetbase.com/, not freely available, CRC press	A major source of chemical, biological, pharmacological and toxicological data on natural products	Historical analysis of structural novelty of natural products[99]

[a] Representative examples of natural product-specific databases and of more comprehensive repositories which include natural products. Sanderson[83] highlighted arguments for and against natural product-specific databases. However, with non-specific databases, it is often difficult to determine which compounds are natural and which are not. For example, in ChEMBL, one can go to the 'Browse Drugs' tab on the home page and search the compounds with criteria 'Natural Product'. However, compounds produced from this search are those that have been classified as 'natural-product derived' (as opposed to 'synthetic small molecule') using a 'natural product'-substructure filter. We believe that the addition of a natural product field to non-natural-product-specific databases would further enhance natural product research. There are many additional TCM-specific databases, many of which are catalogued in Barlow *et al.*[16] A comparison of some of these databases, including some potential limitations, is also given on the TCMSP website http://sm.nwsuaf.edu.cn/lsp/tcmsp.php.

Kunkel *et al.*[45] produced two unbiased mRNA expression signatures of skeletal muscle atrophy, a common and debilitating condition that lacks a pharmacologic therapy, and used these signatures to query the CMap tool for compounds with a signature inversely similar to this disease to identify potential small molecule inhibitors of muscle atrophy. Among the top hits was ursolic acid, a natural compound enriched in apples. Extensive experimental evidence was given there and in a subsequent paper[46] to support the model that ursolic acid reduces obesity, glucose intolerance and fatty liver disease by increasing skeletal muscle and brown fat, thus potentially not only providing a therapy for skeletal muscle atrophy but also a disease preventative solution. Mukherjee *et al.*[102] had earlier in a similar manner used HuR mRNA dynamics as a quantitative response phenotype to be queried against CMap compounds and discovered resveratrol to be a candidate effector of HuR and T-cell activation, these effects being confirmed experimentally *in vitro*.

Si-Wu-Tang (SWT) is a TCM formula which comprises four herbs and at least nine bioactive phytochemicals that has been widely used over a thousand years for women's health, in particular for oestrogen-related diseases. Wen *et al.*[47] produced gene expression profiles of SWT at different doses on MCF7 cells and used canonical pathway analysis of IPA (Ingenuity Pathway Analysis, http://www.ingenuity.com) to show that differentially expressed genes from the high concentration SWT treated group are most significantly enriched in several cancer signalling pathways, in particular the Nrf2-mediated oxidative stress response pathway, which plays an important role in cancer prevention. When the same analysis was performed selectively for dose-responsive genes (as these are most likely to reflect the true pharmacological effect of SWT), the Nrf2-pathway was again top of the list prompting the authors to suggest that SWT is cancer preventative. The gene expression profiles of SWT were then compared to the 1309 compounds using the CMap tool, with estradiol and some chemopreventive compounds amongst the top hits, suggesting a potential oestrogen-like and chemopreventive effect. The oestrogenic activity of SWT was confirmed in a follow-up study[103] using a cell proliferation assay on MCF-7 (ER-positive) and MDA-MB-231 (ER-negative) cells and an oestrogen-responsive element (ERE) luciferase reporter assay in MCF-7 cells.

There are also a few examples where the transcriptional response profile to be compared using the CMap tool is produced using *ex vivo* or *in vivo* experiments, rather than in cell lines like in the traditional CMap approach. For instance, Cheng *et al.*[104] orally administered 15 different TCM formulae to mice and produced gene expression profiles using mouse kidney and liver tissue samples. The molecular signatures of the formulae were then compared with those of molecules in the CMap database to highlight potential common effects of these formulae. Interestingly, the molecular signatures were also compared with the expression signatures of 223 chemical treatments obtained from the Environment, Drugs and Gene Expression (EDGE) database[105] to check for potential nephrotoxicity.

Hassane *et al.*[106] generated a gene expression signature in response to treatment of human acute myelogenous leukemia (AML) cells in primary culture with parthenolide (PTL), which has been shown to ablate AML stem cells as a single agent. With a view to discovering targeted therapies for AML stem cells, the authors searched for similar signatures in CMap using the CMap tool as well as in publicly available gene expression profiles deposited into the Gene Expression Omnibus (GEO, Table 2), using a correlation-based metric. Using both methods, the natural product celastrol was found as a top hit and was shown experimentally to effectively eradicate AML at the bulk, progenitor, and stem cell level. Although, to our knowledge, there are no studies addressing the question of whether MoA predictions are similar across cell line and *in vivo* samples, these studies demonstrate the use of multiple tissue types, including those from patient samples, in CMap-based methods. The work of Hassane *et al.* further demonstrates the applicability of these methods across expression profiles produced at multiple centres, as is the case with the GEO data.

Most of the CMap-based methods employed for natural products use the CMap tool to measure the similarity of transcriptional responses. However, in addition to the "V-score" correlation metric used by Hassane *et al.*,[106] Quan *et al.*[107] also employed an alternative approach to similarity assessment. They use the idea of drug-induced transcriptional modules[108,109] for elucidating the MoA of curcumin and the TCM formula Si-Wu-Tang, mentioned above. In particular, they used a bi-clustering algorithm to identify so-called gene modules in the CMap data matrix, each module consisting of a group of compounds with transcriptional changes on a common set of genes. Each compound was then assigned a binary vector indicating its presence or absence in each of the modules, a high similarity between the binary vectors of two compounds representing similarity in biological effects for those two compounds. The predicted biological effects of curcumin and Si-Wu-Tang were supported by literature evidence.

Other functional approaches to MoA elucidation

There are also many other functional phenotype-based approaches that do not use the CMap resource. Some of the approaches reviewed below do not go as far as suggesting a direct MoA, although it may for example be unrealistic to expect to be able to de-convolute the mechanism of action of a mix of many compounds, as is the case in Shi *et al.*,[48] where individual components themselves are polypharmacologic.

Potts *et al.*[49] developed a method called functional signature ontology (FUSION) to identify MoA for marine-derived natural products, their natural product library being composed of extracts from 92 marine-derived bacterial strains and 20 marine invertebrates. After adding a natural product fraction, which typically consisted of mixtures of two to six compounds, to the human colon cancer cell line HCT116, the authors measured the gene expression changes on six reporter genes chosen to serve as a proxy for the physiological state of the cell. These gene expression signatures were then compared to those produced by 780 small interfering RNAs (siRNAs) targeting human kinases and related proteins and 344 synthetic microRNAs (miRNAs) to identify biologically similar genetic and chemical perturbations. Euclidean distance was mainly used to quantify the similarity between expression profiles and hierarchical and neighbour joining clustering of Euclidean distance was used to draw neighbour joining trees. This analysis produced FUSION maps linking bioactive molecules to the proteins and biological processes they engage in cells. For example, focusing on autophagy, the authors found that metabolites isolated from the *Streptomyces bacillaris* strain SN-B-019 induced reporter gene signatures that positively correlated with knockdown of Unc-51-like autophagy-activating kinase1 (ULK1), and experiments confirmed that four of the five compounds isolated from the SN-B-019 fraction inhibited autophagy. Predicted mechanistic relationships for compounds with functional roles in chemotaxis mediated by discoidin domain receptor 2 (DDR2) or activation of the kinase AKT were also confirmed experimentally.

PHY906 is a four-herb TCM formula with reported anti-inflammatory properties that has been used as an adjuvant to relieve the side effects associated with chemotherapy. Wang *et al.*[50] argued that, when attempting to elucidate the mechanism of action of PHY906, the overall effects of the herb should be investigated *in vivo* using global transcriptional profiling. Thus, the authors examined alterations of the transcriptional program induced by PHY906 following its administration as a single agent and in combination with irinotecan in a preclinical mouse model of colon cancer. The effects were compared to those observed in normal mouse tissues (liver and spleen). When administered alone, PHY906 had no effect on tumour growth and, based on transcriptomic changes in tumour tissues, immune-related canonical pathways were suppressed. However, PHY906 significantly enhanced the anti-tumour activity of irinotecan and this combined treatment in tumour tissue activated pro-apoptotic and pro-inflammatory pathways, an effect not observed in normal tissues. These results led the authors to speculate that PHY906 enhances the anti-tumour properties of chemotherapy with irinotecan by imparting a pro-inflammatory state that is not observed in irinotecan naïve cancerous tissue, suggesting a potentially useful dichotomy in behaviour for PHY906.

Zhao *et al.*[17] considered astragaloside IV (AGS-IV), the main ingredient of a herb widely prescribed in TCM for the treatment of cardiovascular disorders. There is extensive experimental evidence of the cardiovascular-protective effects of AGS-IV, however the MoA of this compound is unknown. The authors produced 33 key pathways involved in cardiovascular disease (CVD) by investigating which pathways are enriched with the targets of FDA-approved small molecule CVD drugs using pathway enrichment analysis. A flexible ligand–protein inverse docking program, INVDOCK, was then applied to the proteins in the 33 enriched pathways to find 39 distinct proteins as putative targets for AGS-IV, three of which were experimentally validated. The authors then proceeded to map the putative targets onto a protein–protein interaction network of the human genome to find that most of the targets link to a single sub-network either through direct interactions or through only

one intermediate protein. A simulated annealing algorithm partitioned this into six topologically compact modules, where the targets are arranged around common hub proteins, agreeing with previous studies suggesting that weak inhibition of multiple targets may be more efficient than potent inhibition of a single hub target.[51] *Via* examination of CVD disease networks it was also observed that some of the putative targets are at crosstalk sites of multiple pathways, thus potentially allowing AGS-IV to intervene in multiple pathways in CVD *via* a limited number of proteins. The authors suggest that AGS-IV produces its therapeutic effect *via* a combination of multiple mechanisms, including anti-oxidation, anti-inflammation, blocking calcium influx and immune regulation, anti-thrombosis and vasodilation.

Shi *et al.*[48] used a similar type of network approach but for a TCM formula Bu-Shen-Huo-Xue (BSHX), which is composed of five herbs, together including 774 known compounds. The formula BSHX is frequently used for treating chronic kidney diseases (CKD) but its MoA is unknown. First, the known targets of the compounds in BSHX and also drug targets related to CKD were obtained from various databases, including the Therapeutic Target Database (see Table 2). Then, the natural product compounds were docked to the CDK target proteins using AutoDock Vina and a natural product–target network was produced. Through analysis of network characteristics, such as node degree, potential effective ingredients and their synergistic mechanism were putatively identified. In particular, it was proposed that BSHX exerts its therapeutic effect by using multi-channel network regulation, such as regulating the coagulation and fibrinolytic balance, and the expression of inflammatory factors, inhibiting abnormal extracellular matrix (ECM) accumulation. Tanshinone IIA, rhein, curcumin, calycosin and quercetin were proposed to be potential effective ingredients of BSHX and their therapeutic effect was validated experimentally, with the key finding that the therapeutic effect of the combined administration of the five components was significantly higher than that of each compound alone, adding support to the proposed multi-component and synergistic mechanisms of TCM. Indeed, one of the major challenges in the modernization of TCM is identifying the active ingredients, which either individually or in combination produce the therapeutic effects or adverse effects and although we do not focus on this particular issue here, we comment that the DMIM[52] methodology introduced below also contributes to this active area of research.

4 Predicting effective therapeutic combinations

There is a strong interest in the use of combinations of drugs to increase therapeutic efficacy and reduce drug toxicity caused by high doses of individual drugs, the rationale being that a combination of drugs may target nodes on compensatory pathways countering problems such as emerging drug resistance. Interestingly, there is evidence to show that the synergistic interaction of different bioactive components of botanical drugs can contribute to the therapeutic effect and is capable of reducing side effects,[53–55] supporting the idea that new botanical drugs should be designed as a combination of several active components. Further, there is experimental support for natural products working with Western drugs to enhance the therapeutic effect and/or limit adverse side effects, with much research focusing on anticancer treatments.[10,56,57] Critically, there are also numerous cases where polypharmacologic natural compounds, extracts and nutraceuticals interfere with the function of approved drugs,[58] leading to adverse drug interactions. Since there is an increasing trend in Western countries towards an integration of the traditional Chinese and Western systems of medicine, more systematic research into such combinations is timely to ensure their safe and effective usage.[59]

In this section, we look at a few selected methods of network pharmacology for prioritization of the most therapeutically effective combinations of compounds for experimental validation *in vitro* or *in vivo*. Such computational methods are important due to the impracticality of testing all the possible drug combinations, even with the availability of robotic drug screening infrastructure, as the number of combinations increases exponentially with the number of drugs to be screened. Again, we illustrate methods already applied to natural products, but stress that there are many generically applicable methods available to the natural products research community that have yet to be applied to this area. These methods are reviewed by Wu *et al.*,[27] Tang *et al.*[60] and Sun *et al.*,[61] where the latter focuses on computational methods for high-throughput biological measurements. It should be noted that molecular response profiles, such as genome-wide gene expression data, for combinations of drugs is currently still scarce, though some efforts toward prediction of such response profiles have been made.[27,62] An attempt has also been made to incorporate side effect prediction and efficacy prediction into a single model to predict the overall effect of a combination.[62]

Li *et al.*[63] introduced a novel method called NIMS (Network target-based Identification of Multicomponent Synergy) to assess the synergistic strength of multicomponent therapeutics. NIMS combines topological features of the targets of the individual compounds in a disease-specific network with a measure of the similarity between the phenotypes associated with the targets of the individual compounds. The underlying hypotheses are that (1) the more important a compound target gene is as a network node, the stronger the effect on the disease that the compound will produce, (2) if a pair of compounds produces synergy then their target genes should be adjacent in the disease network, and (3) compounds with independent action mechanisms but treating similar diseases may be more likely to produce synergistic effect. The method was applied to prioritize synergistic combinations among 63 agents including 61 herbs or herb compounds. Five known synergistic pairs all ranked highly based on predicted synergy scores, with two pairs in the top three. Initially an angiogenesis network was used. However, the method was shown to be relatively robust to the choice of background network. As experimental validation, five compounds were chosen from the 63 agents and combined individually with the anti-angiogenic alkaloid sinomenine to

investigate synergistic effect *in vivo*, the experiments producing an identical rank ordering of synergistic effect to that predicted. In addition, two predicted synergistic agent pairs, sinomenine and matrine, and sinomenine and honokiol, respectively, are the main constituents of TCM herbal formulae such as Qing-Luo-Yin and Tou-Gu-Zhen-Feng. In a previous article,[52] the authors developed another method called the Distance-based Mutual Information Model (DMIM), which uses information from the composition of current TCM formulae to try to elucidate the combination rule in TCM herbal formulae and to suggest TCM compound pairs that may work synergistically to produce anti-angiogenic effects. As with the methods in Section 3, here too both the NIMS and DMIM methods are reliant on the accuracy and completeness of the background information; the target information of the individual components in the former method and knowledge of active components of herbs in the latter.

Zhao *et al.*[64] applied a computational optimization algorithm based on integer programming that uses as its input single-drug efficacies for genetically variant cell subpopulations to predict how drug combinations will affect heterogeneous tumours. The main assumption in the model, which is nicely illustrated in Fig. 1 of the original publication, is that commonly used combinations of chemotherapeutics act as linear averages of each component drug against homogeneous tumours, an assumption that was demonstrated in their earlier work.[65] Tumour heterogeneity is modelled using engineered RNAi based knockdowns. The widely used anti-cancer agent vincristine (a natural product) in combination with vorinostat was predicted as the optimal treatment for a three-component population consisting of the parental Eμ-Myc; p19 $Arf^{-/-}$ lymphoma (no shRNA) and subpopulations expressing either a Chk2 (a DNA damage checkpoint regulator) or a Bok (a Bcl2-family cell death mediator) shRNA, with validation being performed both *in vivo* and *in vitro*. This particular example led to the observation that considering a heterogeneous tumour in its entirety can result in non-intuitive optimal drug combinations containing drugs that are not the best single agent for any of the subpopulations.

Hassane *et al.*[110] used the CMap tool to search for compounds that enhance the therapeutic effect of the plant-derived compound parthenolide (PTL), mentioned in Section 3, which is a suboptimal anti-AML pharmaceutical. The authors hypothesized that compounds with an inversely similar transcriptional profile to that of PTL may counteract some of the cellular cytoprotective responses of PTL. Inhibitors of PI3K and mTOR were significantly enriched in the top hits from CMap and accordingly the authors confirmed that, compared with single agent treatment, exposure of AML cells to the combination of PTL and PI3K/mTOR inhibitors significantly decreased viability of AML cells *in vitro* and reduced tumour burden in murine xenotransplantation models.

We conclude this section by commenting that the rich body of TCM experience in combined use of herbs may provide an excellent model for studying synergistic effects among different components and thus be utilized in the evaluation of different network pharmacology computational methods.[63] It is also critical to validate experimentally the most potent computational predictions, either *in vitro* or ideally *in vivo*, in order to avoid reporting over-optimistic combination results or practically unrealistic model predictions.

5 Conclusions and future directions

Computational tools of network pharmacology are increasingly being developed and applied to drug discovery with a view to combat problems such as lack of efficacy of single-target molecules, side-effects, drug resistance and individual variation in treatment response. At the same time, natural products and their combinations are commonly multi-targeted with targets thought to encompass and exceed the currently limited space of targets of FDA-approved drugs, thus holding potential for new types of therapeutic opportunities.[66] There are many empirical examples of successful therapeutic uses of natural products, in particular *via* TCM formulae, however the lack of knowledge of their MoA diminish their scientific validity and thus limit adoption in Western medicine. We believe that unbiased network pharmacology approaches for uncovering MoA would lend themselves well to natural product research. In practice, recent developments have been relatively slow to be translated into the field. In this review, we have given an overview of the successful approaches applied so far to natural products, giving also references to approaches applied in a different context. The hope is that those involved in natural product research will see the usefulness of these computational tools and adopt them in their own future work.

There are several critical issues to be addressed with regards to applying network pharmacology methods to natural product research. Firstly, the prediction of potential side effects of drugs is an important part of the drug development process. It is known that most drugs come with some side effects and, next to lack of efficacy, they are the leading cause of attrition in clinical trials of new drugs.[4] Any future predictions of efficacy of natural products or combinations thereof should include the prediction and subsequent investigation of potential side effects. Recent efforts, a number of which are highlighted by Wu *et al.*,[27] have been made to develop tools to predict side effects of drugs. Resources, such as the recent work of Kuhn *et al.*[67] on the prediction of proteins that elicit side effects are equally applicable to the natural product research community. To our knowledge, there are, however, no examples of computational side effects prediction being applied to natural products so far, apart from the work of Cheng *et al.*[104] mentioned in Section 3. Similarly, the evaluation of ADME (absorption, distribution, metabolism and excretion) properties is important in drug development. However, *in silico* ADME studies have also been slow to make the transition from drug development to natural product research.[68]

Secondly, foods contain bioactive compounds that can be classified as non-essential nutrients, with evidence of beneficial effects on human health.[69–75] In many cases, the therapeutic effect of individual dietary compounds may be limited due to low bioavailability or low content in foods.[48] However, when combinations of compounds with similar or complementary

effects are considered, a cumulative effect may occur. As an example, when the gene expression profiles of the dietary polyphenols apigenin, luteolin and chrysin were compared to the expression profiles of all 1309 small molecules in the CMap data resource, it was found that the transcriptional responses of these three natural compounds were more similar to each other than to any other compound in the database. Thus, in the diet, there may be multiple compounds producing a similar effect, which may be adding up to a therapeutically effective dose. Also food components having multiple weak targets[76] may have an important role to play in disease prevention and there is scope for the methods described here to be used in discovering which dietary compounds, alone or in combination, play a part in which preventive/therapeutic mechanisms.

Thirdly, for more comprehensive understanding of the compounds' MoA, methodologies incorporating other types of functional response profiles, including, for instance, proteomics, metabolomics and DNA methylation, as well as dynamic or longitudinal data[77] would broaden the limited view captured by the single time point transcriptomic responses alone. Such multi-phenotype data resources are emerging, notably from community efforts such as the new Library of Integrated Network-based Cellular Signatures project (LINCS, http://www.lincsproject.org/), which is aiming to provide a unique comprehensive reference data resource of cellular response signatures to a wide spectrum of small molecule and genetic perturbations. LINCS includes, for example, biochemical protein binding profiles and various cellular phenotypic response profiles in addition to the genome-wide transcriptional signatures. Such a data resource should facilitate the development of more realistic, systems-level models of disease mechanisms and drug action. Some initial case studies which have utilized the integrated diverse data sets from LINCS are presented in Vidović *et al.*[78]

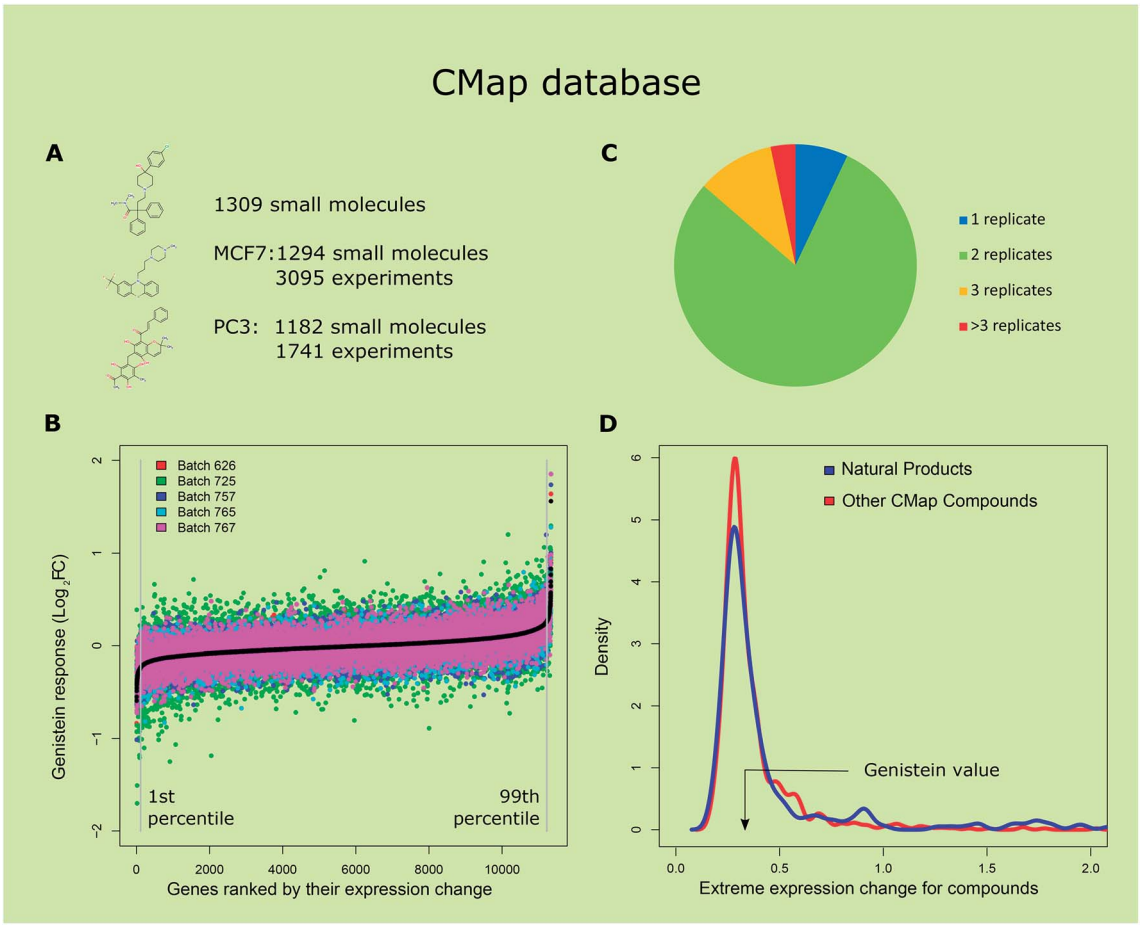


Fig. 3 The Connectivity Map Database. (A) Experiments present in the current Connectivity Map database, which were performed in numerous batches over the period of one year (here information for the two main cell lines is given). (B) An example of the gene expression response of a CMap small molecule, here the Fold Change (FC) produced by genistein 10 μM on MCF7 measured in 5 batches using the most abundant platform (HT_HG_U133A). The black curve shows the average value over the 5 measurements and the grey vertical lines the portion of the top and bottom 1% of the changes (1st and 99th percentiles). (C) Pie chart of the number of replicates for small molecules on the MCF7 cell line and HT_HG_U133A platform in the CMap data. For PC3, the majority of the compounds have just one replicate on this platform (72% of the compounds). (D) The distribution of the average of the extreme absolute $\log_2$ FC changes within the 98th percentile for all CMap compounds tested on MCF7, demonstrating that both natural product and non-natural product compound profiles have comparable expression changes. The 98th percentile was chosen here since, as expected, for a given compound most genes show very little expression change.

Finally, it is important to elucidate the polypharmacological behaviour of natural products that does not arise from direct binding to specific target proteins, in order to ascertain when such indirect effects are beneficial and when detrimental. It has recently been suggested that phytochemicals derive much of their broad pharmacological effects *via* cell membrane perturbations, rather than through binding to specific proteins.[79] Further, several epigenetic responses to treatment with natural products have been reported,[80] but the mechanisms by which epigenetic changes regulate the pathogenesis and progression of diseases such as cancer is still largely unknown. Dashwood *et al.*[81] put forward the idea that dietary HDAC inhibitors, as weak ligands, might subtly regulate the expression of genes involved in cell growth and apoptosis. Ehrman *et al.*[24] showed that most plant compounds tend to have moderate to weak affinities for their targets and suggest that they may therefore serve more as modulators of target function rather than strong inhibitors. Csermely *et al.*[51] discussed the possible advantages of low-affinity multi-target drugs already in 2005. It remains an open question whether these types of weak but important polypharmacological effects can be captured by the functional response data, or whether we need a new type of data to elucidate which compounds behave in this promiscuous way. More importantly, can these effects and their combinations be converted to proven clinical effects to treat diseases?

Fig. 3 provides an overview of the CMap data and its potential limitations. One possible limitation, highlighted already in the original CMap paper,[28] is the presence of potential batch effect, *i.e.* the similarity of gene expression profiles observed for unrelated stimuli in cells grown or processed at the same time (Fig. 3A and B). Attempts to remedy batch effects have been made, for example, in the methods of Iorio *et al.*[39] and Iskar *et al.*[114] Another potential issue with the CMap data is that for most compounds there is only one replicate per cell line on the main microarray platform used in CMap (Fig. 3C). This poses some challenges to the statistical analysis, such as finding differentially expressed genes for these compounds. Knowing which genes are significantly differentially expressed for a given compound would help, for example, in choosing the appropriate, significance-based cut-off for the number of genes taken as up- or down-regulated; these genes being then used in the signature comparison as part of some CMap methods, such as the method of Iorio *et al.*[39]

Of the many network pharmacology computational tools already available, it is not at all clear which methods or options are the most suited to natural product research. Even for the CMap-based phenotypic approaches discussed in Section 3, there are a multitude of variations on each stage of the approaches, ranging from data pre-processing[28,34] to expression signature comparison.[34,39] However, it seems that the response profiles from natural product and non-natural product compounds have relatively comparable expression changes (Fig. 3D). This knowledge makes it possible to reliably compare natural products and other compounds with computational approaches that use the FC-measures, like those based on the novel probabilistic method of Khan *et al.*[34] Similarly, improvements on the original LINCS data analysis pipeline have already been proposed.[111] It would be therefore important to evaluate how much these different variations affect the results, and whether there would be an optimal combination of methods and options specifically for natural product research.

What is clear though is that although the current methods already show great promise, they are still constantly developing. Similarly, the background information which many of the methods discussed here depend upon, such as target information, protein interaction networks and molecular pathways, are constantly being updated and becoming more comprehensive and accurate. Drug discovery is moving towards a precision medicine treatment approach and it is likely that we will see many more personalized approaches to disease prevention in the future. A number of critical challenges remain, however, including addressing the impact on drug response of both target and non-target associated genetic and/or epigenetic alterations, along with understanding the effect of candidate lead compounds in specific disease settings and appropriate cellular, tissue and organism environments that are therapeutically important.[82] As new network pharmacology methods become available and we find new ways to integrate complementary information from different measurement types, it is important that the advances in the field are translated also to natural product research in order to tap the huge potential of natural products to provide more effective and safe disease treatment and prevention applications.

6 Acknowledgements

MK would like to thank Francesco Iorio, Julio Saez-Rodriguez, Kalai Jayaseelan, Patricia Bento and Anne Hersey for many interesting and useful discussions whilst visiting the European Bioinformatics Institute (EMBL-EBI), Cambridge, UK, in the Spring of 2013. We would also like to thank Kalai Jayaseelan and Sushil Shakyawar for initial investigations into the natural products present in the CMap data and Agnieszka Szwajda, Bhagwan Yadav and Suleiman Khan for their help with the diagrams. Funding: Aalto Science Institute Exploratory grant (MK), Academy of Finland (SM, NS, JT, TA, KW), the Finnish Cancer Organisations (JT, TA, KW), Jane and Aatos Erkko Foundation (KW). Part of this work was carried out whilst MK was also employed at the Department of Mathematics and Systems Analysis, Aalto University, Finland.

7 Notes and references

1 J. Gertsch, Botanical drugs, synergy, and network pharmacology: forth and back to intelligent mixtures, *Planta Med.*, 2011, **77**, 1086–1098.
2 A. L. Harvey, Natural products in drug discovery, *Drug Discovery Today*, 2008, **13**, 894–901.
3 D. C. Swinney, Phenotypic *vs.* Target-Based Drug Discovery for First-in-Class Medicines, *Clin. Pharmacol. Ther.*, 2013, **93**, 299–301.
4 I. Kola and J. Landis, Can the pharmaceutical industry reduce attrition rates?, *Nat. Rev. Drug Discovery*, 2004, **3**, 711–716.

5 L. Hutchinson and R. Kirk, High drug attrition rates—where are we going wrong?, *Nat. Rev. Clin. Oncol.*, 2011, **8**, 189–190.
6 H. Kitano, A robustness-based approach to systems-oriented drug design, *Nat. Rev. Drug Discovery*, 2007, **6**, 202–210.
7 X. Chen, *et al.*, Database of traditional Chinese medicine and its application to studies of mechanism and to prescription validation, *Br. J. Pharmacol.*, 2006, **149**, 1092–1103.
8 A. L. Hopkins, Network pharmacology, *Nat. Biotechnol.*, 2007, **25**, 1110–1111.
9 J. Lehár, *et al.*, Synergistic drug combinations tend to improve therapeutically relevant selectivity, *Nat. Biotechnol.*, 2009, **27**, 659–666.
10 W. Lam, *et al.*, The Four-Herb Chinese Medicine PHY906 Reduces Chemotherapy-Induced Gastrointestinal Toxicity, *Sci. Transl. Med.*, 2010, **2**, 45ra59.
11 R. Stone, Lifting the Veil on Traditional Chinese Medicine, *Science*, 2008, **319**, 709–710.
12 A. S. Azmi, R. M. Mohammad and F. H. Sarkar, Can network pharmacology rescue neutraceutical cancer research?, *Drug Discovery Today*, 2012, **17**, 807–809.
13 F. Iorio, T. Rittman, H. Ge, M. Menden and J. Saez-Rodriguez, Transcriptional data: a new gateway to drug repositioning?, *Drug Discovery Today*, 2013, **18**, 350–357.
14 A. Buriani, *et al.*, Omic techniques in systems biology approaches to traditional Chinese medicine research: present and future, *J. Ethnopharmacol.*, 2012, **140**, 535–544.
15 A. A. Lagunin, *et al.*, Chemo- and bioinformatics resources for *in silico* drug discovery from medicinal plants beyond their traditional use: a critical review, *Nat. Prod. Rep.*, 2014, **31**, 1585–1611.
16 D. J. Barlow, *et al.*, *In silico* studies in Chinese herbal medicines' research: evaluation of *in silico* methodologies and phytochemical data sources, and a review of research to date, *J. Ethnopharmacol.*, 2012, **140**, 526–534.
17 J. Zhao, *et al.*, Therapeutic Effects of Astragaloside IV on Myocardial Injuries: Multi-Target Identification and Network Analysis, *PLoS One*, 2012, **7**, e44938.
18 I. A. Ramallo, M. O. Salazar, L. Mendez and R. L. E. Furlan, Chemically engineered extracts: source of bioactive compounds, *Acc. Chem. Res.*, 2011, **44**, 241–250.
19 J. Clardy and C. Walsh, Lessons from natural molecules, *Nature*, 2004, **432**, 829–837.
20 K. V. Jayaseelan, P. Moreno, A. Truszkowski, P. Ertl and C. Steinbeck, Natural product-likeness score revisited: an open-source, open-data implementation, *BMC Bioinf.*, 2012, **13**, 106.
21 M. Füllbeck, E. Michalsky, M. Dunkel and R. Preissner, Natural products: sources and databases, *Nat. Prod. Rep.*, 2006, **23**, 347–356.
22 T. Liu, Y. Lin, X. Wen, R. N. Jorissen and M. K. Gilson, BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities, *Nucleic Acids Res.*, 2007, **35**, D198–D201.
23 J. Gu, *et al.*, Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology, *PLoS One*, 2013, **8**, e62839.
24 T. M. Ehrman, D. J. Barlow and P. J. Hylands, Virtual Screening of Chinese Herbs with Random Forest, *J. Chem. Inf. Model.*, 2007, **47**, 264–278.
25 F. Luo, J. Gu, L. Chen and X. Xu, Systems pharmacology strategies for anticancer drug discovery based on natural products, *Mol. BioSyst.*, 2014, **10**, 1912–1917.
26 M. N. Patel, M. D. Halling-Brown, J. E. Tym, P. Workman and B. Al-Lazikani, Objective assessment of cancer genes for drug discovery, *Nat. Rev. Drug Discovery*, 2013, **12**, 35–50.
27 Z. Wu, Y. Wang and L. Chen, Network-based drug repositioning, *Mol. BioSyst.*, 2013, **9**, 1268–1281.
28 J. Lamb, *et al.*, The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease, *Science*, 2006, **313**, 1929–1935.
29 X. A. Qu and D. K. Rajpal, Applications of Connectivity Map in Drug Discovery and Development, *Drug Discovery Today*, 2012, **17**, 1289–1298.
30 A. Subramanian, *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15545–15550.
31 N. S. Jahchan, *et al.*, A Drug Repositioning Approach Identifies Tricyclic Antidepressants as Inhibitors of Small Cell Lung Cancer and Other Neuroendocrine Tumors, *Cancer Discovery*, 2013, **3**(12), 1–14.
32 G. Laenen, L. Thorrez, D. Börnigen and Y. Moreau, Finding the targets of a drug by integration of gene expression data with a protein interaction network, *Mol. BioSyst.*, 2013, **9**, 1676–1685.
33 Z. Wu, Y. Wang and L. Chen, Drug repositioning framework by incorporating functional information, *IEE Proc.: Syst. Biol.*, 2013, **7**, 188–194.
34 S. A. Khan, *et al.*, Identification of structural features in chemicals associated with cancer drug response: a systematic data-driven analysis, *Bioinformatics*, 2014, **30**, i497–i504.
35 G. Jin, *et al.*, A novel method of transcriptional response analysis to facilitate drug repositioning for cancer therapy, *Cancer Res.*, 2012, **72**, 33–44.
36 J. A. Parkkinen and S. Kaski, Probabilistic drug connectivity mapping, *BMC Bioinf.*, 2014, **15**, 113.
37 J. R. Pritchard, P. M. Bruno, M. T. Hemann and D. A. Lauffenburger, Predicting cancer drug mechanisms of action using molecular network signatures, *Mol. Biosyst*, 2013, **9**, 1604–1619.
38 S. Banerjee, Y. Li, Z. Wang and F. H. Sarkar, Multi-targeted therapy of cancer by genistein, *Cancer Lett.*, 2008, **269**, 226–242.
39 F. Iorio, *et al.*, Discovery of drug mode of action and drug repositioning from transcriptional responses, *Proc. Natl. Acad. Sci. U. S. A*, 2010, **107**, 14621–14626.
40 D. Carrella, *et al.*, Mantra 2.0: an online collaborative resource for drug mode of action and repurposing by network analysis, *Bioinformatics*, 2014, **30**, 1787–1788.

41 J. Tang, *et al.*, Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis, *J. Chem. Inf. Model.*, 2014, **54**, 735–743.
42 P. Rajendran, E. Ho, D. E. Williams and R. H. Dashwood, Dietary phytochemicals, HDAC inhibition, and DNA damage/repair defects in cancer cells, *Clin. Epigenet.*, 2011, **3**, 4.
43 E. Horia and B. A. Watkins, Complementary actions of docosahexaenoic acid and genistein on COX-2, PGE2 and invasiveness in MDA-MB-231 breast cancer cells, *Carcinogenesis*, 2007, **28**, 809–815.
44 R. E. Tiedemann, *et al.*, Identification of a potent natural triterpenoid inhibitor of proteosome chymotrypsin-like activity and NF-kappaB with antimyeloma activity *in vitro* and *in vivo*, *Blood*, 2009, **113**, 4027–4037.
45 S. D. Kunkel, *et al.*, mRNA Expression Signatures of Human Skeletal Muscle Atrophy Identify a Natural Compound that Increases Muscle Mass, *Cell Metab.*, 2011, **13**, 627–638.
46 S. D. Kunkel, *et al.*, Ursolic Acid Increases Skeletal Muscle and Brown Fat and Decreases Diet-Induced Obesity, Glucose Intolerance and Fatty Liver Disease, *PLoS One*, 2012, **7**, e39332.
47 Z. Wen, *et al.*, Discovery of Molecular Mechanisms of Traditional Chinese Medicinal Formula Si-Wu-Tang Using Gene Expression Microarray and Connectivity Map, *PLoS One*, 2011, **6**, e18278.
48 S. Shi, *et al.*, A Network Pharmacology Approach to Understandingthe Mechanisms of Action of Traditional Medicine: Bushenhuoxue Formula for Treatment of Chronic Kidney Disease, *PLoS One*, 2014, **9**, e89123.
49 M. B. Potts, *et al.*, Using Functional Signature Ontology (FUSION) to Identify Mechanisms of Action for Natural Products, *Sci. Signaling*, 2013, **6**, ra90.
50 E. Wang, *et al.*, Interaction of a traditional Chinese Medicine (PHY906) and CPT-11 on the inflammatory process in the tumor microenvironment, *BMC Med. Genomics*, 2011, **4**, 38.
51 P. Csermely, V. Agoston and S. Pongor, The efficiency of multi-target drugs: the network approach might help drug design, *Trends Pharmacol. Sci.*, 2005, **26**, 178–182.
52 S. Li, B. Zhang, D. Jiang, Y. Wei and N. Zhang, Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae, *BMC Bioinf.*, 2010, **11**(11), S6.
53 T. Xue and R. Roy, Studying Traditional Chinese Medicine, *Science*, 2003, **300**, 740–741.
54 L. Wang, *et al.*, 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**, 4826–4831.
55 H. Fang, *et al.*, Bioinformatics Analysis for the Antirheumatic Effects of Huang-Lian-Jie-Du-Tang from a Network Perspective, *J. Evidence-Based Complementary Altern. Med.*, 2013, e245357.
56 P. Pyrko, A. H. Schönthal, F. M. Hofman, T. C. Chen and A. S. Lee, The unfolded protein response regulator GRP78/BiP as a novel target for increasing chemosensitivity in malignant gliomas, *Cancer Res.*, 2007, **67**, 9809–9816.
57 H. Jo, *et al.*, Natural Product Celastrol Destabilizes Tubulin Heterodimer and Facilitates Mitotic Cell Death Triggered by Microtubule-Targeting Anti-Cancer Drugs, *PLoS One*, 2010, **5**, e10318.
58 E. Chan, M. Tan, J. Xin, S. Sudarsanam and D. E. Johnson, Interactions between traditional Chinese medicines and Western therapeutics, *Curr. Opin. Drug Discovery Dev.*, 2010, **13**, 50–65.
59 J.-L. Tang, B.-Y. Liu and K.-W. Ma, Traditional Chinese Medicine, *Lancet*, 2008, **372**, 1938–1940.
60 J. Tang and T. Aittokallio, Network Pharmacology Strategies Toward Multi-Target Anticancer Therapies: From Computational Models to Experimental Design Principles, *Curr. Pharm. Des.*, 2014, **20**, 23–36.
61 X. Sun, S. Vilar and N. P. Tatonetti, High-throughput methods for combinatorial drug discovery, *Sci. Transl. Med.*, 2013, **5**, 205rv1.
62 Z. Wu, X.-M. Zhao and L. Chen, A systems biology approach to identify effective cocktail drugs, *BMC Syst. Biol.*, 2010, **4**, S7.
63 S. Li, B. Zhang and N. Zhang, Network target for screening synergistic drug combinations with application to traditional Chinese medicine, *BMC Syst. Biol.*, 2011, **5**(1), S10.
64 B. Zhao, J. R. Pritchard, D. A. Lauffenburger and M. T. Hemann, Addressing genetic tumor heterogeneity through computationally predictive combination therapy, *Cancer Discovery*, 2013, **4**, 166–174.
65 J. R. Pritchard, *et al.*, Defining principles of combination drug mechanisms of action, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 403–404.
66 L. Tao, *et al.*, Nature's contribution to today's pharmacopeia, *Nat. Biotechnol.*, 2014, **32**, 979–980.
67 M. Kuhn, *et al.*, Systematic identification of proteins that elicit drug side effects, *Mol. Syst. Biol.*, 2013, **9**, 663.
68 C. Huang, *et al.*, Systems pharmacology in drug discovery and therapeutic insight for herbal medicines, *Briefings Bioinf.*, 2014, **15**, 710–733.
69 A. R. M. R. Amin, *et al.*, Enhanced Anti-tumor Activity by the Combination of the Natural Compounds (−)-Epigallocatechin-3-gallate and Luteolin, *J. Biol. Chem.*, 2010, **285**, 34557–34565.
70 C. Chen and A.-N. T. Kong, Dietary cancer-chemopreventive compounds: from signaling and gene expression to pharmacological effects, *Trends Pharmacol. Sci.*, 2005, **26**, 318–326.
71 L. Gibellini, *et al.*, Quercetin and Cancer Chemoprevention, *J. Evidence-Based Complementary Altern. Med.*, 2011, **2011**, 591356.
72 A. Das, N. L. Banik and S. K. Ray, Flavonoids activated caspases for apoptosis in human glioblastoma T98G and U87MG cells but not in human normal astrocytes, *Cancer*, 2010, **116**, 164–176.
73 D. Ornish, *et al.*, Effect of comprehensive lifestyle changes on telomerase activity and telomere length in men with

biopsy-proven low-risk prostate cancer: 5-year follow-up of a descriptive pilot study, *Lancet Oncol.*, 2013, **14**, 1112–1120.
74 M.-H. Pan, *et al.*, Black tea in chemo-prevention of cancer and other human diseases, *Food Science and Human Wellness*, 2013, **2**, 12–21.
75 M. H. Yang, J. Kim, I. A. Khan, L. A. Walker and S. I. Khan, Nonsteroidal anti-inflammatory drug activated gene-1 (NAG-1) modulators from natural products as anti-cancer agents, *Life Sci.*, 2014, **100**, 75–84.
76 D. Arango, *et al.*, Molecular basis for the action of a dietary flavonoid revealed by the comprehensive identification of apigenin human targets, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, E2153–E2162.
77 C. Zhao, J. Hua, M. L. Bittner, I. Ivanov and E. R. Dougherty, Identifying mechanistic similarities in drug responses, *Bioinformatics*, 2012, **28**, 1902–1910.
78 D. Vidović, A. Koleti and S. C. Schürer, Large-scale integration of small molecule-induced genome-wide transcriptional responses, Kinome-wide binding affinities and cell-growth inhibition profiles reveal global trends characterizing systems-level drug action, *Front. Genet.*, 2014, **5**, 342.
79 H. I. Ingólfsson, *et al.*, Phytochemicals Perturb Membranes and Promiscuously Alter Protein Function, *ACS Chem. Biol.*, 2014, **9**, 1788–1798.
80 F. L. Cherblanc, R. W. M. Davidson, P. Di Fruscia, N. Srimongkolpithak and M. J. Fuchter, Perspectives on natural product epigenetic modulators in chemical biology and medicine, *Nat. Prod. Rep.*, 2013, **30**, 605–624.
81 R. Dashwood, M. C. Myzak and E. Ho, Dietary HDAC inhibitors: time to rethink weak ligands in cancer chemoprevention?, *Carcinogenesis*, 2006, **27**, 344–349.
82 L. Xie, *et al.*, Towards Structural Systems Pharmacology to Study Complex Diseases and Personalized Medicine, *PLoS Comput. Biol.*, 2014, **10**, e1003554.
83 K. Sanderson, Databases aim to bridge the East-West divide of drug discovery, *Nat. Med*, 2011, **17**, 1531.
84 M. Sirota, *et al.*, Discovery and preclinical validation of drug indications using compendia of public gene expression data, *Sci. Transl. Med.*, 2011, **3**, 96ra77.
85 A. Gaulton, *et al.*, ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.*, 2012, **40**, D1100–D1107.
86 J. Zou, *et al.*, Neighbor communities in drug combination networks characterize synergistic effect, *Mol. BioSyst.*, 2012, **8**, 3185–3196.
87 M. P. Gleeson, A. Hersey, D. Montanari and J. Overington, Probing the links between *in vitro* potency, ADMET and physicochemical parameters, *Nat. Rev. Drug Discovery*, 2011, **10**, 197–208.
88 K. C. Bulusu, J. E. Tym, E. A. Coker, A. C. Schierz and B. Al-Lazikani, canSAR: updated cancer research and drug discovery knowledgebase, *Nucleic Acids Res.*, 2014, **42**, D1040–D1047.
89 X. Chen, Z. L. Ji and Y. Z. Chen, TTD: Therapeutic Target Database, *Nucleic Acids Res.*, 2002, **30**, 412–415.
90 F. Zhu, *et al.*, Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 12943–12948.
91 J. Jeon, *et al.*, A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening, *Genome Med.*, 2014, **6**, 57.
92 C. Y.-C. Chen, TCM Database@Taiwan: The World's Largest Traditional Chinese Medicine Database for Drug Screening *In silico*, *PLoS One*, 2011, **6**, e15939.
93 S.-C. Yang, S.-S. Chang, H.-Y. Chen and C. Y.-C. Chen, Identification of potent EGFR inhibitors from TCM Database@Taiwan, *PLoS Comput. Biol.*, 2011, **7**, e1002189.
94 T. M. Ehrman, D. J. Barlow and P. J. Hylands, Phytochemical Databases of Chinese Herbal Constituents and Bioactive Plant Compounds with Known Target Specificities, *J. Chem. Inf. Model.*, 2007, **47**, 254–263.
95 T. M. Ehrman, D. J. Barlow and P. J. Hylands, Phytochemical informatics of traditional Chinese medicine and therapeutic relevance, *J. Chem. Inf. Model.*, 2007, **47**, 2316–2334.
96 T. M. Ehrman, D. J. Barlow and P. J. Hylands, *In silico* search for multi-target anti-inflammatories in Chinese herbs and formulas, *Bioorg. Med. Chem.*, 2010, **18**, 2204–2218.
97 S. Paoletta, *et al.*, Screening of herbal constituents for aromatase inhibitory activity, *Bioorg. Med. Chem.*, 2008, **16**, 8466–8470.
98 J. Ru, *et al.*, TCMSP: a database of systems pharmacology for drug discovery from herbal medicines, *J. Cheminf.*, 2014, **6**, 13.
99 D.-X. Kong, M.-Y. Guo, Z.-H. Xiao, L.-L. Chen and H.-Y. Zhang, Historical Variation of Structural Novelty in a Natural Product Library, *Chem. Biodiversity*, 2011, **8**, 1968–1977.
100 J. P. Overington, B. Al-Lazikani and A. L. Hopkins, How many drug targets are there?, *Nat. Rev. Drug Discovery*, 2006, **5**, 993–996.
101 H. Hieronymus, *et al.*, Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators, *Cancer Cell*, 2006, **10**, 321–330.
102 N. Mukherjee, *et al.*, Coordinated posttranscriptional mRNA population dynamics during T-cell activation, *Mol. Syst. Biol.*, 2009, **5**, 288.
103 M. Liu, *et al.*, Transcriptional profiling of Chinese medicinal formula Si-Wu-Tang on breast cancer cells reveals phytoestrogenic activity, *BMC Complementary Altern. Med.*, 2013, **13**, 11.
104 H.-M. Cheng, *et al.*, Application of bioactivity database of Chinese herbal medicine on the therapeutic prediction, drug development, and safety evaluation, *J. Ethnopharmacol.*, 2010, **132**, 429–437.
105 K. R. Hayes, *et al.*, EDGE: a centralized resource for the comparison, analysis, and distribution of toxicogenomic information, *Mol. Pharmacol.*, 2005, **67**(4), 1360–1368.
106 D. C. Hassane, *et al.*, Discovery of agents that eradicate leukemia stem cells using an *in silico* screen of public gene expression data, *Blood*, 2008, **111**, 5654–5662.

107 Y. Quan, *et al.*, Elucidating Pharmacological Mechanisms of Natural Medicines by Biclustering Analysis of the Gene Expression Profile: A Case Study on Curcumin and Si-Wu-Tang, *Int. J. Mol. Sci.*, 2015, **16**, 510–520.
108 M. Iskar, *et al.*, Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding, *Mol. Syst. Biol.*, 2013, **9**, 662.
109 M. Xiong, Identification of transcription factors for drug-associated gene modules and biomedical implications, *Bioinformatics*, 2014, **30**(3), 305–309.
110 D. C. Hassane, *et al.*, Chemical genomic screening reveals synergism between parthenolide and inhibitors of the PI-3 kinase and mTOR pathways, *Blood*, 2010, **116**, 26.
111 C. Liu, *et al.*, Compound signature detection on LINCS L1000 big data, *Mol. BioSyst.*, 2015, **11**, 714–722.
112 M.-H. Chen, *et al.*, Gene Expression-Based Chemical Genomics Identifies Potential Therapeutic Drugs in Hepatocellular Carcinoma, *PLoS One*, 2011, **6**, e27186.
113 A. Franceschinia, *et al.*, Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**(12), 4548–4553.
114 M. Iskar, *et al.*, Drug-Induced Regulation of Target Expression, *PLoS Comput. Biol.*, 2010, **6**(9), e1000925.
115 R. Edgar, *et al.*, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, *Nucleic Acids Res.*, 2002, **30**(1), 207–210.