Research Article

For reprint orders, please contact reprints@future-science.com

Future Medicinal Chemistry

Extending *in silico* mechanism-of-action analysis by annotating targets with pathways: application to cellular cytotoxicity readouts

Background: An *in silico* mechanism-of-action analysis protocol was developed, comprising molecule bioactivity profiling, annotation of predicted targets with pathways and calculation of enrichment factors to highlight targets and pathways more likely to be implicated in the studied phenotype. **Results:** The method was applied to a cytotoxicity phenotypic endpoint, with enriched targets/pathways found to be statistically significant when compared with 100 random datasets. Application on a smaller apoptotic set (10 molecules) did not allowed to obtain statistically relevant results, suggesting that the protocol requires modification such as analysis of the most frequently predicted targets/annotated pathways. **Conclusion:** Pathway annotations improved the mechanism-of-action information gained by target prediction alone, allowing a better interpretation of the predictions and providing better mapping of targets onto pathways.

Sonia Liggi[1], Georgios Drakakis[1], Alexios Koutsoukas[1], Isidro Cortes–Ciriano[2], Patricia Martínez–Alonso[3,4], Thérèse E Malliavin[2], Adrian Velazquez-Campoy[3,4,5], Suzanne C Brewerton[6], Michael J Bodkin[6], David A Evans[6], Robert C Glen[1], José Alberto Carrodeguas[3,4] & Andreas Bender*[,1]

[1]Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

[2]Unité de Bioinformatique Structurale, Institut Pasteur and CNRS UMR 3825, Structural Biology and Chemistry Department, 25–28, rue du Dr. Roux, 75 015 Paris, France

[3]Institute for Biocomputation and Physics of Complex Systems, Mariano Esquillor s/n, Edificio I + D, Zaragoza 50018, Spain

[4]Department of Biochemistry and Molecular and Cellular Biology, School of Science, Pedro Cerbuna, University of Zaragoza, Zaragoza 50009, Spain

[5]Fundacion ARAID, Diputacion General de Aragon, Spain

[6]Eli Lilly UK, Erl Wood Manor, Windlesham, Surrey GU206PH, UK

*Author for correspondence:
Tel.: +44(0) 1223 762983
Fax: +44 (0) 1223 763076
ab454@cam.ac.uk

Drug discovery research has changed dramatically during its history. Traditionally, it was focused on phenotypic readouts: in this kind of experiments compounds are tested on organisms, tissues or cells showing a particular phenotype [1]. The development of the 'magic bullet' theory by Ehrlich [2], which suggests as a goal the development of a drug that is highly selective against a particular target for a given disease, and more recently the advent of the genomic era have led to a shift away from phenotypic approaches and toward target-based approaches. Shortcomings are present in both phenotypic and target-based methodologies: the former by its nature neglects information on the biological targets of the tested compounds [3], since only the final response caused by a molecule administration is observed. On the other hand, in the latter, efficacy of compounds can often not be sufficiently established until later clinical stages, since a number of important factors, such as the full bioactivity profile of a compound and its ADMET/DMPK (absorption, distribution, metabolism, elimination and toxicity/drug metabolism and pharmacokinetics) properties are not taken into account early on. Hence, both approaches are affected by a lack of information about compound action, which may relate both to efficacy and toxicity. This lack of information represents one of the reasons why drugs often fail during early clinical trials, a key problem for the current state of pharmaceutical industry [4,5]. Recently, phenotypic assays have experienced a resurgence [1,6], in parallel with a movement from the 'one drug–one disease' approach to considering the full polypharmacology of compounds [7–10]. This is due to several reasons, primarily the lack of efficacy of modulating single targets in many disease areas (although also exceptions, such as infectious diseases, exist). The importance of multi-target activity is well established, for example, in CNS-related diseases such as mood disorders and psychiatric diseases, where modulation of multiple G-protein-coupled receptors (GPCRs) is of crucial importance [11]. Phenotypic readouts are now able to take the full bioactivity profile of compounds into account by their very nature, given that a large number of targets which may or may not be modulated are also present in the assay, and their modulation is (or rather should be in case of

FUTURE SCIENCE part of fsg

Key terms

Mechanism of action: Biological response derived from a drug administration, due to the interaction of the molecule with specific protein targets. It defines the functional changes on a molecular level, in contrast with 'mode of action', which instead refers to the changes observed after administration of a substance on a cellular level.

Chemogenomics: Discipline studying the interaction between small molecules and protein targets on a large scale (i.e., considering the activity of multiple ligands against multiple targets at once, as opposed to considering only individual ligand–target interactions).

Target prediction: *In silico* pattern recognition techniques used to infer small molecule–protein interaction by utilizing known bioactivity data of ligands.

Systems biology: Interdisciplinary (biological, mathematical, analytical, etc.) approach that derives from the recognized complexity of living organisms, whose behavior can be fully understood only through consideration of all the dynamic interactions between components of the biological system.

Cytotoxicity: Effects exerted by an agent (either molecules or other cells, such as lymphocytes) that leads to cellular damage and/or death through a mechanism of necrosis or apoptosis.

Apoptosis: Organized process of programmed cell death. In contrast to necrosis, which is the result of acute cell damage and affects neighbor cells, apoptosis does not affect negatively the surrounding tissue.

relevance) also be present in the readouts. However, one of the shortcomings of phenotypic assays is the lack of information about the precise mechanism of action (MoA) of a compound. In this context, chemogenomics has recently been of increasing interest, where it is able to contribute by giving new insights into the MoA [12–16] as well as polypharmacological properties of a molecule [17], thus in addition allowing for drug repositioning [9,18–21], as well as a better understanding of compound side effects [22–26]. In particular, *in silico* target prediction methods appear to be important in filling in the gaps present in chemical bioactivity libraries, which are biased toward a limited number of therapeutically relevant protein targets and do not report the complete bioactivity spectrum of molecules [27]. However, the lack of a complete knowledge regarding bioactivity of compounds represents also a limitation of *in silico* bioactivity deconvolution methodologies. Since target prediction is based on the data available, predicting bioactivity spectra is limited by the ability of the training set to cover both chemical and biological space. Hence, despite their successes in the recent past, *in silico* bioactivity prediction approaches are still presented with some shortcomings that need to be addressed. For example, public data sources can be affected by a number of errors, spanning from the molecular structure to physicochemical properties and the accompanying biological annotations [28]. Thus, the reliability of the data sources used needs to be considered when evaluating the output of an algorithm, in combination with the completeness of chemical and biological space covered by the data [29]. Moreover, another severe shortcoming is that current *in silico* bioactivity profiling tools do not consider the pharmacokinetics/pharmacodynamics (PK/PD) properties of compounds [30], but rather assume that molecules are able to reach and interact with all proteins. Thus, they convert biological responses that depend to a large extent on time and dose, to simplified, binary 'yes/no' decisions about the interaction of a compound with a target.

Furthermore, a number of 'systems biology' factors need to be considered when trying to understand the MoA of a compound better. As shown in Figure 1 (green text), it is fundamental to take into account the chemical structure of the compound studied, the protein(s) it interacts with, as well as the complete network of interacting proteins which can then be linked to the phenotypic response (instead of the protein target itself, which only represents one player along a chain of events). Such a 'systems biology' approach would fill the gap between chemistry and biology, thus potentially helping the drug discovery process in the future [31,32].

We will now summarize current approaches to *in silico* MoA analysis which consider analyses of chemical space, target space and phenotypic readouts, before making a case for using pathway annotations to improve MoA analyses and presenting our results in this area.

A recent application of *in silico* target predictions to *in vitro* and *in vivo* readouts was performed on phenotypic observations from a zebrafish behavioral model [33]. In this study, 681 molecules were selected for their ability to cause characteristic movements in response to a series of light stimuli (a photomotor response). The Similarity Ensemble Approach [34] was then applied to predict protein targets, and the novel molecule–protein pairs obtained were selected for *in vitro* affinity measurement. Eleven compounds, out of the 20 tested, showed activity on 22 of the 31 predicted targets. Further validation of the implication of these targets in the phenotype comes from the comparable photomotor response profile observed for a set of chemically diverse ligands, which were still interacting with the same protein targets. Thus, the combination of target prediction with target-based as well as phenotypic readouts allowed the elucidation of the MoA of the tested compounds. The same target prediction method, Similarity Ensemble Approach, has also been used in what is probably the largest prospective validation of *in silico* target prediction methods. In this case,

predictions were combined with side-effect information on a set 656 US FDA approved drugs, with the aim to elucidate the underlying mechanism of side effects [24]. PK/PD properties of the compound–target–side effect triplets were manually checked to retain only the ones containing compounds likely to reach and interact with the target. This study was able to elucidate the biological mechanism underlying several side effects, such as the relationship between abdominal pain observed after treatment with the synthetic estrogen chlorotrianisene, which could be linked to the inhibition of cyclooxygenase–1 (PTGS1). This study hence demonstrates that also high-level *in vivo* readouts can be analyzed *in silico*; however, care needs to be taken also of other compound properties such as their PK/PD profiles. The above studies do not consider information about the underlying biological network though when analyzing *in silico* MoA predictions, and relatively few studies exist in this area as of now. One of the few applications of a network-based approach comprises the integration of a set of proteins known to be targeted by protoxins, their partners in protein–protein interaction and the pathways and biological processes in which these proteins are implicated [35]. The annotation of these proteins with gene ontology (GO) terms followed by subsequent enrichment calculations found a limited number of biological processes significantly enriched as well as strongly implicated in the toxicity phenotype (such as pathways related to protein folding, immune system regulation and the MAP kinase signaling pathway). Overall, this study shows how the analysis of the targets together with the proteins with which they interact leads to greater insights into the mechanism of toxicity than by considering targets alone, both from the statistical and interpretability point of view – here 186 proteins could be summarized by a much smaller number of biological processes as annotated in GO. In a similar way, annotation of pathways to the known targets of a set of 16 FDA approved drugs helped in better elucidating their MoA [36]. In particular, 11 drugs were associated with enriched pathways relevant for their therapeutic effects. Moreover, pathways implicated in different diseases, which could be explored for drug repositioning, were found to be enriched for four of these drugs. An example is represented by celecoxib, inhibitor of prostaglandin-endoperoxide synthase 2 (COX2): apart from arachidonic acid metabolism and eicosanoid synthesis, the list of enriched pathways for this drug included several cancer-related pathways whose activation by celecoxib was reported in previous experimental studies. Hence, the integration of pathway information appears to be a viable avenue to improve MoA elucidation studies, as well as to identify new therapeutic effects of molecules.

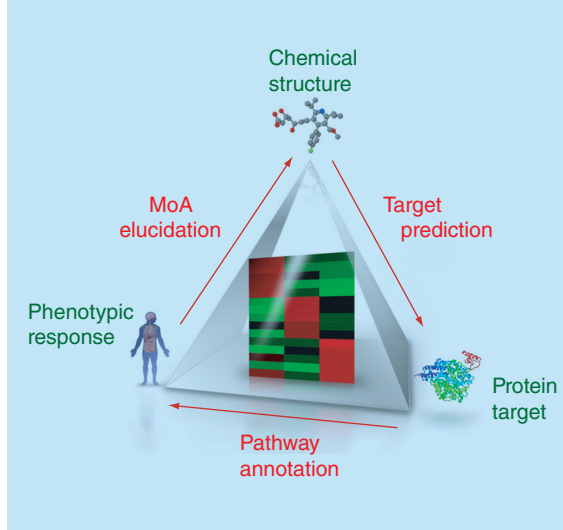


Figure 1. Chemical space (chemical structure of a drug), protein space (proteins: the drug targets) and the phenotypic readout space (symbolized by a human organism) are interconnected and cannot be considered as independent units. The lack of complete information present in each of the three data sources makes the links between them more important: the protocol presented here aims to integrate the information coming from the three vertexes, helping to draw the edges of the triangle. MoA: Mechanism of action.
Reproduced with permission from [22] John Wiley and Sons © (2014).
For color images please see online www.future-science.com/doi/full/10.4155/FMC.14.137

Please note that no *in silico* target prediction was part of the two studies described above; however, the subsequent pathway analysis is similar to what we suggest in our work, namely, a combination of both of those parts, *in silico* target predictions, followed by pathway annotation of the predictions. Integration of pathway information combined to *in silico* bioactivity profiling was performed by Scheiber *et al.* [37], demonstrating how such approaches can help to obtain new insights into the side effects of compounds. A computational pipeline was developed to infer protein targets for a set of molecules responsible for the same side effect, and annotate them with pathways. Application of this pipeline to several datasets allowed the creation of links between side effects and perturbed pathways, as exemplified in the hypotension dataset by the estrogen receptor activation pathway.

In this study, we aim to now describe and validate the integration of *in silico* target prediction methods with subsequent pathways analysis, and to underline the value of integrating both approaches. This will be performed on two sets of compounds which led to cytotoxicity and apoptosis in cell-based screenings, and therefore not influenced by compound PK/PD properties (as would be the case in more complex biological

Key term

Biological pathway: Set of interactions amongst proteins and molecules which leads to a particular cellular, tissue, organ or organism response

systems, such as full organisms). The two datasets have additionally been chosen because of the importance of cytotoxicity, and more specifically apoptosis, in cancer therapeutics. (Tumor cells usually rely on glycolytic energy metabolism instead on mitochondrial oxidative metabolism [38], which allows them to escape the death control usually mediated by the mitochondria [39]. Chemotherapeutic agents should ideally act through apoptosis in order to specifically kill tumor cells and avoid the surrounding tissue damage that would follow a necrotic process.) To achieve this aim, the integration of *in silico* target predictions with automated pathway annotations [16] is used to obtain improved information regarding the MoA of molecules responsible for the cytotoxicity and apoptosis phenotypes (Figure 1, red text). An *in silico* protocol was developed (shown in Figure 2), comprising the first step in bioactivity profile prediction (performed using a Laplacian-modified Naïve Bayes classifier described by Koutsoukas *et al.* [40]) for a dataset of molecules responsible for the same phenotypic response. The predicted protein targets were then annotated with the biological pathways they belong to according to KEGG [41], GO [42] and GO Slim [43] annotations. Subsequently, statistical measures such as enrichment factors were calculated to highlight targets and pathways most likely to be implicated in the phenotype studied. Finally, the results obtained for the cytotoxic dataset were compared with 100 random sets to understand whether integration of pathway information and enrichment calculation on a phenotypic dataset could give more relevant results than on random sets.

Materials & methods

Experimental methods

Apoptotic dataset

Twenty-two compounds were selected from the Prestwick Chemical Library [44] because of their activity in killing embryonic mouse stem cells detected

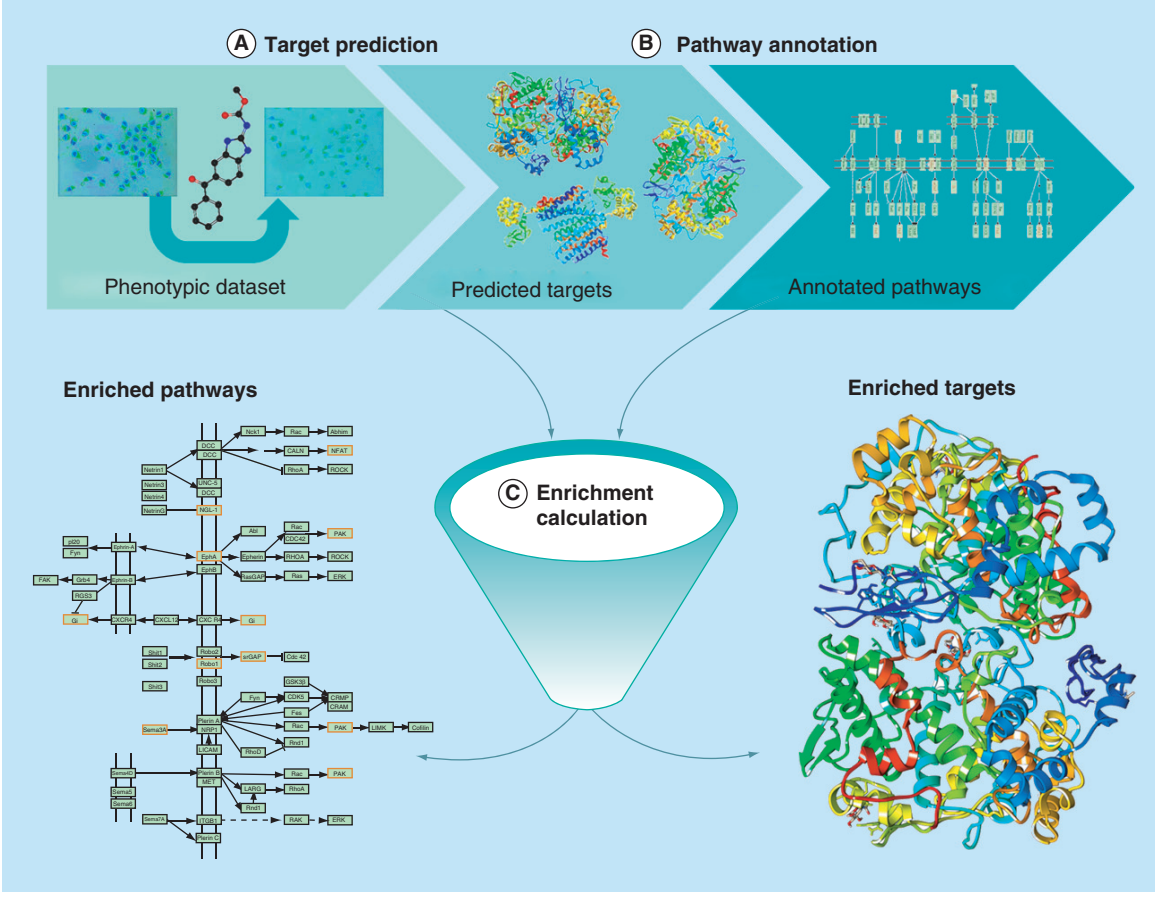


Figure 2. Pipeline of the developed protocol. (A) *In silico* bioactivity prediction is performed on a dataset of small molecules sharing the same phenotype. **(B)** The predicted targets are then annotated with the biological pathways where they are expressed, and finally **(C)** enrichment calculation is performed on both targets and pathways to select the ones statistically more relevant to the phenotype of study to subject to further analysis.

by a colorimetric assay based on metabolic activity performed with the Cell Proliferation Kit II (XTT, Roche), as reported by Conesa *et al.* [45]. These compounds were then used to treat HeLa cells in culture in order to determine if they could induce cell death in these tumor cells. Time lapse fluorescence microscopy was used to characterize the type of cell death and differentiate between apoptosis and necrosis (Leica DMI6000B inverted microscope and MetaMorph software). Cells were cultured in Dulbecco's modified Eagle's medium, supplemented with 10% fetal bovine serum, 2 mM glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin (all reagents from Invitrogen) a 37°C in 5% CO_2 atmosphere. Nuclei and mitochondria were stained by incubating the cells with 250 nM Hoechst 33342 (Invitrogen) and with MitoTracker Green FM (Invitrogen) at 250 nM, respectively, for 30 min in cultured media. The media was then replaced with fresh media, and a 10 µM final concentration of each compound was added; one compound per well. Camptothecin was used as a positive control for apoptosis and DMSO, at 10 µM final concentration, as negative control. Images were acquired every 30 min for 24 h. Three images were acquired at 200× magnification at each time point, using fluorescence filter sets for Hoechst and for MitoTracker Green, and bright field. The three images at each time point were then combined to build a composite image and images from all time points were assembled in a video using MetaMorph software. Cells with shrunk or fragmented nuclei, compacted mitochondria and membrane blebbing, without release of cytosolic content, were considered apoptotic. Ten compounds were clearly inducing apoptotic cell death, and they were therefore selected for further computational analysis.

Cytotoxic dataset

1094 cytotoxic compounds were extracted from PubChem by selecting 186 bioassays describing molecules which had proved cytotoxic to HeLa cells in cell-based assays (Supplementary Table S10) as described by Cortes-Ciriano *et al.* [46]. The targets of these compounds were predicted by the same Laplacian-modified Naïve Bayes classifier [40] used on this work. However, in Cortes-Ciriano *et al.* the classifier was trained on a dataset extracted from ChEMBL version 10 [47] composed of 105,946 compounds and covering 894 human targets. Other differences in the methodology consist in the score cutoffs and enrichment calculation (please refer to the following section 'Computational Methods' for details on the method used here). In the work of Cortes-Ciriano, the predicted targets likely to interact with a compound were selected applying a general score cutoff of 35, and only the top 15 targets were retained and considered enriched. The authors hypothesized that targets enriched from the ones generated by the target prediction algorithm were likely to be mechanistically implicated in cytotoxicity. To validate the hypothesis, the same algorithm was applied to the HitFinder chemical library [48]. Ten compounds exhibiting high scores for two targets previously identified as implicated in cytotoxicity, P-glycoprotein 1 and Topoisomerase I, were prospectively tested to validate or invalidate the hypothesis. The screening led to identify five moderate inhibitors of Topoisomerase I and two inhibitors of PGP 1, confirming the implication of these targets in cytotoxicity. Given the available results obtained by Cortes-Ciriano *et al.*, we decided to use this library to explore pathway annotations on this established dataset.

Computational methods

Compound preprocessing

A pipeline to standardize the datasets was built using KNIME 2.7.4 [49] comprising MOE extensions for KNIME v2.1.8 [50] nodes. The procedure consisted of conversion of structures in 2D (using the node 'SMILES to Molecules') salts stripping and molecules neutralization ('Wash'), aromaticity and tautomerism canonicalization and duplicate removal (both steps performed with the 'Filter Duplicates' node).

Target prediction

In silico bioactivity profiles for the compounds were generated using the Laplacian-modified Naïve Bayes classifier described by Koutsoukas *et al.* [40]. Molprint2D [51] circular fingerprints of depth 3 were then generated using the OpenBabel [52]. The training set consists of 189,147 ligand–protein pairs extracted from ChEMBL version 14 covering 477 human protein targets [16]. The model outputs a confidence score for each target (class) on a per-compound basis. Class-specific score cutoffs were derived by calculating the optimal balanced accuracy (precision and recall trade-off) on a per target class basis, and were used to retain protein targets likely to interact with the compounds in the dataset [16].

Pathway annotation

Protein-pathway annotations were extracted from three different sources, namely, KEGG [41] (release 58.1), GO biological processes [42] (extracted from UniProt 2012_01 on 25 January 2012) and GO Slim biological processes [42] (retrieved using the QuickGO [43] application on 15 October 2012). Different sources were used in parallel to overcome the lack of comprehensive information contained in databases [28–29,53–57], which arises from the different definition of KEGG pathways and GO biological processes. More specifically, KEGG

pathways are defined as molecular interaction/reaction network diagram including genes, proteins, RNAs, chemical compounds, glycans, chemical reactions, disease genes and drug targets. On the other hand, GO biological processes are strictly linked to genes, and represent a recognized series of events or molecular functions with a defined beginning and end, whose alteration can lead to a variation of the phenotype.

It was possible to annotate 405 out of 477 targets with 176 KEGG pathways, while all targets were annotated with 2785 GO biological processes and 173 GO Slim biological processes.

Enrichment calculation

Enrichment calculation was performed to avoid over- and under-predictions/annotations to the extent possible, which could derive from biases in the data used. To this end, a background distribution was assembled in order to create a diverse library of compounds covering large chemical space. Approved drugs from DrugBank, drug-like molecules from PubChem [58], human metabolites from HDMB [59], natural products from ZINC [60] and computationally generated compounds from GDB13 [61] were randomly sampled, yielding 194,849 molecules with a molecular weight ranging from 100 to 900. The background distribution was subjected to compound preprocessing and target prediction to obtain a final dataset of 194,433 molecules.

The *Odds Ratio* for a given target/pathway is calculated by comparing the frequency of prediction/annotation in the test set (F_t) and the frequency of prediction/annotation in the background distribution (F_b), taking into consideration the total number of predictions/annotations in each set (N).

$$Odds\ Ratio = \frac{F_t / N_t}{F_b / N_b}$$

Equation 1

The *Average Ratio* is similar to the *Odds Ratio*, but on a larger scale since 10,000 (n) datasets are used for the calculation. These datasets have the same number of compounds of the test set, and they are randomly generated from the background distribution. The *Average Ratio* for a given target/pathway is then calculated by comparing the frequency of prediction/annotation in each random dataset (R_i) with the frequency of prediction/annotation in the test dataset:

$$Average\ Ratio = \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{F_t}$$

Equation 2

The *Estimation Score* value describes the fraction of random samples whose frequency for a given target/pathway is above the observed one. It is calculated by counting how many times the frequency of prediction/annotation for target/pathway in the aforementioned random datasets is higher or equal to the one in the test dataset ($R_i \geq F_t$). The absolute frequency (C) obtained is then divided by the total number of random dataset n yielding a value between 0 and 1.

$$Estimation\ score = \frac{c}{n}$$

Equation 3

The *Estimation Score* is the first method used to rank targets and pathways according to their enrichment, and it estimates the statistical relevance of predicted targets or annotated pathways, respectively. More specifically, if the *Estimation Score* for a given target equals 1, its frequency of prediction in the random datasets was always found to be higher than the one observed in the test set (and hence it is not found to be significant). Vice versa, if a target presents an *Estimation Score* of 0, its frequency of prediction in the random datasets was never found to be higher than the one observed in the test set (and hence it is statistically significant; the same applies to pathways as well). Hence, if the *Estimation Score* equals 1 it is likely that the high frequency of encountering the target/pathway was a chance event, whereas a value close to 0 suggests a statistically significant target/pathway to be further analyzed and to be explored with respect to its potential biological meaning. In our case, targets and pathways with *Estimation Score* ≤ 0.01 are considered enriched. In case of targets/pathways having the same *Estimation Score*, the two other methods can be indistinctly used to discriminate between them, given their correlation (see Supplementary Information, Enrichment methods comparison). Please note that interpretation of the *Estimation Score* is purely probabilistic, not causal: meaningfulness of the results ultimately needs to be experimentally validated.

Data visualization

A total of 100 random datasets containing the same number of molecules as the cytotoxicity set were extracted from the background distribution using an in-house script. They were subjected to the computational pipeline previously described, and the average number of enriched items per class (targets, KEGG pathways, GO and GO Slim biological processes) was plotted for the random sets and the cytotoxicity set.

Compound-structure, compound-target and compound-pathway binary matrices were generated by linking all compounds with the fingerprint features used to describe them, the targets they are predicted to interact with and the pathways they are annotated with. These matrices were used to construct distance matrices based on the binary method implemented in R [62]. For both binary and distance matrices the

heatmap 2 function of the ggplot package was used, where the complete linkage method was selected to perform hierarchical clustering.

Results & discussion

Cytotoxicity dataset

A library of 1094 compounds cytotoxic for HeLa cells was subjected to the protocol outlined in detail in the Materials & Methods section. The dataset was chosen due to its number of data points, which allows statistically significant results to be obtained, as well the readout being relatively free from PK/PD and other parameters (at least compared with many other, more complex phenotypic readouts). Out of 1094 compounds, 948 successfully passed the standardization protocol, leading to a total of 400 unique predicted targets annotated with 172 unique KEGG pathways, 2565 unique GO biological processes and 168 unique GO Slim biological processes. Enriched targets and pathways will be discussed and linked to the phenotype being studied in the next paragraphs.

Target analysis

The enrichment calculation for 400 predicted targets yielded 151 proteins statistically associated with the cytotoxic phenotype (*Estimation Score* ≤ 0.01), out of which for 110 the *Estimation Score* equals 0 (see Supplementary Table S1 for details). The considerable number of enriched targets is likely due to a combination of the size of the dataset as well as different MoAs involved in the cytotoxicity phenotype. Cytotoxicity can indeed be caused by a large number of compound–protein interactions; a recent study by Flachner *et al.* [63] found a total of ~300 proteins to be targeted by compounds cytotoxic in cancerous cells (265 molecules) as well as healthy cells (251 molecules).

As can be seen in Supplementary Table S1, high frequency of prediction does not always correspond to a high enrichment (see also Supplementary Information, Enrichment methods comparison), thereby underlining the importance of normalizing predictions to a background distribution to eliminate biases in the chemical space, as well as the models, to the extent possible. Overall, 48% of the targets predicted for at least 5% of the compounds (i.e., targets whose absolute frequency of prediction is higher than 47) were found to be not enriched (data not shown). On the other hand, 37% of the enriched targets were predicted for at least 5% of the compounds. Two extreme cases are represented by the proteins thymidylate synthase (TYMS) and carbonic anhydrase 6 (CA6). For the former the *Estimation Score* equals 0 (indicating significance) while the absolute frequency encountered in the dataset is only 5 (see Supplementary Table S1); whereas for the latter the *Estimation Score* equals 1 while the absolute frequency of prediction is as high as 116.

The top enriched targets (defined in the further analysis as those proteins whose *Estimation Score* is equal to 0), are listed in Table 1 along with literature links to cytotoxic phenotype. The most enriched target is NAD(P)H dehydrogenase [quinone] 1 (NQO1, *Average Ratio* = 0.0052). NQO1 is an enzyme that catalyzes the two-electron reduction of quinones to hydroquinones. The cytotoxic activity of the hydroquinones is exploited in cancer therapy, where a number of side effects can be observed because of cytotoxicity in healthy tissue [64]. The Multidrug resistance protein 1 (ABCB1, *Average Ratio* = 0.0818), also known as P-Glycoprotein (PGP), is also important in tumor treatment. ABCB1 is a transmembrane protein responsible for xenobiotics efflux, whose overexpression causes multi-drug resistance (MDR) in response to chemotherapeutic treatment [65]. Moreover, implication of ABCB1 in cytotoxicity was also proven in the study combining *in silico* bioactivity profiling and *in vitro* measurements from which the library analyzed in this work was extracted [46].

A substantial number of phosphorylating proteins (kinases and phosphatases) were found to be enriched. Protein kinases are enzymes whose role consists of adding a phosphate group to other proteins in order to regulate their activity. More than 500 kinases are found in the human genome [77], and their broad expression, together with their implication in a large number of cellular processes, renders them an important family of drug targets in several conditions. An important example is cancer [78], where kinase inhibitors often exert their activity through cytotoxicity. More specifically, modulation of serine/threonine-protein kinase Chk1 (CHEK1, most enriched kinase, *Average Ratio* = 0.0384) and Wee1-like protein kinase (WEE1, *Average Ratio* = 0.0743) both alone and synergistically was recently found to be important in blocking neuroblastoma cell growth [69]. Similarly, the cytotoxic effect responsible for the activity of several anticancer agents has been shown to be linked to inhibition of tyrosine–protein kinase JAK1 (JAK1, *Average Ratio* = 0.0801) [73] and activation of proto-oncogene tyrosine–protein kinase Src (SRC, *Average Ratio* = 0.0881) [76], whereas the role played by tyrosine-protein kinase Lck (LCK, *Average Ratio* = 0.0536) in apoptosis modulation [70] could be responsible for the anticancer activity of natural compounds [71]. Hence, these targets show a clear connection with the cytotoxicity phenotype.

A second group of enriched targets comprises proteins related to DNA regulation, which is essential for cell replication and cell death, and accordingly

Table 1. The top enriched targets for the cytotoxicity dataset are listed along with *Average Ratio* and literature links to the phenotype.

Target	*Average Ratio*	Implication in cytotoxicity	Ref.
NAD(P)H dehydrogenase [quinone] 1 (NQO1)	0.0052	Enzyme that catalyzes the two-electron reduction of quinones to hydroquinones, whose cytotoxic activity is exploited in cancer therapy	[64]
Histone deacetylase 8 (HDAC8)	0.0138	Important in cell cycle and apoptosis regulation. HDAC inhibitors are used in cancer treatment	[66]
Histone deacetylase 10 (HDAC10)	0.0240		
Vasopressin V1b receptor (AVPR1B)	0.0250	AVPR1B plays a proinflammatory role in the gastrointestinal tract, and antagonists of this receptor are able to prevent induced cellular damage	[67]
Telomerase reverse transcriptase (TERT)	0.0372	TERT is fundamental in preserving the telomere length. It is implicated in apoptosis, transcription, metabolic reprogramming, etc	[68]
Serine/threonine-protein kinase Chk1 (CHEK1)	0.0384	Modulation of CHEK1 important in blocking neuroblastoma cell growth	[69]
Tyrosine-protein kinase Lck (LCK)	0.0536	LCK plays an important role in apoptosis modulation which could be responsible for the anticancer activity of natural compounds	[70,71]
Thymidylate synthase (TYMS)	0.0594	TYMS inhibitors are used in cancer therapy given their toxicity caused by interference with mRNA translation and DNA replication	[72]
Wee1-like protein kinase (WEE1)	0.0743	Modulation of WEE1 important in blocking neuroblastoma cell growth	[69]
Tyrosine-protein kinase JAK1 (JAK1)	0.0801	Inhibition of JAK1 is responsible for the cytotoxic effect caused by several anticancer agents	[73]
Multi-drug resistance protein 1 (ABCB1)	0.0818	Transmembrane protein responsible for xenobiotics efflux, whose overexpression causes multi-drug resistance (MDR) in response to chemotherapeutic treatment	[65]
DNA topoisomerase 1 (TOP1)	0.0832	TOP1 inhibitors interfere with mRNA translation and DNA replication. Implication in cytotoxicity was recently proven in a study combining in silico and *in vitro* measurements	[46,74]
P2X purinoceptor 7 (P2RX7)	0.0864	Expressed on the cellular membrane of inflammatory cells, where they interact with extracellular ATP and mediate a number of responses including cytotoxicity. It is upregulated in a number of tumors	[75]
Proto-oncogene tyrosine-protein kinase Src (SRC)	0.0881	Activation of SRC is responsible for the cytotoxic effect caused by several anticancer agents	[76]

All the targets present *Estimation Score* = 0, hence they are ranked according to their *Average Ratio* values. All the most enriched targets found a clear link with the cytotoxicity phenotype.

the proteins contained in this class are in many cases current targets in cancer therapy. The most enriched targets in this cluster are Histone deacetylase 8 and 10 (HDAC8 and HDAC10, whose *Average Ratio* equals 0.0138 and 0.0240, respectively): given the importance of HDACs in cell cycle and apoptosis regulation, HDAC inhibitors have therapeutic application in cancer treatment. [66] Interference with the processes of mRNA translation as well as DNA replication and repair are responsible for the cytotoxic activity of TYMS (*Average Ratio* = 0.0594) inhibitors [72] and DNA topoisomerase 1 (TOP1, *Average Ratio* = 0.0832) inhibitors. [74] As for the previously discussed ABCB1, TOP1 is implicated in cytotoxicity as recently observed by Cortes-Ciriano *et al.* [46]. Hence, also in this second MoA cluster clear links between enriched targets and the cytotoxic phenotype could be established.

The importance of both kinase activity and DNA modulation in cytotoxicity was recently demonstrated by combining *in silico* and *in vitro* screenings to discriminate between cytotoxicity in cancer and healthy cells [63]. In particular, 52% of the cancer selective proteins were kinases, and SRC was found to be one of the most enriched proteins from this group. Moreover, 21% of the enriched proteins played an important role in transcription regulation, with several members of the histone deacetylase family found to be enriched. Even though the dataset used in our analysis was derived from a different source, the results are consistent with the above study, thus

supporting the importance of the targets obtained in cytotoxicity.

Overall, the top enriched targets from all three groups have explicit literature links with the cytotoxic phenotype, confirming the validity of the methodology. In the following we will now compare our analysis of targets to the analysis of annotated pathways.

Pathways analysis

The predicted targets were annotated with pathways extracted from KEGG, GO and GO Slim, and the annotated pathways were subjected to enrichment calculation (see Materials & Methods for details). A total of 77 out of 171 KEGG pathways were found to be enriched, and for 59 of those the *Estimation Score* equals 0 (Supplementary Table S2). Regarding GO Slim biological processes, 62 out of 167 were found to be enriched, for 54 of which the *Estimation Score* equals 0 (Supplementary Table S3). Finally, 1071 out of 2565 GO biological processes were found enriched, 863 of which with an *Estimation Score* of 0 (see Supplementary Table S4). As already observed for targets, a high number of enriched terms is due to both size of the dataset and multiple MoA of the molecules contained in it [63]. Moreover, also in this case a high frequency of annotation for a pathway does not always correspond to a high enrichment value. Taking into consideration the KEGG pathways whose frequency of annotation falls in the top 5% (i.e., whose frequency of annotation is higher than 157), 56% are enriched (data not shown). On the other hand, only for 71% of the enriched KEGG pathways the absolute frequency of annotation is in the highest 5%, so also some pathways relatively rare in absolute numbers can be significant when compared with the background distribution. An example of an enriched KEGG pathway with a low annotation frequency is 'hsa00531 glycosaminoglycan degradation' (*Estimation Score* = 0, frequency of annotation = 10, see Supplementary Table S2), whereas the pathway 'hsa04080 Neuroactive ligand–receptor interaction' presents both high *Estimation Score* (1) and frequency of annotation (3147, data not shown). Similar situations are found also for the GO and GO Slim annotation, confirming the importance of the enrichment calculation to normalize prediction frequencies with regard to a background distribution.

The enriched pathways and biological processes from all ontologies (KEGG, GO and GO Slim) will in the following be discussed jointly, given the considerable overlap that exists between them. In all cases, the enriched pathways were found to be implicated in similar and strongly correlated processes, namely, cancer, immune response, DNA replication and repair and mitosis. In order to focus on the most significant pathways from here on only pathways whose *Estimation Score* equals 0 will be included in the analysis (and pathways are further sorted by the *Average Ratio* measure, see Materials & Methods for details).

The biggest group identified comprises pathways important in cancer development and/or immune response (Table 2), and because of the established link between dysfunction of immunity and carcinogenicity [79] the two classes will here be considered together. Inside this group, several overlaps between annotations coming from the three sources used can be noticed, such as the enriched KEGG pathway 'hsa02010 ABC transporters' (*Average Ratio* = 0.1851), as well as the two enriched, yet more general, GO Slim biological process 'GO:0051181 cofactor transport' (*Average Ratio* = 0.0864) and 'GO:0015893 drug transport' (*Average Ratio* = 0.1166). These entries are lined to processes such as cell detoxification as well as MDR [80] in which the enriched target PGP takes part. Similarly, a subcluster comprising processes important in the activity modulation of Natural Killer cells (NK) can be identified. In particular, the 'hsa04640 Hematopoietic cell lineage' (*Average Ratio* = 0.3520) is a process that starts from hematopoietic stem cell to give several blood cell types, including NK cells [81]. On the other hand, impaired NK function is responsible for 'hsa05340 primary immunodeficiency' (*Average Ratio* = 0.2153), a condition characterized by increased predisposition to infections, autoimmune diseases and cancer [82]. Finally the KEGG pathway 'hsa04520 Adherens junction' (*Average Ratio* = 0.3691) and their components 'hsa04514 Cell adhesion molecules (CAMs)' (*Average Ratio* = 0.3658) all regulate NK function. Nectin, a CAM, is a ligand for the immunoglobulin superfamily of receptors: these receptors are expressed on the outer membrane of NK, and by interacting with nectin modulate the functions of these cells, known to be importantly correlated to pathogenic and tumorigenic responses [83]. As the name suggests, NK cells exert cytotoxic activity on infected, neoplastic or other immune cells [82], thus enrichment of this subcluster of processes regulating their activity fits into the studied phenotype.

The second cluster of pathways, as for targets, is composed of pathways related to DNA and cell cycle (Table 3), and has some general exponents in the GO Slim terms 'GO:0006259 DNA metabolic process' (*Average Ratio* = 0.3956). A strong correlation is observed because of their fundamental role played on DNA repair for the general GO Slim entry 'GO:0006281 DNA repair' (enriched for both GO and GO Slim annotation, with an *Average Ratio* of 0.3026 and 0.3009, respectively), the two KEGG pathways 'hsa00562 inositol phosphate metabolism' and 'hsa03450 Nonhomologous end-joining' (whose

Average Ratio equals 0.1698 and 0.1175, respectively), as well as the GO counterpart of the latter 'GO:0010569 regulation of double-strand break repair via homologous recombination' (*Average Ratio* = 0.0384). As nonhomologous and homologous end joining are two fundamental processes of DNA repair following double-strand breaks [91], they are important in cytotoxicity caused by agents damaging DNA. On the other

Table 2. The most enriched KEGG pathways, GO and GO Slim biological processes related to cancer and immune response.

Enriched pathways and biological processes	Average Ratio	Cancer and immune response	Ref.
KEGG			
hsa00531 Glycosaminoglycan degradation	0.1435	Glycosaminoglycans are targeted by anticancer drugs. They induce cytotoxicity by binding soluble proteins to form amyloid plaques	[84,85]
hsa02010 ABC transporters	0.1851	Linked to processes such as cell detoxification as well as multidrug resistance (MDR)	[80]
hsa05100 Bacterial invasion of epithelial cells	0.1872	The biological processes triggered by pathogens involve the same pathways implicated in autoimmunity. The process is thus compatible with the cytotoxicity phenotype	
hsa04330 Notch signaling pathway	0.2064	Dysfunctions of this pathway play an important role in the pathogenesis of several rumors. It is responsible for the cytotoxic effects of γ-secretase inhibitors, originally designed against Alzheimer's and furthermore been exploited to target various forms of cancer	[86]
hsa05340 Primary immunodeficiency	0.2153	Condition characterized by increased predisposition to infections, autoimmune diseases and cancer	[82]
hsa04630 Jak-STAT signaling pathway	0.2262	Implicated in both autoimmunity and canceroggenity. The enriched member of this pathway JAK1 mediates cytotoxicity after inhibition by anticancer agent	[73, 87–88]
hsa04640 Hematopoietic cell lineage	0.3520	Process that starts from hematopoietic stem cell to give several blood cell types, including NK cells	[81]
hsa04514 Cell adhesion molecules (CAMs)	0.3658	CAMs, in particular nectin, modulate the pathogenic and tumorigenic responses mediated by NK cells	[83]
hsa04520 Adherens junction	0.3691		
hsa04660 T cell receptor signaling pathway	0.6824	The role of T cells is fundamental in the immune response and their relation to cytotoxic activity has been established previously	[89,90]
GO Slim			
GO:0051181 cofactor transport	0.0864	Linked to processes such as cell detoxification as well as multi-drug resistance (MDR)	[80]
GO:0015893 drug transport	0.1166		
GO			
GO:0050862 positive regulation of T cell receptor signaling pathway	0.0536	The role of T cells is fundamental in the immune response and their relation to cytotoxic activity has been established previously	[89,90]
GO:0006027 glycosaminoglycan catabolic process	0.1435	Glycosaminoglycans are targeted by anticancer drugs. They induce cytotoxicity by binding soluble proteins to form amyloid plaques	[84,85]
GO:0030203 glycosaminoglycan metabolic process	0.1536		
GO:0007219 Notch signaling pathway	0.2017	Dysfunctions of this pathway play an important role in the pathogenesis of several rumors. It is responsible for the cytotoxic effects of γ-secretase inhibitors, originally designed against Alzheimer's and furthermore been exploited to target various forms of cancer	[86]

All the pathways and biological processes in this table present *Estimation Score* = 0, hence they are ranked according to their *Average Ratio* values. Given the fundamental role played by these processes in life of cells, they have a clear link with the cytotoxicity phenotype.
GO: Gene Ontology; NK: Natural Killer.

hand, the role of the inositol phosphate metabolic pathway in cytotoxicity is related to both DNA repair and immune-mediated cytotoxicity. For the former, the terminal product inositol hexaphosphate (IP6, polyphosphorylated carbohydrate) is important in activating the Non-homologous end-joining pathway, while the latter consists in activation of NK cells cytolytic activity by the intermediate inositol-1,4,5-trisphosphate (IP3) [92]. A second sub-group is represented by the strongly interconnected processes 'GO:0006333 chromatin

Table 3. The most enriched KEGG pathways, GO and GO Slim biological processes related to DNA regulation and cell cycle are listed along with their *Average Ratio* values and the biological function they mediate.

Enriched pathways and biological processes	*Average Ratio*	DNA regulation and cell cycle	Ref.
KEGG			
hsa03450 Nonhomologous end-joining	0.1175	Fundamental process of DNA repair following double-strand breaks; important in cytotoxicity caused by agents damaging DNA	[91]
hsa00562 Inositol phosphate metabolism	0.1698	The terminal product of this pathway inositol hexaphosphate (IP6) is important in activating the Nonhomologous end-joining pathway, whereas the intermediate inositol-1,4,5-trisphosphate (IP3) activates NK cells cytolytic activity	[92]
GO Slim			
GO:0016458 gene silencing	0.1870	Process that regulates genes expression	
GO:0007059 chromosome segregation	0.2689	Process taking place during DNA replication and in mitosis	[94]
GO:0006397 mRNA processing	0.2918	Process taking place during transcription	
GO:0006281 DNA repair	0.3009	Process taking place after DNA damage	
GO:0006259 DNA metabolic process	0.3956	Any metabolic process involving DNA	
GO			
GO:0006333 chromatin assembly or disassembly	0.0138	Chromatin function is fundamental in DNA transcription	[93]
GO:0022616 DNA strand elongation	0.0372	Telomeres play a fundamental role in DNA replication; alteration of these processes might cause a number of diseases as well as cancer	[68,95]
GO:0007004 telomere maintenance via telomerase	0.0372		
GO:0032203 telomere formation via telomerase	0.0372		
GO:0045839 negative regulation of mitosis	0.0384	Process that stops cellular division	
GO:0010569 regulation of double-strand break repair via homologous recombination	0.0384	Fundamental process of DNA repair following double-strand breaks, important in cytotoxicity caused by agents damaging DNA	[91]
GO:0046602 regulation of mitotic centrosome separation	0.0384	Centrosome plays an important role in mitosis	[96]
GO:0010767 regulation of transcription from RNA polymerase II promoter in response to UV-induced DNA damage	0.0384	Important in preventing the apoptotic processes triggered by UV-induced DNA damage	[97]
GO:2000615 regulation of histone H3-K9 acetylation	0.0384	Histone function is important in cell cycle and DNA transcription	[93]
GO:0035407 histone H3-T11 phosphorylation	0.0384		
GO:0007059 chromosome segregation	0.0832	Process taking place during DNA replication and in mitosis	[94]
GO:0006281 DNA repair	0.3026	Process taking place after DNA damage	

All the pathways and biological processes inthis table present *Estimation Score* = 0, hence they are ranked according to their *Average Ratio* values. Given the fundamental role played by these processes in life/death balance of cells, they have a clear link with the cytotoxicity phenotype.
GO: Gene Ontology; NK: Natural Killer; UV: Ultraviolet.

assembly or disassembly' (*Average Ratio* = 0.0138), 'GO:2000615 regulation of histone H3-K9 acetylation' (*Average Ratio* = 0.0384) and 'GO:0035407 histone H3-T11 phosphorylation' (*Average Ratio* = 0.0379), which can be considered as part of this group given the implication of histone modifications and chromatin function in DNA transcription [93].

Given the fundamental role played by these processes in the life of cells, there is also a clear link to the cytotoxicity phenotype being analyzed here.

As discussed in the previous section, kinases are implicated in a vast number of processes fundamental to cell health maintenance, thus extrapolating their precise action on the cytotoxicity phenotype is not trivial. For this reason, an explicit comparison was performed between the 50 enriched proteins belonging to the kinases class and their annotated 91 KEGG pathways (see Figure 3 for a schematic representation). The 58 enriched KEGG pathways annotated with the enriched kinases mediate the previously discussed functions, namely, immune response, cancer and cell cycle. Hence, analysis of the enriched pathways allows easier interpretation of the data as a number of pathways merge into the few principal biological functions whose regulation is important for the studied phenotype.

To further investigate the validity of the enrichment calculation, an analysis of the KEGG pathways annotated with targets whose *Estimation Score* equals 1 was performed. Overall, it was found that the 74 non-enriched targets were annotated with a total of 94 KEGG pathways. However, a diverse pathway profile was observed for this subset of targets, since only two pathways were present with a relatively high frequency value (namely, 'hsa04080 neuroactive ligand–receptor interaction', whose frequency equals 31, and 'hsa04020 calcium signaling pathway', whose frequency equals 17). It is also remarkable that between the 10 most frequently annotated pathways for this subset (shown in Supplementary Table S5), only one is enriched also for the cytotoxicity dataset, namely the 'hsa05200 Pathways in cancer'. However, it is annotated only with four of the non-enriched targets, which makes it an underrepresented pathway for the subset analyzed. Hence, none of the enriched KEGG pathways is significantly represented among the non-enriched targets, lending credibility that the

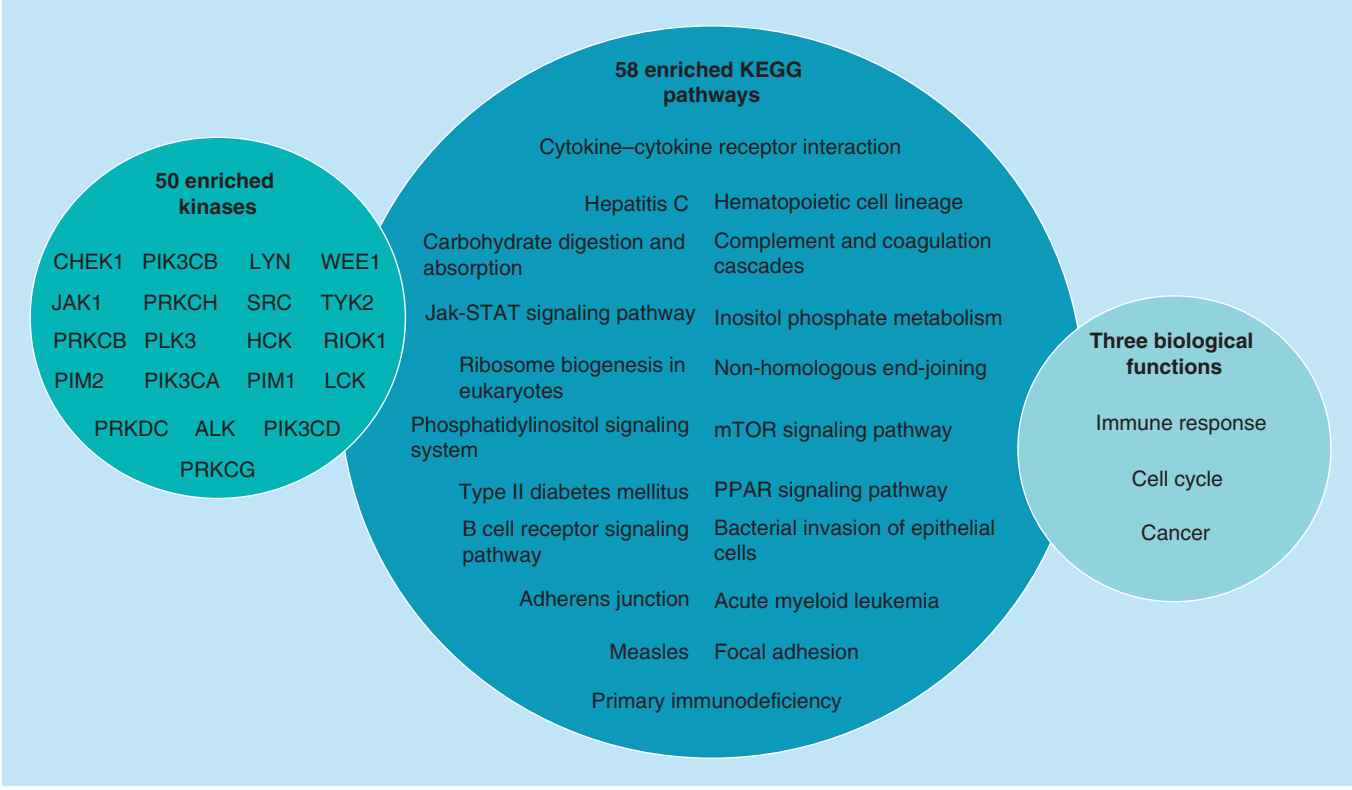


Figure 3. A large number of targets enriched for the cytotoxicity dataset are members of the kinase protein family, a class of proteins implicated in a number of biological processes. All the enriched KEGG pathways annotated with the enriched kinases were collected in order to identify the role of kinases in the cytotoxicity phenotype. In this way, it was possible to observe how three strictly correlated functions are mediated, namely, immune response, cell cycle and cancer (see main text for details). In this representation, only the top 20 enriched targets and KEGG pathways are shown, for the complete list please refer to Supplementary Tables S1 & S2.

enrichment of large number of pathways is unlikely to be the case by chance alone.

To compare our results to existing literature (which did not involve any *in silico* target predictions), most of it was found to be in agreement with a previously reported study integrating a set of proteins known to be targeted by protoxins (responsible for cellular damage through necrosis and/or apoptosis), their partners in protein–protein interactions and the pathways these proteins are involved in [35]. Analysis of the pathways annotated with the 171 proteins directly targeted by protoxins revealed 38 GO biological processes and 15 KEGG pathways to be enriched. The number of enriched terms significantly decreased when considering proteins adducted by reactive metabolites of at least three protoxins and their direct partners (186 targets in total), with three GO processes and eight KEGG pathways significantly enriched. Although these numbers might seem considerably lower than the results obtained here (where we have found 1071 GO processes and 77 KEGG pathways to be enriched), a numerical comparison cannot be performed since neither the complete list nor the total number of annotated terms was released by Fang *et al.* [35]. Moreover, several differences in the methodologies (such as different enrichment calculation and cutoff used) might be responsible for the numerically divergent results. Furthermore, none of the most representatives entries (namely, 'GO:0006457 protein folding', 'GO:0006915 apoptotic process', 'hsa04612 Antigen processing and presentation', 'hsa04010 MAPK signaling pathway') in this previous study has been obtained in our analysis. Nevertheless, several links between both analyses can also be identified. In particular, the term 'GO:0006915 apoptotic process' is equivalent to the KEGG term 'hsa04210 apoptosis' (*Average Ratio* = 0.6449, *Estimation Score* = 0). The fact that the GO term referring to apoptosis was not enriched in our analysis might seem surprising given its obvious relationship with the cytotoxicity phenotype. However, it is a demonstration of the importance of merging several pathway data sources given the incomplete information contained in the available ones [28–29,53–57]. Moreover, 'hsa04612 antigen processing and presentation' can be linked to the enriched immune regulation processes previously discussed. Finally, 'hsa04010 MAPK signaling pathway' is a sub-process of 'hsa05200 pathways in cancer', and can be considered in the same category of the previously discussed pathways implicated in tumor modulation. Thus, in spite of the different nature and MoA of protoxins compared with the molecules in our dataset, the most significantly enriched pathways belong to the same categories. This fact emphasizes the importance of these processes in the life/death balance of cells, as well as further validating our results.

Visualization of distance in structure, target & pathway space between cytotoxic & random sets

In this part of the study we aimed to understand whether pathway annotations, together with enrichment calculation, are able to perform a meaningful mapping of predicted targets onto pathways, in order to allow an easier analysis of the biological processes involved. To address this, 100 datasets of the same size of the cytotoxicity set were randomly selected from the background distribution and processed with the computational pipeline involving target prediction, pathway annotation and enrichment calculation. Binary (Figure 4) and distance (Figure 5) matrices were generated for the test set as well as for one of the randomly selected datasets, in order to analyze similarities in chemical structure, target and pathway space, and understand whether annotation with biological pathways adds any information compared with the classic *in silico* bioactivity profiling where only protein targets are taken into account. In both heatmaps generated from the studied phenotype (Figure 4A & Figure 5A) it is possible to observe how the size of clusters increases when moving from structure space to target space, and finally to pathway space. In particular, in the cytotoxicity compound-fingerprint binary matrix (Figure 4A, left-hand side), the clusters are very sparse and small, as the biggest cluster in the bottom left contains only 7% of the compounds in the dataset and 1.9% of the fingerprint features. In the compound-target matrix (Figure 4A, center) a higher degree of clustering is observed, with the biggest cluster (center-left) comprising 5% of the compounds and 3% of the predicted targets. In the compound-KEGG pathway binary matrix (Figure 4A, right-hand side) the situation is different again, where the largest cluster in the top-right corner contains 41% of the annotated pathways as well as 48% of the compounds. However, a similar behavior of clustering is observed also in the heatmaps obtained for the random dataset (Figure 4B). While no noticeable clustering is observed for the random compound-fingerprint binary matrix (Figure 4B, left), the number and dimension of clusters increases in target and pathway space. In particular, a big cluster is noticeable on the left-hand side of the random compound-target binary matrix (Figure 4B, center), which comprises 7% of the predicted targets and 19% of the compounds. Similarly, the cluster on the top right of the random compound-KEGG heatmap (Figure 4B, right) comprises 48% of both compounds and KEGG pathways. Considering the heatmaps representing distance between compounds in structure space, a similar increase in number of clusters is observed moving from structure to pathways. Moreover, a different profile between cytotoxic (Figure 5A, left) and random (Figure 5B, left) compounds is noticeable, which indicates both a moderate structural

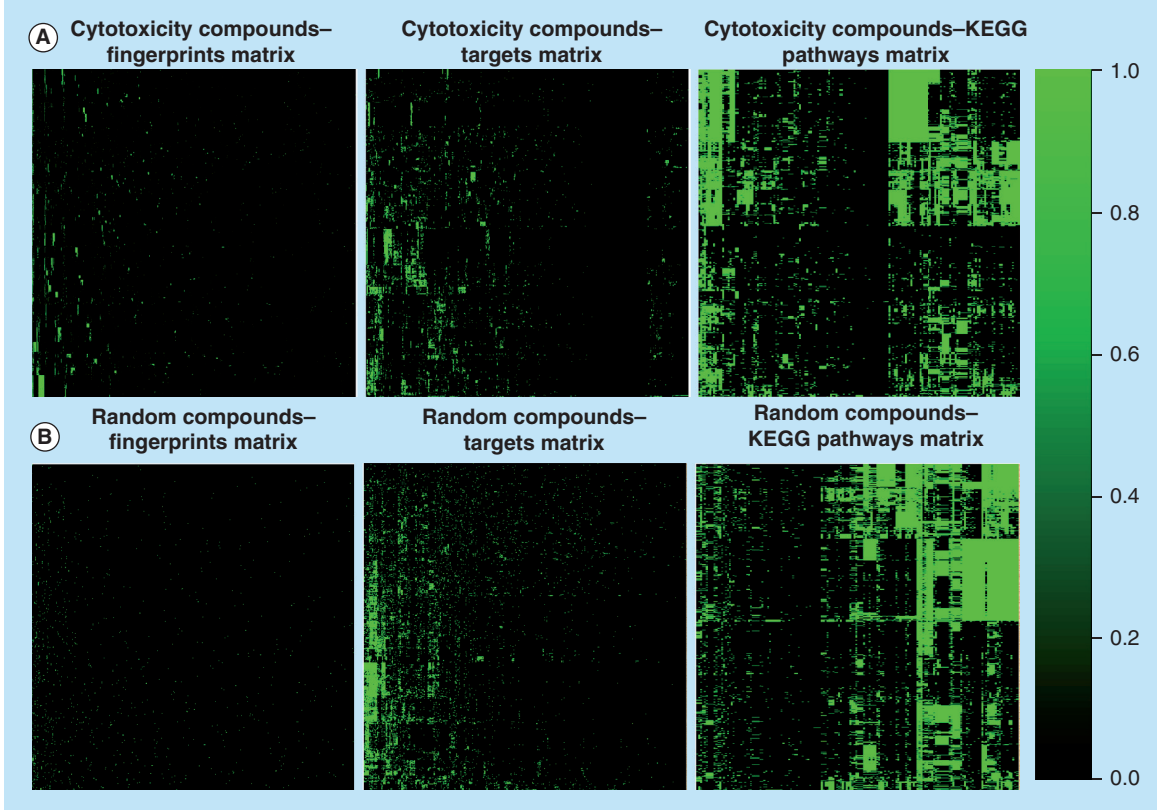


Figure 4. Heatmap representations of binary matrices of cytotoxic compounds. Fingerprints (left), predicted protein targets (center) and annotated KEGG pathways (right) for: **(A)** the cytotoxicity dataset, compared with **(B)**, a random dataset. For every compound, a light green cell indicates presence of the feature, whereas a black cell indicates absence of the feature. The number of clusters in increases moving from the left (fingerprints) to the right (KEGG pathways), as a result of the decreasing number of features describing the compounds. Hence, this indicates that interpretability of mechanism of action is increased by annotating targets with pathways.

similarity among cytotoxic compounds and a successful sampling of diverse molecules in the random set. This could be further explained by analyzing the total number of fingerprint features annotated with the two sets, the frequency of annotation/prediction of these frequency, and the mean and standard deviation among them (Table 4). For cytotoxic compounds, the total number of fingerprint features (1881) is considerably lower than for the random set (3018), whereas the opposite is observed for maximum frequency (474 for cytotoxicity and 179 for random set) and standard deviation (26.97 for cytotoxicity and 6.79 for random set). The higher maximum frequency and lower number of fingerprint features observed for the cytotoxic set indicates a higher similarity among compounds (since more compounds share the same chemical features). On the other hand, the lower values of maximum frequency and standard deviation observed in the random set indicate diversity among compounds, since few compounds share the same substructures.

A higher degree of similarity between cytotoxic molecules (Figure 5A) and between random molecules (Figure 5B) is observed also in target (Figure 5A, center) and pathway (Figure 5A, right) space. However, the number of clusters increases when moving from structure to target to pathway space also in the random set. Although a bigger difference between random and cytotoxic profiles would be expected, it is important to remember that given the dimension of the dataset and the multiple MoA responsible for the cytotoxicity phenotype (as discussed in section 2.1), the total number of predicted targets and annotated pathways will not differ significantly from what observed in random sets (Table 4). Moreover, compounds are described in domains very different in dimension: whereas the number of fingerprint features representing the molecules in the datasets is in the order of thousands, the number of features decreases considerably when considering targets (a maximum of 477 targets can be predicted) and KEGG pathways (a maximum of 176 pathways can be annotated to targets). Hence, the multiple MoAs and the different domains of features used to describe compounds appear to be responsible for the partly similar profiles observed in the cytotoxic and random sets. Further analysis on a set of molecules sharing a more specific mode of action, as well as an

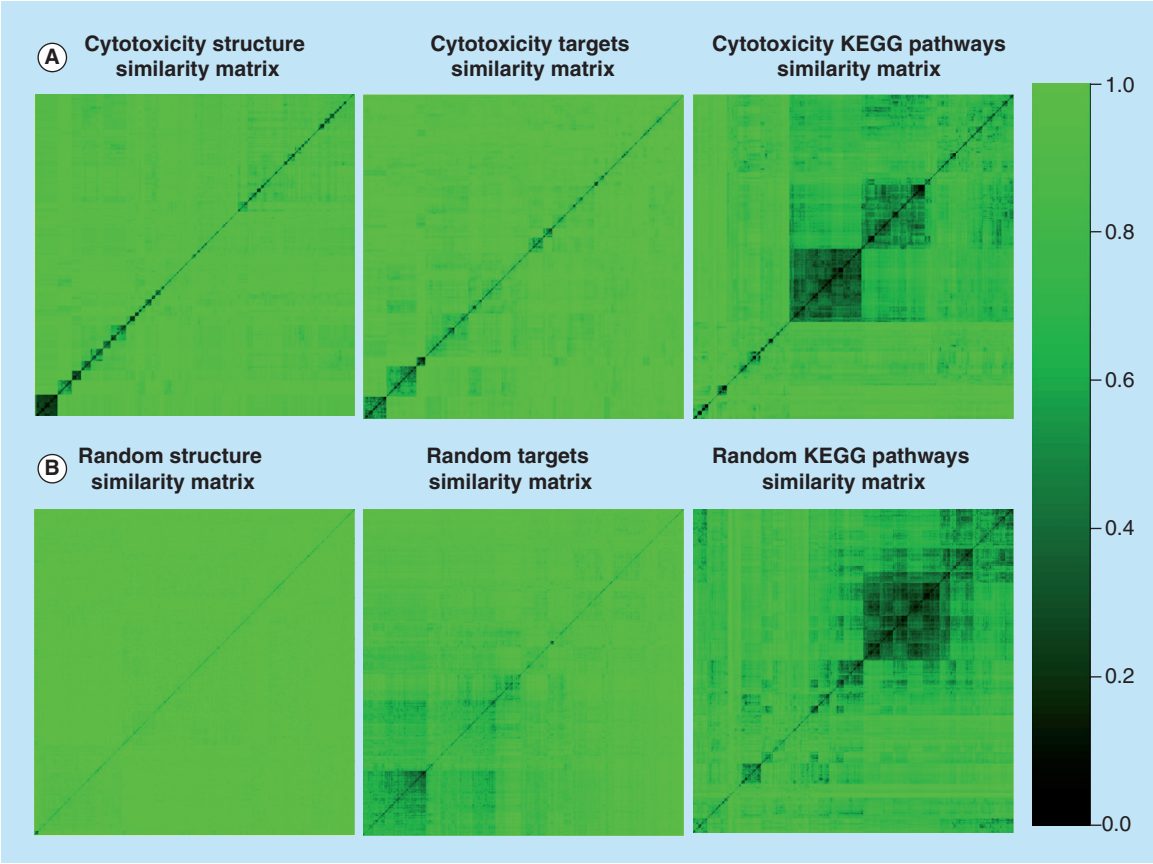


Figure 5. Heatmap representations of distance matrices between cytotoxic compounds. Fingerprints (left), protein targets (center) and KEGG pathways (right) for: **(A)** the cytotoxicity dataset, compared with **(B)** a random dataset. For every compound the darker the cell, the more similar the compounds are. The number of clusters, as well as their dimension, increases moving from the left (fingerprints) to the right (KEGG pathways). This is particularly evident in **(A)**, whereas a smaller degree of clustering is observed in **(B)**. This finding indicates that the information added by pathway annotation is important in order to group compounds having a similar phenotypic readout.

increased number of members in the protein and pathway spaces (which is unfortunately not available at the moment) would help with clarifying this. Nonetheless, the increasing number of clusters observed in pathway space suggests a mapping of multiple targets (as well as multiple compounds) onto the same pathway (confirmed in most of the cases for enriched pathways and biological processes, see Supplementary Figures S1 & S2), hence attesting the help provided by the annotation step in facilitating the analysis of the results.

The validity of our approach on a phenotypic dataset can be further confirmed by considering the number of enriched items for the 100 random sets, which was significantly lower than in the cytotoxicity set. In particular, the maximum number of enriched items in a random set was 12 protein targets, 17 KEGG pathways,

Table 4. The total number of fingerprint features, targets and KEGG pathways predicted/annotated with the cytotoxicity and random datasets.

Dataset	Feature	No of features	Minimum frequency	Maximum frequency	Mean	Standard deviation
Cytotoxicity	Fingerprints	1881	1	474	10.73	26.97
	Targets	400	1	297	36.81	44.69
	Pathways	171	1	3147	378.38	447.77
Random	Fingerprints	3018	1	179	3.33	6.79
	Targets	388	1	503	47.99	87.11
	Pathways	168	1	6332	384.13	677.89

68 GO biological processes and 12 GO Slim biological processes, against 151 targets, 77 KEGG pathways, 1071 GO biological processes and 62 GO Slim found to be enriched in the cytotoxicity set (Figure 6). Moreover, none of these items results particularly over-enriched (Supplementary Figure S3). Hence, this demonstrates how the high number of enriched items for the cytotoxicity dataset does not happen by chance.

Overall, our initial hypothesis that pathway annotation together with enrichment calculations adds information to the MoA elucidation of a set of compounds, compared with target prediction alone, has been confirmed here by showing a better mapping of predicted targets onto annotated pathways and a non-casual enrichment for the cytotoxicity set.

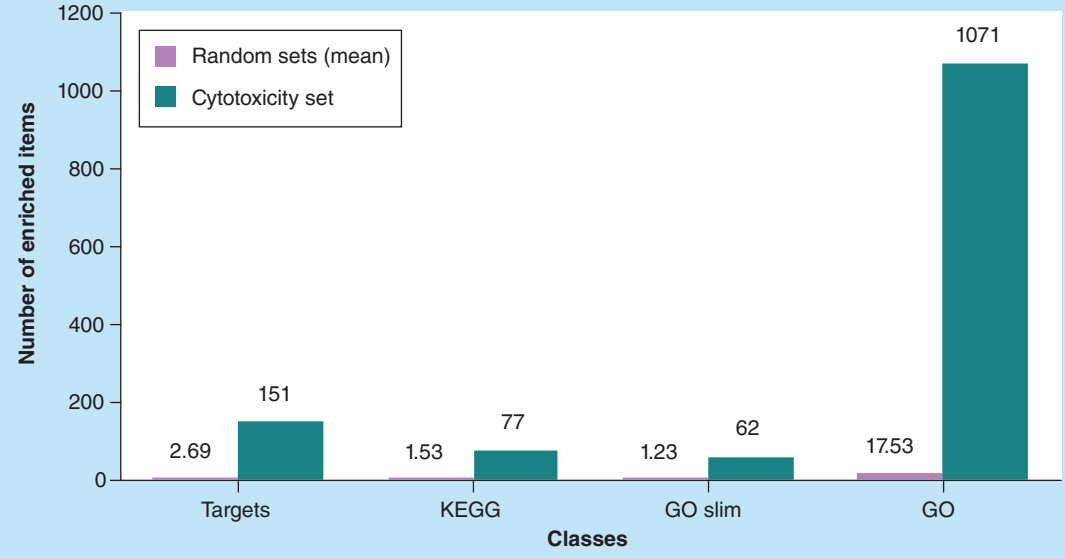
However, it should also be noted that the analysis performed here depends on the size of dataset studied, as we will also outline in more detail in the following section.

Apoptosis dataset

Ten compounds (see Table 5 for the list of compounds along with their biological properties) were selected because of their ability to induce apoptosis. The dataset was subjected to the computational pipeline described in Materials & Methods, yielding 108 unique predicted targets and 109 KEGG, 135 GO Slim and 1180 GO unique annotated processes. Given the diverse bioactivity profiles of the molecules in the apoptotic dataset, together with its small dimension, the enrichment calculations did not lead to statistically significant results. This will be discussed in the following sections, together with the analysis of the most frequently predicted targets and most frequently annotated pathways. Moreover, for each compound the targets and pathways predicted and annotated will be considered individually.

Target analysis

The enrichment calculation gave rise to a limited number of enriched targets (shown in Table 6), with only four proteins (3.7% of the predicted targets) being enriched. This value is particularly low when compared with the cytotoxicity dataset, where 37.8% of the predicted targets were found enriched. Moreover, the enriched targets are not representative of the dataset,



	Targets	KEGG	GO slim	GO
Standard deviation in random sets	2.162	2.692	2.266	15.202
Minimum	0	0	0	0
Maximum	12	17	12	68

Figure 6. The bar plot shows the number of enriched items for the cytotoxicity dataset and the average number of enriched items for the random sets along with the standard deviation in each class (targets, KEGG, GO and GO Slim) and the outliers. The number of enriched items in the cytotoxicity dataset is considerably higher than in the random sets, confirming a non-casual enrichment for the cytotoxicity set.

given that only one was predicted for three compounds, while the remaining three were only predicted for an individual compound from the set. Similar results were also obtained for pathways analysis (see following section), making an enrichment-based analysis on this small dataset difficult. An analysis of the most frequently predicted targets in absolute terms does not lead to more successful results either: the most recurrent target was predicted four times, while frequency for the subsequent seven targets equals three, for the following 21 targets it equals two and finally for the last 79 targets it equals only one (see Table 6 for the top 10 most predicted targets). This observation, together with the lack of significant clusters in the compound-fingerprint and compound-target heatmaps (shown in Figure 7), highlights the multiple and diverse bioactivity profiles of the molecules in the dataset as well as our need for an additional individual analysis of links between compounds and their MoAs.

The ten targets most frequently predicted in absolute terms were subjected to literature search to understand whether they might still be implicated in apoptotic mechanisms (Table 7). As already observed in the cytotoxicity case study, it is possible to cluster targets according to their bioactivity. In particular, seven out of the 10 protein targets are neurotransmitter receptors, namely α-2 adrenergic receptors (ADRA2A, ADRA2B and ADRA2C), D3 dopamine receptor (DRD3), 5-hydroxytryptamine receptor 2B (HTR2B), muscarinic acetylcholine receptors (CHRM1 and CHRM2). ADRA2C is the most predicted receptor (with a frequency of prediction equal to 4), followed by the other two members of the same family ADRA2B and ADRA2C (three predictions in both cases). Implication of ADRA receptors in apoptotic mechanisms is indeed supported by literature: Brimonidine, an ADRA2 agonist, has a protective action against apoptosis induced by glutamate and H_2O_2 in spiral ganglion neurons, which can be reversed by the ADRA2 antagonist yohimbine [113]. On the other hand, yohimbine has been associated with protection from myocardial apoptosis in mice presenting cardiac dysfunction through multiple mechanisms involving inhibition of ADRA2 [114]. Although these two pieces of evidences might seem contrasting, the different tissues in which the measurements were carried out

Table 5. The compounds in the apoptotic dataset are listed along with their main use and their implication in apoptosis.

Compound	Activity	Implication in apoptosis	Ref.
Quinacrine	Antimalarial drug	Suppression of nuclear factor-kappa B (NFKB1) and activation of p53 (TP53) signaling	[98]
Perhexiline	Antianginal agent	Mitochondrial protection from oxidative stress caused by fatty acids loading; interaction with Serine/threonine-protein kinase mTOR	[99,100]
Hycanthone	Antiparasitic, intercalating agent	Inhibition of DNA-(apurinic or apyrimidinic site) lyase (APEX1)	[101]
Puromycin	Antibiotic	Increases caspase 3-mediated apoptosis	[102]
Ellipticine	Antineoplastic agent	Inhibition of RNA polymerase I (POL-I), suppression of TOP2, apoptosis promotion through both caspase-dependent and caspase-independent mechanisms	[103,104]
Benzethonium chloride	Antiseptic drug	Able to cause apoptosis in stem cells but not fibroblasts, increase of cytosolic Ca^{2+} in erythrocytes	[45,105]
Mefloquine	Antimalarial agent	Neurotoxicity associated with apoptosis mediated through activation of Non-receptor tyrosine-protein kinase TYK2, induces apoptosis in breast cancer	[106,107]
Ursolic acid	Natural product used in food and cosmetic industry	Disruption of the anti-apoptotic Glucose-6-phosphate isomerase (PGI) signaling pathway in hepatoma cells, apoptosis induced in colon cancer cells through multiple apoptotic pathways modulation	[108,109]
Tetrandrine	Calcium-channel blocker	Caspase activation, reactive oxygen species (ROS) production, modulation of RAC serine/threonine-protein kinase (AKT) pathway, and inhibition of mitochondrial antiapoptosis proteins	[110]
Cycloheximide	Inhibitor of protein biosynthesis in eukaryotic organisms	Involvement of c-Jun N-terminal kinases (JNKs) and caspase activation in apoptosis induced by combination of Rumor necrosis factor-α (TNF) and cycloheximide, down regulation of FLIP (CASP8 and FADD-like apoptosis regulator, CFLAR) in human fat cells	[111,112]

From the literature links between the compounds and apoptosis it is possible to notice the diverse MoA spectrum of the dataset.
MoA: Mechanism of action.

Table 6. Enriched and ten most frequently predicted targets for the apoptosis dataset.

Targets	Frequency	*Average Ratio*	*Estimation Score*	*Odds Ratio*
Enriched targets				
Equilibrative nucleoside transporter 1	1	0.0057	0.0057	231.6807
Inosine-5′-monophosphate dehydrogenase 1	1	0.0076	0.0076	177.1676
Dopamine D3 receptor	3	0.1558	0.0076	8.8219
C-C chemokine receptor type 4	1	0.0086	0.0086	158.5184
Top 10 predicted targets				
α-2c adrenergic receptor	4	0.3985	0.0605	3.4800
Dopamine D3 receptor	3	0.1558	0.0076	8.8219
Potassium voltage-gated channel subfamily H member 2	3	0.1857	0.0158	7.5993
α-2a adrenergic receptor	3	0.4032	0.1068	3.4376
α-2b adrenergic receptor	3	0.5732	0.2389	2.4103
Serotonin 2b (5-HT2b) receptor	3	0.8509	0.4904	1.6355
Muscarinic acetylcholine receptor M2	3	1.4715	0.8968	0.9449
Muscarinic acetylcholine receptor M1	3	1.5083	0.9063	0.9211
Serine/threonine-protein kinase PIM2	2	0.0868	0.0143	16.2510
Testosterone 17-β-dehydrogenase 3	2	0.0903	0.0127	15.5651

Enriched targets are ranked according firstly to their *Estimation Score* values, and secondly to their *Average Ratio* values, whereas the top 10 targets are ranked according firstly to their frequency of prediction, and secondly to their *Odds Ratio* value. The frequency of prediction of the enriched targets has a low absolute value in every case, indicating that the enriched targets on the apoptotic dataset cannot be considered representative. While the 10 targets most frequently predicted in absolute terms also have a low frequency of prediction, they still cover more compounds and could also be linked to the phenotype in a literature analysis. This illustrates that depending on the size and nature of the dataset a different mechanism of action analysis needs to be performed.

(neurons in the former, myocardial cells in the latter) seem to play a role, and in both cases they show a type of involvement of the ADRA2 receptor in apoptosis. Importance in apoptotic modulation is recognized also for the cholinergic [115], serotoninergic [116] dopaminergic [117] receptors frequently predicted (please refer to Table 7 for details). Similarly, links to apoptosis were found also for the final three protein targets belonging to the top 10 most predicted, namely, testosterone 17-β-dehydrogenase 3 (HSD17B3, two predictions in the dataset) [118], potassium voltage-gated channel subfamily H member 2 (KCNH2, three predictions) [119] and the Serine/threonine-protein kinase pim-2 (PIM2, two predictions) (Table 7) [120].

Hence, all the top 10 predicted targets found literature links with apoptosis. However, none of these targets found a literature annotation with the processes responsible for the apoptotic response mediated by the compounds in the dataset. This finding should not necessarily be interpreted as a failure of prediction, but instead it confirms the importance of *in silico* bioactivity profiling methods in filling in the gaps in the knowledge of the full bioactivity spectrum of molecules, and in suggesting MoA hypotheses in addition to the ones already known. In this particular case, target prediction contributes to the MoA analysis by suggesting novel targets that could contribute to the apoptotic response triggered by the compounds in the dataset.

Individual analysis of the non-frequent targets predicted for each compound was performed to understand whether the predicted targets overlap with the experimentally annotated ones. However, also here only a limited number of MAP kinases, important in apoptotic processes [123], were predicted for few compounds, namely, ellipticine, cycloheximide, hycanthone and quinacrine (data not shown). This result highlights the common limitation of *in silico* bioactivity profiling methodologies that rely on ligand–protein interaction data (see [16] for a recent overview of the field). In this context, consideration of the annotated pathways appears even more important in helping to elucidate MoAs, and this is what we will describe in the following section.

Pathway analysis

A similar situation to the one described for targets was found also in case of the pathways analysis. No KEGG pathways presented an *Estimation Score* higher than the usual cutoff of 0.01, whereas only two GO Slim and 19 GO biological processes were enriched (Supplementary Table S6). This trend is different from that observed in the cytotoxicity analysis, where 45% of the annotated KEGG pathways, 37%

of the annotated GO Slim biological processes and 42% of the annotated GO biological processes were found enriched (compared with 0, 1.5 and 1.6% for the apoptosis phenotype, respectively). Furthermore, the enriched items are not always representative of the dataset, since low absolute frequencies of annotation are recorded. The frequency of annotation equals six for the enriched GO Slim biological process 'GO:0050879 multicellular organismal movement', while it equals only one for 'GO:0015931 nucleobase-containing compound transport'. The situation is slightly different for the GO annotations, where 19 terms are enriched; however, their frequency values span from one to four only, which is as sparse as in case of the enriched targets and GO Slim terms. On the other hand, the absolute frequency of annotation for KEGG (Supplementary Table S7), GO Slim (SupplementaryTableS8)andGO(SupplementaryTableS9) are considerably higher than those recorded for targets, given that a large number of targets merge into fewer but highly populated pathways. However, the high frequency of annotation observed for pathways and biological processes is influenced not only by frequent targets, but also from targets predicted only once which map onto the same pathway. An example is the entry 'hsa04080 neuroactive ligand–receptor interaction', whose absolute frequency of annotation is equal to 56 and is annotated to 35 predicted targets (Supplementary Table S7). However, only 13 of the annotated targets were predicted for more than one compound in the dataset (data not shown). Hence, the absolute frequency of annotation is also not ideal in representing important pathways for the apoptotic dataset. Nonetheless, it is possible to observe a bigger clustering in the compound-KEGG than in the compound-fingerprint and compound-target distance matrices (Figure 7), demonstrating similarity between compounds in pathways space. This suggests

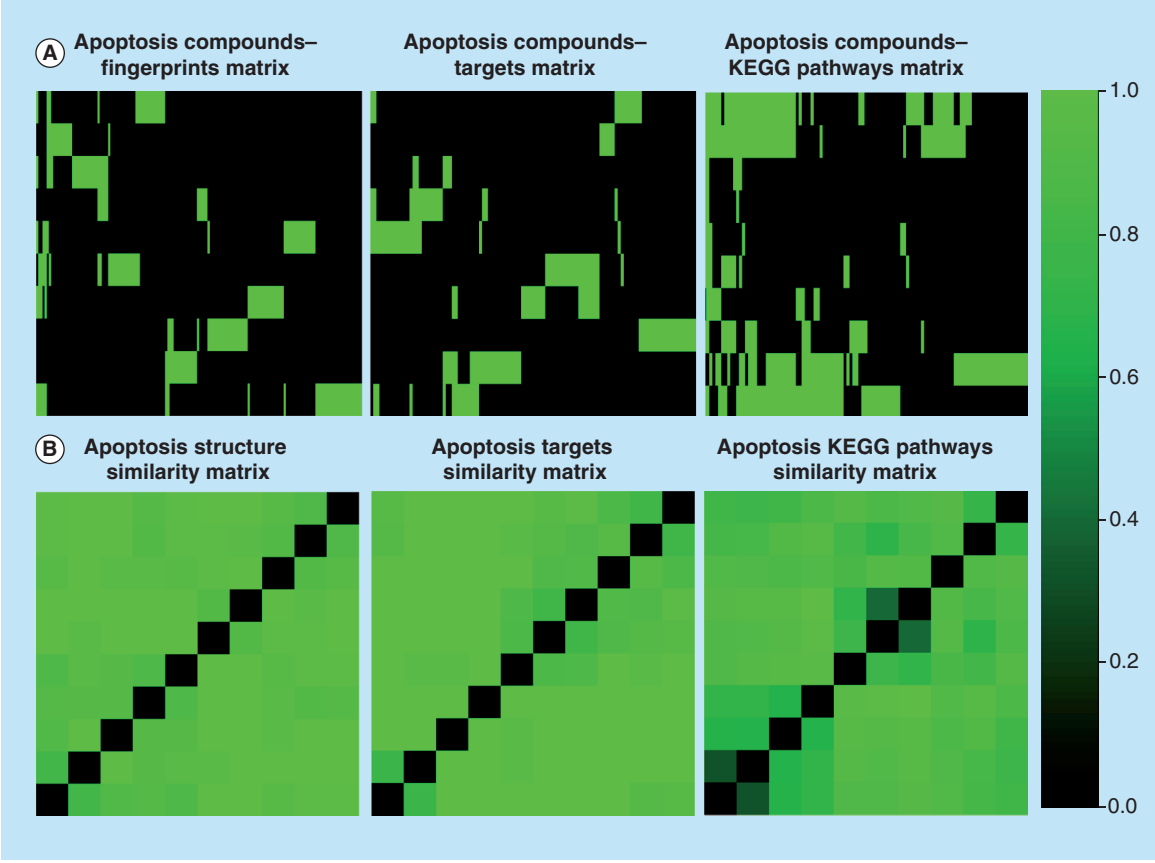


Figure 7. Heatmap representations of matrices between apoptotic compounds based on their fingerprints (left), protein targets (center) and KEGG pathways (right). (A) Binary matrix, every compound in the a light green cell indicates presence of the feature, whereas a black cell indicates absence of the feature. **(B) D**istance matrix, the darker the cell, the more similar the compounds are. In both **(A)** and **(B)** the number of clusters increases upon moving from left (fingerprints) to right (KEGG pathways). The clusters in the target heatmaps are smaller than those in the KEGG heatmaps, highlighting the multiple bioactivity profiles of the molecules in the dataset. This finding confirms the importance of pathway annotation in emphasizing the processes whose modulation by the molecules in the dataset leads to a given biological outcome, overcoming the limitation given by the diverse mechanism of action of the molecules in the dataset.

Table 7. The ten most frequently predicted targets for the apoptosis dataset along with their implication in apoptosis, which was supported by literature for all the protein targets.

Target	Link to apoptosis	Ref.
α-2c adrenergic receptor α-2a adrenergic receptor α-2b adrenergic receptor	Implication of adrenergic receptors in apoptotic mechanisms is supported by literature. In particular, ADRA2 agonist has a protective action against apoptosis in spiral ganglion neurons, whereas ADRA2 antagonist have been associated with protection from myocardial apoptosis	[113,114]
Dopamine D3 receptor	DRD3 mediates the protective response to serum deprivation in peripheral nerve sheath rumor (MPNST) cells through the inhibition of neurofibromatosis type 1 (NF1, proapoptotic) gene expression	[117]
Serotonin 2b (5-HT2b) receptor	Selective antagonism of HTR2B was recently found to cause apoptosis in fibrogenic hepatic stellate cells	[116]
Muscarinic acetylcholine receptor M2 Muscarinic acetylcholine receptor M1	Cholinergic receptors are implicated in apoptotic modulation. In particular, agonism of CHRM2 was shown to block tumor growth through activation of apoptotic mechanisms	[115,121]
Potassium voltage-gated channel subfamily H member 2	KCNH2 is known to be expressed in a number of cancer cells where it controls cell proliferation and apoptosis	[119]
Serine/threonine-protein kinase PIM2	PIM2 is up-regulated in many cancer types (both solid and hematological), where it exerts antiapoptotic regulation	[120]
Testosterone 17-β-dehydrogenase 3	HSD17B3 is an enzyme fundamental for the synthesis of androgens and estrogens, hormones that play a role in modulating pathways implicated in cell viability and rumor cells growth. Inhibitors of this enzyme are being exploited as cancer therapies given their ability to block rumor growth in several types of cancer by an apoptotic mechanism	[118,122]

that pathway annotations are useful to gain improved insight into the MoAs underlying phenotypic readouts of the apoptotic set, as already observed in the case of cytotoxicity. Hence, only the pathways and biological processes annotated with at least two of the top 10 targets will be considered for further (Table 8).

As already observed for the cytotoxicity case study, the most frequent items coming from different data sources can be grouped into multiple related processes. More precisely, a number of processes implicated in ligand–receptor interaction are often annotated with the most frequent targets, namely, the two KEGG annotations 'hsa04080 neuroactive ligand–receptor interaction' (annotated to seven targets) and 'hsa04725 cholinergic synapse' (two annotations), the GO terms 'GO:0007186 G-protein-coupled receptor signaling pathway' (seven annotations), 'GO:0071875 adrenergic receptor signaling pathway' (three annotations), 'GO:0032811 negative regulation of epinephrine secretion' (three annotations), 'GO:0010700 negative regulation of norepinephrine secretion' (three annotations) and finally the term 'GO:0007165 signal transduction' observed for both GO and GO Slim annotation (eight annotations in both cases). The KEGG entry 'hsa04080 neuroactive ligand–receptor interaction' is not a standard pathway, but rather a collection of neuroactive ligands and the receptors with which they interact. However, its high frequency of annotation among frequent targets, and the similarly high frequency of annotation of other processes linked to the ligand–receptor interaction, highlights the activation of neurotransmitters signaling caused by the molecules in the apoptotic dataset already observed with the targets analysis (as discussed in the previous section).

On the other hand, several terms annotated with the most frequent targets can be directly correlated to apoptosis through literature, such as the KEGG pathway 'hsa04020 calcium signaling pathway' (three annotations). Ca^{2+} is a secondary messenger that regulates a plethora of cellular functions, including apoptotic processes through activation of caspases [142]. Not surprisingly, given the important role played by caspases in apoptosis regulation, their activation has been observed for most of the compounds in the dataset, namely, quinacrine [98], puromycin [102], ellipticine [103], mefloquine [106], ursolic acid [108], tetrandrine [110], cycloheximide [111] and benzethonium chloride [45]. Similarly, several frequent pathways and biological processes could be related to the bioactivity of a number of compounds in the dataset (please refer to Table 8 for a detailed discussion).

Table 8. The most frequent KEGG pathways, GO and GO Slim biological processes annotated with the subset of top 10 predicted targets.

Pathways and biological processes	Frequency	Link to apoptosis	Ref.
KEGG			
hsa04080 Neuroactive ligand–receptor interaction	7	Apoptosis is regulated by the predicted targets mapping to this pathway, namely ADRA2, CHRM1 and CHRM2, HTR2B and DRD3	[113–117, 121]
hsa04020 Calcium signaling pathway	3	Ca^{2+} mediated-apoptosis through activation of caspases has been observed for quinacrine, puromycin, ellipticine, mefloquine, ursolic acid, tetrandrine, cycloheximide and benzethonium chloride	[45,98, 102–103, 106,108, 110–111]
hsa04725 Cholinergic synapse	2	Cholinergic receptors are important in apoptosis mediation in cancer cells	[115,121]
hsa04810 Regulation of actin cytoskeleton	2	The actin cytoskeleton is involved in apoptosis. Its regulation is mediated by NADPH oxidase, which can be modulated by perhexiline and puromycin. Moreover, benzethonium chloride and ursolic acid affect cytoskeleton stability	[124–129]
GO			
GO:0007165 signal transduction	8	Biological processes linked to the frequently predicted receptors modulating apoptosis	
GO:0007186 G-protein-coupled receptor signaling pathway	7		
GO:0008284 positive regulation of cell proliferation	4	Any process that activates cell proliferation	
GO:0071883 activation of MAPK activity by adrenergic receptor signaling pathway	3	MAPK signaling is important in apoptotic processes, and can be modulated by cycloheximide as well as ursolic acid pathway	[35,130–131]
GO:0071875 adrenergic receptor signaling pathway	3	Linked to the apoptotic activity mediated by the frequently predicted adrenergic receptors	[113,114]
GO:0044281 small molecule metabolic process	3	Process activated by xenobiotics in general, thus expected to be frequently annotated	
GO:0035625 EGFR transactivation by G-protein-coupled receptor signaling pathway	3	EGFR regulates apoptosis, and disruption of its pathway is responsible for cancer. Quinacrine resistance was observed after activation of EGFR, whereas ellipticine and EGF exhibited a synergistic effect in causing growth inhibition of EGFR-overexpressing breast cancer. Moreover, modulation of EGFR could contribute to the activity of ursolic acid and tetrandrine	[131–135]
GO:0032148 activation of protein kinase B activity	3	Protein kinase B promotes cell survival by inhibiting apoptotic signaling. Quinacrine, mefloquine, ursolic acid and tetrandrine antiproliferative activity on tumor cells is mediated by modulation of the protein kinase B pathway	[135–139]
GO:0010700 negative regulation of norepinephrine secretion	3	Linked to the apoptotic activity mediated by the frequently predicted adrenergic receptors	[113,114]
GO:0032811 negative regulation of epinephrine secretion	3		
GO Slim			
GO:0032501 multicellular organismal process	9	The high frequency of this process derives from misannotation in the data source	
GO:0065007 biological regulation	9	Generic term indicating any process that modulates a biological function	

Table 8. The most frequent KEGG pathways, GO and GO Slim biological processes annotated with the subset of top 10 predicted targets (cont.).

Pathways and biological processes	Frequency	Link to apoptosis	Ref.
GO:0007165 signal transduction	8	Cellular process deriving from reception of a signal which regulates a downstream cellular process. Can be linked to GPCRs activity, and hence to the predicted receptors	
GO:0048856 anatomical structure development	7	Generic term indicating the development of any anatomical structure, which includes apoptotic processes involved in development	
GO:0007267 cell–cell signaling	5	Alteration of cell–cell signaling can lead to apoptosis	[140]
GO:0050896 response to stimulus	5	General term indicating cellular response to an external stimulus	
GO:0030154 cell differentiation	4	Proteins and pathways implicated in apoptosis are proven to be important in activating the cell differentiation process	[141]
GO:0044281 small molecule metabolic process	4	Process activated by xenobiotics in general, thus expected to be frequently annotated	
GO:0009987 cellular process	4	General term referring to any process happening at the cellular level	
GO:0006468 protein phosphorylation	4	Protein kinases modulates a plethora of functions, hence they are fundamental in cell life	[78]

Overall, analysis of the pathways has allowed us to elucidate the MoA of the apoptotic compounds in the dataset in an improved way than considering targets alone. Although the most frequently predicted targets do not represent the experimentally annotated proteins targeted by the molecules in the dataset, they have an implication in apoptosis, thus representing a novel hypothesis on the apoptotic MoA of the compounds. Moreover, the pathways annotated to the most frequent targets reflect the biological mechanisms responsible for the apoptotic response to the compounds in the dataset. Furthermore, the larger clusters observable in both KEGG binary and distance matrices compared with fingerprints and targets (Figure 7) confirm the importance of pathway annotation: they demonstrate similarity between compounds in pathways space, thus highlight the processes whose modulation leads to a given biological outcome, apoptosis in this case. This allows us to overcome the limitation of targets analysis alone, which are caused by the multiple MoA of the molecules in the dataset.

Apart from improving the MoA analysis itself, it should be noted that our pathway annotation approach also permitted us to notice a misannotation in the GO database. More specifically, the GO Slim biological process 'GO:0032501 multicellular organismal process', according to the version of GO Slim used here, is annotated with the predicted protein targets β-2 adrenergic receptor (ADRB2), Ribosomal protein S6 kinase β-1 (RPS6KB1) and D-3 dopamine receptor (DRD3). However, in the current version (checked online on 17 June 2014) annotation of these proteins with the aforementioned biological process is not present, reflecting the incomplete information contained in biological data sources as well as their evolving nature [28–29,53–57].

Conclusion

As shown in this study, the annotation of predicted targets with pathways can give new insights into understanding the MoA of drugs, underlining pathways implicated in the studied phenotype which would not have been detected by only considering the predicted protein targets. This has been evaluated on two datasets, namely, a large cytoxicity dataset, as well as a considerably smaller dataset comprising compounds which elicit an apoptotic response in cells. In the cytotoxicity case study, enrichment calculations of significant targets were found to be fundamental in eliminating biases in the targets predicted. In the second step, pathway annotations lead to an improved understanding of the different pathways that might be implicated in the phenotype. Hence, target prediction increases interpretability for the readouts analyzed here, but even more so the subsequent pathway annotations combined with enrichment calculation, as shown by the comparison of results with random sets. On the other hand, the analysis performed on the apoptotic dataset highlighted that analyses on small datasets need to be performed differently than on larger datasets, and that an analysis of absolute numbers of predicted targets seems to be more appropriate. It is also important to specify that this methodology aims to help the MoA

elucidation of a set of molecules causing a known phenotypic response. For example, the KEGG pathway 'hsa04020 calcium signaling pathway' was found to be the second most annotated pathway for the apoptotic dataset, and indeed its importance in regulating apoptotic processes is well known [142]. However, an interpretation of this finding such as 'every compound interacting with the calcium signaling pathway has an apoptotic activity' would not be valid, since the final cellular effect depends on which step (protein) of the pathway is modulated and the way it is modulated (activation or inhibition, for instance; as well as other factors such as concentration of the compound, etc.). Finally, we have been able to identify inconsistencies in GO annotations (which was corrected in a more recent version) when looking into the most frequently annotated biological processes for the apoptotic dataset, underlining the importance of using reliable data for biological analyses.

Overall, we were able to show that pathway annotations, together with enrichment calculation, add value to *in silico* MoA analyses of phenotypic datasets. Although the two datasets were very different in nature, the added information regarding pathways improved interpretability of the data in both cases. However, statistical significance of the enrichment methodology was proven only for the larger cytotoxicity dataset, illustrating the need to adapt the analysis performed to the specific case studied. This is likely to be the case generally when trying to deconvolute targets from phenotypic screening using *in silico* target prediction algorithms.

Future perspective

The development of computational power is likely to make *in silico* bioactivity analyses even more accessible and easier to perform in the future. At the same time, data sources containing information about the bioactivity of compounds are increasing in size at a high speed. Thus, the combination of both factors is very likely to lead to quantitatively and qualitatively better *in silico* MoA analyses of phenotypic endpoints.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/full/10.4155/FMC.14.137

Financial & competing interests disclosure

This research was supported by Unilever (S.L., A.K., R.C.G. and A.B.), European Research Council (A.B., Grant ERC-2013-StG-336159 MIXTURE), EPSRC (G.D.), Eli Lilly (G.D., S.C.B., M.J.B., D.A.E.), Pasteur–Paris University International PhD Program (I.C.C.), Ibercaja/Universidad de Zaragoza (J.A.C., Grant UZ2011–BIO–02), Institute for Biocomputation and Physics of Complex Systems (J.A.C. and P.M.A., Intramural grant), Spanish Ministerio de Ciencia e Innovacion (A.V.C., Grant BFU2010–19451), Centre National de la Recherche Scientifique (T.E.M.), and the Pasteur Institute (I.C.C. and T.E.M.). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- We evaluated the annotation of pathways onto *in silico* target predictions on two different datasets with very different characteristics – a cytotoxicity dataset containing more than 1000 members and a apoptosis dataset which only comprised 10 compounds.
- We have found that the integration of compound structures, protein targets modulated, as well as pathway information gives a better understanding of the mechanism of action of a compound than using activities against protein targets alone.
- Pathways also show significantly more clustering than compound structure similarities, or target/bioactivity profile similarities, hence facilitating the analysis of subclusters which cause the same phenotype, but via different mechanisms of action.
- The analysis performed needs to be tailored to the particular phenotypic dataset, as clearly demonstrated when describing the differences between the cytotoxic and apoptotic endpoints considered in the current study.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

1 Eggert US. The why and how of phenotypic small-molecule screens. *Nat. Chem. Biol.* 9(4), 206–209 (2013).

• **Discusses importance of phenotypic screenings in drug discovery research.**

2 Drews J. Case histories, magic bullets and the state of drug discovery. *Nat. Rev. Drug Discov.* 5(8), 635–640 (2006).

3 Feng Y, Mitchison TJ, Bender A, Young DW, Tallarico JA. Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. *Nat. Rev. Drug Discov.* 8(7), 567–578 (2009).

• **Summarizes techniques used for phenotypic profiling, and the importance of integrating these methodologies with in silico approaches in the drug discovery process.**

4 Swinney DC, Anthony J. How were new medicines discovered? *Nat. Rev. Drug Discov.* 10(7), 507–519 (2011).

5 Sams-Dodd F. Is poor research the cause of the declining productivity of the pharmaceutical industry? – An industry in need of a paradigm shift. *Drug Discov. Today* 18(5–6), 211–217 (2013).

6 Schenone M, Dančík V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.* 9(4), 232–240 (2013).

7 Mencher SK, Wang LG. Promiscuous drugs compared with selective drugs (promiscuity can be a virtue). *BMC Clin. Pharmacol.* 5(3) (2005).

8 Morphy R, Rankovic Z. Fragments, network biology and designing multiple ligands. *Drug Discov. Today* 12(3–4), 156–160 (2007).

9 Medina-Franco JL, Giulianotti MA, Welmaker GS, Houghten RA. Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discov. Today* 18(9–10), 495–501 (2013).

10 Brown JB, Okuno Y. Systems biology and systems chemistry: new directions for drug discovery. *Chem. Biol.* 19(1), 23–28 (2012).

11 Bianchi MT, Botzolakis EJ. Targeting ligand-gated ion channels in neurology and psychiatry: is pharmacological promiscuity an obstacle or an opportunity? *BMC Pharmacol.* 10(3) (2010).

12 Koutsoukas A, Simms B, Kirchmair J *et al.* From *in silico* target prediction to multi-target drug design: current databases, methods and applications. *J. Proteomics* 74(12), 2554–2574 (2011).

•• **Summarizes *in silico* approaches, data sources and applications.**

13 Gedeck P, Kramer C, Ertl P. Computational analysis of structure-activity relationships. *Prog. Med. Chem.* 49(10), 113–160 (2010).

14 Glick M, Jacoby E. The role of computational methods in the identification of bioactive compounds. *Curr. Opin. Chem. Biol.* 15(4), 540–546 (2011).

15 Chua HN, Roth FP. Discovering the targets of drugs via computational systems biology. *J. Biol. Chem.* 286(27), 23653–23658 (2011).

16 Liggi S, Drakakis G, Hendry AE *et al.* Extensions to *in silico* bioactivity predictions using pathway annotations and differential pharmacology analysis: application to xenopus laevis phenotypic readouts. *Mol. Inf.* 32(11–12), 1009–1024 (2013).

•• **Reviews *in silico* bioactivity profiling methods, applications and pitfalls, together with the presentation of two extensions applied on a phenotypic dataset.**

17 Xie L, Xie L, Kinnings SL, Bourne PE. Novel computational approaches to polypharmacology as a means to define responses to individual drugs. *Annu. Rev. Pharmacol. Toxicol.* 52, 361–379 (2012).

18 Bender A, Young D, Jenkins J *et al.* Chemogenomic data analysis: prediction of small-molecule targets and the advent of biological fingerprints. *Comb. Chem. High Throughput Screen.* 10(8), 719–731 (2007).

19 Jacoby E. Computational chemogenomics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1(1), 57–67 (2011).

20 Oprea TI, Nielsen SK, Ursu O *et al.* Associating drugs, targets and clinical outcomes into an integrated network affords a new platform for computer-aided drug repurposing. *Mol. Inf.* 30(2–3), 100–111 (2011).

21 Napolitano F, Zhao Y, Moreira VM *et al.* Drug repositioning: a machine-learning approach through data integration. *J. Cheminform.* 5(1), 30 (2013).

22 Bender A, Scheiber J, Glick M *et al.* Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem.* 2(6), 861–873 (2007).

• ***In silico* bioactivity profiling is used to combine preclinical safety pharmacology and adverse drug reactions information, leading to new hypothesis linking drugs, targets and side effects.**

23 Tatonetti NP, Liu T, Altman RB. Predicting drug side-effects by chemical systems biology. *Genome Biol.* 10(9), 238 (2009).

24 Lounkine E, Keiser MJ, Whitebread S *et al.* Large-scale prediction and testing of drug activity on side-effect targets. *Nature* 486(7403), 361–367 (2012).

•• **Large-scale prospective validation of target prediction algorithms which shows how target prediction can help understanding promiscuity.**

25 Brouwers L, Iskar M, Zeller G, van Noort V, Bork P. Network neighbors of drug targets contribute to drug side-effect similarity. *PLoS ONE* 6(7), e22187 (2011).

26 Yamanishi Y, Kotera M, Kanehisa M, Goto S. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 26(12), i246–i254 (2010).

27 Mestres J, Gregori-Puigjané E, Valverde S, Solé R V. Data completeness – the Achilles heel of drug-target networks. *Nat. Biotech.* 26(9), 983–984 (2008).

28 Williams AJ, Ekins S, Tkachenko V. Towards a gold standard: regarding quality in public domain chemistry databases and approaches to improving the situation. *Drug Discov. Today* 17(13–14), 685–701 (2012).

29 Tiikkainen P, Franke L. Analysis of commercial and public bioactivity databases. *J. Chem. Inf. Model.* 52(2), 319–326 (2012).

30 Ortwine DF, Aliagas I. Physicochemical and DMPK in silico models: facilitating their use by medicinal chemists. *Mol. Pharm.* 10(4), 1153–1161 (2013).

31 Hood L, Perlmutter RM. The impact of systems approaches on biological problems in drug discovery. *Nat. Biotechnol.* 22(10), 1215–1217 (2004).

32 Schrattenholz A, Soskič V. What does systems biology mean for drug development? *Curr. Med. Chem.* 15(15), 1520–1528 (2008).

33 Laggner C, Kokel D, Setola V *et al.* Chemical informatics and target identification in a zebrafish phenotypic screen. *Nat. Chem. Biol.* 8(2), 144–146 (2012).

34 Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* 25(2), 197–206 (2007).

- **The *in silico* bioactivity profiling method similarity ensemble approach, extensively used in a number of applications, is here presented.**

35 Fang J, Koen YM, Hanzlik RP. Bioinformatic analysis of xenobiotic reactive metabolite target proteins and their interacting partners. *BMC Chem. Biol.* 9, 5 (2009).

36 Pan Y, Cheng T, Wang Y, Bryant SH. Pathway analysis for drug repositioning based on public database mining. *J. Chem. Inf. Model.* 54(2), 407–418 (2014).

37 Scheiber J, Chen B, Milik M *et al.* Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis. *J. Chem. Inf. Model.* 49(2), 308–317 (2009).

38 Mak L, Liggi S, Tan L *et al.* Anti-cancer drug development: computational strategies to identify and target proteins involved in cancer metabolism. *Curr. Pharm. Des.* 19(4), 532–577 (2013).

39 Tait SWG, Green DR. Mitochondria and cell death: outer membrane permeabilization and beyond. *Nat. Rev. Mol. Cell Biol.* 11(9), 621–632 (2010).

40 Koutsoukas A, Lowe R, Kalantarmotamedi Y *et al.* In silico target predictions: defining a benchmarking data set and comparison of performance of the multiclass naïve bayes and parzen-rosenblatt window. *J. Chem. Inf. Model.* 53(8), 1957–1966 (2013).

41 Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28(1), 27–30 (2000).

42 The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25(1), 25–29 (2000).

43 Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 25(22), 3045–3046 (2009).

44 Prestwick Chemical, Illkirch, France. www.prestwickchemical.com

45 Conesa C, Doss MX, Antzelevitch C, Sachinidis A, Sancho J, Carrodeguas JA. Identification of specific pluripotent stem cell death--inducing small molecules by chemical screening. *Stem Cell Rev. Rep.* 8(1), 116–127 (2012).

46 Cortes-Ciriano I, Koutsoukas A, Abian O, Glen RC, Velazquez-Campoy A, Bender A. Experimental validation of in silico target predictions on synergistic protein targets. *Medchemcomm* 4(1), 278–288 (2013).

47 Gaulton A, Bellis LJ, Bento AP *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 44(13), 1–8 (2011).

48 HitFinder: Maybridge HitFinder Library version 11 (14,400 compounds, accessed 8 Nov 2011).

49 Berthold MR, Cebron N, Dill F, Gabriel TR. KNIME: the Konstanz information miner. In: *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007).* Springer, 319–326 (2008).

50 MOE Molecular Operating Environment. Chemical Computing Group. www.chemcomp.com.

51 Bender A, Mussa HY, Glen RC, Reiling S. Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance. *J. Chem. Inf. Comput. Sci.* 44(5), 1708–1718 (2004).

52 O'Boyle NM, Banck M, James C a, Morley C, Vandermeersch T, Hutchison GR. Open babel: an open chemical toolbox. *J. Cheminf.* 3(1), 33 (2011).

53 Williams AJ, Ekins S. A quality alert and call for improved curation of public chemistry databases. *Drug Discov. Today* 16(17–18), 747–750 (2011).

54 Hettne K, Cases M, Boyer S, Mestres J. Connecting small molecules to nuclear receptor pathways. *Curr. Top. Med. Chem.* 7(15), 1530–1536 (2007).

55 Shmelkov E, Tang Z, Aifantis I, Statnikov A. Assessing quality and completeness of human transcriptional regulatory pathways on a genome-wide scale. *Biol. Direct* 6, 15 (2011).

56 Soh D, Dong D, Guo Y, Wong L. Consistency, comprehensiveness, and compatibility of pathway databases. *BMC Bioinformatics* 11, 449 (2010).

57 Altman T, Travers M, Kothari A, Caspi R, Karp PD. A systematic comparison of the MetaCyc and KEGG pathway databases. *BMC Bioinformatics* 14(1), 112 (2013).

58 Bolton EE, Wang Y, Thiessen PA. PubChem: integrated platform of small molecules and biological activities. *Annu. Rep. Comput. Chem.* 4, 217–241 (2008).

59 Wishart DS, Jewison T, Guo AC *et al.* HMDB 3.0 – the human metabolome database in 2013. *Nucleic Acids Res.* 41(Database issue), D801–D807 (2013).

60 Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* 52(7), 1757–1768 (2012).

61 Blum LC, Reymond J-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.* 131(25), 8732–8733 (2009).

62 R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

63 Flachner B, Lörincz Z, Carotti A *et al.* A chemocentric approach to the identification of cancer targets. *PLoS ONE* 7(4), e35582 (2012).

64 Siegel D, Yan C, Ross D. NAD(P)H: quinone oxidoreductase 1 (NQO1) in the sensitivity and resistance to antitumor quinones. *Biochem. Pharmacol.* 83(8), 1033–1040 (2012).

65 Gillet J-P, Efferth T, Remacle J. Chemotherapy-induced resistance by ATP-binding cassette transporter genes. *Biochim. Biophys. Acta.* 1775(2), 237–262 (2007).

66 Jazirehi AR. Regulation of apoptosis-associated genes by histone deacetylase inhibitors: implications in cancer therapy. *Anticancer Drugs* 21(9), 805–813 (2010).

67 Ferrier L, Serradeil-Le Gal C, Schulte AM *et al.* Proinflammatory role of vasopressin through V1b receptors in hapten-induced experimental colitis in rodents: implication in IBD. *Am. J. Physiol. Gastrointest. Liver Physiol.* 299(6), G1298–G1307 (2010).

68 Low KC, Tergaonkar V. Telomerase: central regulator of all of the hallmarks of cancer. *Trends Biochem. Sci.* 38(9), 426–434 (2013).

69 Russell MR, Levin K, Rader J *et al.* Combination therapy targeting the Chk1 and Wee1 kinases shows therapeutic efficacy in neuroblastoma. *Cancer Res.* 73(2), 776–784 (2013).

70 Patrussi L, Giommoni N, Pellegrini M, Gamberucci A, Baldari CT. p66Shc-dependent apoptosis requires Lck and CamKII activity. *Apoptosis* 17(2), 174–186 (2012).

71 Robles-Escajeda E, Lerma D, Nyakeriga AM *et al.* Searching in mother nature for anti-cancer activity: anti-proliferative and pro-apoptotic effect elicited by green barley on leukemia/lymphoma cells. *PLoS ONE* 8(9), e73508 (2013).

72 Garg D, Henrich S, Salo-Ahen OMH, Myllykallio H, Costi MP, Wade RC. Novel approaches for targeting thymidylate synthase to overcome the resistance and toxicity of anticancer drugs. *J. Med. Chem.* 53(18), 6539–6549 (2010).

73 Kao C, Chao A, Tsai CL *et al.* Phosphorylation of signal transducer and activator of transcription 1 reduces bortezomib-mediated apoptosis in cancer cells. *Cell Death Dis.* 4(2), e512 (2013).

74 Pommier Y. Drugging topoisomerases: lessons and challenges. *ACS Chem. Biol.* 8(1), 82–95 (2013).

75 Jelassi B, Anchelin M, Chamouton J *et al.* Anthraquinone emodin inhibits human cancer cell invasiveness by antagonizing P2X7 receptors. *Carcinogenesis* 34(7), 1487–1496 (2013).

76 Slingerland M, Cerella C, Guchelaar HJ, Diederich M, Gelderblom H. Cardiac glycosides in cancer therapy: from preclinical investigations towards clinical trials. *Invest. New Drugs* 31(4), 1087–1094 (2013).

77 Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science* 298(5600), 1912–1934 (2002).

78 Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* 9(1), 28–39 (2009).

79 Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell* 140(6), 883–899 (2010).

80 Kapoor K, Sim HM, Ambudkar S V. Multidrug resistance in cancer: A tale of ABC drug transporters. In: *Molecular Mechanisms of Tumor Cell Resistance to Chemotherapy Resistance to Targeted Anti-Cancer Therapeutics.* Bonavida B (Ed.). Springer, New York, NY, USA, 1–34 (2013).

81 Vivier E, Raulet DH, Moretta A *et al.* Innate or adaptive immunity? The example of natural killer cells. *Science* 331(6013), 44–49 (2011).

82 Wood SM, Ljunggren H-G, Bryceson YT. Insights into NK cell biology from human genetics and disease associations. *Cell. Mol. Life Sci.* 68(21), 3479–3493 (2011).

83 Chan CJ, Andrews DM, Smyth MJ. Receptors that interact with nectin and nectin-like proteins in the immunosurveillance and immunotherapy of cancer. *Curr. Opin. Immunol.* 24(2), 246–251 (2012).

84 Yip GW, Smollich M, Götte M. Therapeutic value of glycosaminoglycans in cancer. *Mol. Cancer Ther.* 5(9), 2139–2148 (2006).

85 Di Domizio J, Zhang R, Stagg LJ *et al.* Binding with nucleic acids or glycosaminoglycans converts soluble protein oligomers to amyloid. *J. Biol. Chem.* 287(1), 736–747 (2012).

86 Groth C, Fortini ME. Therapeutic approaches to modulating Notch signaling: current challenges and future prospects. *Semin. Cell Dev. Biol.* 23(4), 465–472 (2012).

87 Sansone P, Bromberg J. Targeting the interleukin-6/Jak/stat pathway in human malignancies. *J. Clin. Oncol.* 30(9), 1005–1014 (2012).

88 O'Shea JJ, Plenge R. JAK and STAT signaling molecules in immunoregulation and immune-mediated disease. *Immunity* 36(4), 542–550 (2012).

89 Brownlie RJ, Zamoyska R. T cell receptor signaling networks: branched, diversified and bounded. *Nat. Rev. Immunol.* 13(4), 257–269 (2013).

90 Milstein O, Hagin D, Lask A *et al.* CTLs respond with activation and granule secretion when serving as targets for T-cell recognition. *Blood* 117(3), 1042–1052 (2011).

91 Granata M, Panigada D, Galati E *et al.* To trim or not to trim: progression and control of DSB end resection. *Cell Cycle* 12(12), 1848–1860 (2013).

92 Micucci F, Capuano C, Marchetti E *et al.* PI5KI-dependent signals are critical regulators of the cytolytic secretory pathway. *Blood* 111(8), 4165–4172 (2008).

93 Lee J-S, Smith E, Shilatifard A. The language of histone crosstalk. *Cell* 142(5), 682–685 (2010).

94 Verdaasdonk JS, Bloom K. Centromeres: unique chromatin structures that drive chromosome segregation. *Nat. Rev. Mol. Cell Biol.* 12(5), 320–332 (2011).

95 Nandakumar J, Cech TR. Finding the end: recruitment of telomerase to telomeres. *Nat. Rev. Mol. Cell Biol.* 14(2), 69–82 (2013).

96 Mahen R, Venkitaraman AR. Pattern formation in centrosome assembly. *Curr. Opin. Cell Biol.* 24(1), 14–23 (2012).

97 Muñoz MJ, Pérez Santangelo MS, Paronetto MP *et al.* DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* 137(4), 708–720 (2009).

98 Wu X, Wang Y, Wang H *et al.* Quinacrine inhibits cell growth and induces apoptosis in human gastric cancer cell line SGC-7901. *Curr. Ther. Res.* 73(1–2), 52–64 (2012).

99 Nobuhara M, Saotome M, Watanabe T *et al.* Mitochondrial dysfunction caused by saturated fatty acid loading induces myocardial insulin-resistance in differentiated H9c2 myocytes: a novel *ex vivo* myocardial insulin-resistance model. *Exp. Cell Res.* 319(7), 955–966 (2013).

100 Balgi AD, Fonseca BD, Donohue E *et al.* Screen for chemical modulators of autophagy reveals novel therapeutic inhibitors of mTORC1 signaling. *PLoS ONE* 4(9), e7124 (2009).

101 Naidu MD, Agarwal R, Pena LA *et al.* Lucanthone and its derivative hycanthone inhibit apurinic endonuclease-1 (APE1) by direct protein binding. *PLoS ONE* 6(9), e23679 (2011).

102 Ding X, Zhu F, Li T, Zhou Q, Hou FF, Nie J. Numb protects renal proximal tubular cells from puromycin aminonucleoside-induced apoptosis through inhibiting Notch signaling pathway. *Int. J. Biol. Sci.* 7(3), 269–278 (2011).

103 Kim JY, Lee SG, Chung J-Y *et al.* Ellipticine induces apoptosis in human endometrial cancer cells: the potential involvement of reactive oxygen species and mitogen-activated protein kinases. *Toxicology* 289(2–3), 91–102 (2011).

104 Andrews WJ, Panova T, Normand C, Gadal O, Tikhonova IG, Panov KI. Old drug, new target: ellipticines selectively inhibit RNA polymerase I transcription. *J. Biol. Chem.* 288(7), 4567–5582 (2013).

105 Lang E, Jilani K, Zelenak C *et al.* Stimulation of suicidal erythrocyte death by benzethonium. *Cell. Physiol. Biochem.* 28(2), 347–354 (2011).

106 Sharma N, Thomas S, Golden EB *et al.* Inhibition of autophagy and induction of breast cancer cell death by mefloquine, an antimalarial agent. *Cancer Lett.* 326(2), 143–154 (2012).

107 Milatovic D, Jenkins JW, Hood JE, Yu Y, Rongzhu L, Aschner M. Mefloquine neurotoxicity is mediated by non-receptor tyrosine kinase. *Neurotoxicology* 32(5), 578–585 (2011).

108 Wang J, Liu L, Qiu H *et al.* Ursolic acid simultaneously targets multiple signaling pathways to suppress proliferation and induce apoptosis in colon cancer cells. *PLoS ONE* 8(5), e63872 (2013).

109 Shih W-L, Yu F-L, Chang C-D, Liao M-H, Wu H-Y, Lin P-Y. Suppression of AMF/PGI-mediated tumorigenic activities by ursolic acid in cultured hepatoma cells and in a mouse model. *Mol. Carcinog.* 52(10), 800–812 (2013).

110 Liu C, Gong K, Mao X, Li W. Tetrandrine induces apoptosis by activating reactive oxygen species and repressing Akt activity in human hepatocellular carcinoma. *Int. J. Cancer* 129(6), 1519–1531 (2011).

111 Minero VG, Khadjavi A, Costelli P, Baccino FM, Bonelli G. JNK activation is required for TNFα-induced apoptosis in human hepatocarcinoma cells. *Int. Immunopharmacol.* 17(1), 92–98 (2013).

112 Fischer-Posovszky P, Keuper M, Nagel S *et al.* Downregulation of FLIP by cycloheximide sensitizes human fat cells to CD95-induced apoptosis. *Exp. Cell Res.* 317(15), 2200–2209 (2011).

113 Cai J, Li J, Liu W, Han Y, Wang H. α2-adrenergic receptors in spiral ganglion neurons may mediate protective effects of brimonidine and yohimbine against glutamate and hydrogen peroxide toxicity. *Neuroscience* 228, 23–35 (2013).

114 Wang Y, Yu X, Wang F *et al.* Yohimbine promotes cardiac NE release and prevents LPS-induced cardiac dysfunction via blockade of presynaptic α2A-adrenergic receptor. *PLoS ONE* 8(5), e63622 (2013).

115 Resende RR, Adhikari A. Cholinergic receptor pathways involved in apoptosis, cell proliferation and neuronal differentiation. *Cell Commun. Signal.* 7, 20 (2009).

116 Ebrahimkhani MR, Oakley F, Murphy LB *et al.* Stimulating healthy tissue regeneration by targeting the 5-HT2B receptor in chronic liver disease. *Nat. Med.* 17(12), 1668–1673 (2011).

117 Castorina A, Giunta S, Agata VD. Protective effect of the dopamine D3 receptor agonist (7-OH-PIPAT) against apoptosis in malignant peripheral nerve sheath tumor (MPNST) cells. *Int. J. Oncol.* 37(4), 927–934 (2010).

118 Ajeawung NF, Maltais R, Jones C, Poirier D, Kamnasaran D. Viability screen on pediatric low grade glioma cell lines unveils a novel anti-cancer drug of the steroid biosynthesis inhibitor family. *Cancer Lett.* 330(1), 96–105 (2013).

119 Jehle J, Schweizer PA, Katus HA, Thomas D. Novel roles for hERG K(+) channels in cell proliferation and apoptosis. *Cell Death Dis.* 2(8), e193 (2011).

120 Blanco-Aparicio C, Carnero A. Pim kinases in cancer: diagnostic, prognostic and treatment opportunities. *Biochem. Pharmacol.* 85(5), 629–643 (2013).

121 Ferretti M, Fabbiano C, Di Bari M *et al.* M2 receptor activation inhibits cell cycle progression and survival in human glioblastoma cells. *J. Cell. Mol. Med.* 17(4), 552–566 (2013).

122 Poirier D. New cancer drugs targeting the biosynthesis of estrogens and androgens. *Drug Dev. Res.* 69(6), 304–318 (2008).

123 Li Z, Yang Y, Ming M, Liu B. Mitochondrial ROS generation for regulation of autophagic pathways in cancer. *Biochem. Biophys. Res. Commun.* 414(1), 5–8 (2011).

124 Smethurst DGJ, Dawes IW, Gourlay CW. Actin – a biosensor that determines cell fate in yeasts. *FEMS Yeast Res.* 14, 89–95 (2013).

125 Munnamalai V, Weaver CJ, Weisheit CE *et al.* Bidirectional interactions between NOX2-type NADPH oxidase and the F-actin cytoskeleton in neuronal growth cones. *J. Neurochem.* 130(4), 526–540 (2014).

126 Jaquet V, Scapozza L, Clark RA, Krause K-H, Lambeth JD. Small-molecule NOX inhibitors: ROS-generating NADPH oxidases as therapeutic targets. *Antioxid. Redox Signal.* 11(10), 2535–2552 (2009).

127 Wang Z, Wei X, Zhang Y *et al.* NADPH oxidase-derived ROS contributes to upregulation of TRPC6 expression in puromycin aminonucleoside-induced podocyte injury. *Cell. Physiol. Biochem.* (24), 619–626 (2009).

128 Uesono Y, Araki T, Toh-E A. Local anesthetics, antipsychotic phenothiazines, and cationic surfactants shut down intracellular reactions through membrane perturbation in yeast. *Biosci., Biotechnol., Biochem.* 72(11), 2884–2894 (2014).

129 Li G, Zhou T, Liu L *et al.* Ezrin dephosphorylation/downregulation contributes to ursolic acid-mediated cell death in human leukemia cells. *Blood Cancer J.* 3, e108 (2013).

130 Ogura H, Tsukumo Y, Sugimoto H, Igarashi M, Nagai K, Kataoka T. ERK and p38 MAP kinase are involved in downregulation of cell surface TNF receptor 1 induced by acetoxycycloheximide. *Int. Immunopharmacol.* 8(6), 922–926 (2008).

131 Shan J, Xuan Y, Zheng S, Dong Q, Zhang S. Ursolic acid inhibits proliferation and induces apoptosis of HT-29 colon cancer cells by inhibiting the EGFR/MAPK pathway. *J. Zhejiang Univ. Sci. B.* 10(9), 668–674 (2009).

132 Han W, Lo H-W. Landscape of EGFR signaling network in human cancers: biology and therapeutic response in relation to receptor subcellular locations. *Cancer Lett.* 318(2), 124–134 (2012).

133 Guo C, Stark GR. FER tyrosine kinase (FER) overexpression mediates resistance to quinacrine through EGF-dependent activation of NF- κ B. *Proc. Natl Acad. Sci. USA* 108(19), 7968–7973 (2011).

134 Lee H, Hu M, Reilly RM, Allen C. Apoptotic epidermal growth factor (EGF)-conjugated block copolymer micelles as a nanotechnology platform for targeted combination therapy. *Mol. Pharm.* 4(5), 769–781 (2007).

135 Wu Z, Wang G, Xu S *et al.* Effects of tetrandrine on glioma cell malignant phenotype via inhibition of ADAM17. *Tumor Biol.* 35(3), 2205–2210 (2014).

136 Zhang X, Tang N, Hadden TJ, Rishi AK. Akt, FoxO and regulation of apoptosis. *Biochim. Biophys. Acta.* 1813(11), 1978–1986 (2011).

137 Guo C, Gaspariana V, Zhuang Z *et al.* 9-Aminoacridine-based anticancer drugs target the PI3K/AKT/mTOR, NF-kappaB and p53 pathways. *Oncogene* 28(8), 1151–1161 (2009).

138 Yan K-H, Yao C-J, Hsiao C-H *et al.* Mefloquine exerts anticancer activity in prostate cancer cells via ROS-mediated modulation of Akt, ERK, JNK and AMPK signaling. *Oncol. Lett.* 5(5), 1541–1545 (2013).

139 Gao N, Cheng S, Budhraja A *et al.* Ursolic acid induces apoptosis in human leukaemia cells and exhibits anti-leukaemic activity in nude mice through the PKB pathway. *Br. J. Pharmacol.* 165(6), 1813–1826 (2012).

140 Livshits G, Kobielak A, Fuchs E. Governing epidermal homeostasis by coupling cell–cell adhesion to integrin and growth factor signaling, proliferation, and apoptosis. *Proc. Natl Acad. Sci. USA* 109(13), 4886–4891 (2012).

141 Solá S, Aranha MM, Rodrigues CMP. Driving apoptosis-relevant proteins toward neural differentiation. *Mol. Neurobiol.* 46(2), 316–331 (2012).

142 Zhivotovsky B, Orrenius S. Calcium and cell death mechanisms: a perspective from the cell death community. *Cell Calcium* 50(3), 211–221 (2011).