nature biotechnology

Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms

Nicole C Kleinstreuer[1], Jian Yang[2], Ellen L Berg[2], Thomas B Knudsen[1], Ann M Richard[1], Matthew T Martin[1], David M Reif[1], Richard S Judson[1], Mark Polokoff[2], David J Dix[1], Robert J Kavlock[1] & Keith A Houck[1]

Addressing the safety aspects of drugs and environmental chemicals has historically been undertaken through animal testing. However, the quantity of chemicals in need of assessment and the challenges of species extrapolation require the development of alternative approaches. Our approach, the US Environmental Protection Agency's ToxCast program, utilizes a large suite of *in vitro* and model organism assays to interrogate important chemical libraries and computationally analyze bioactivity profiles. Here we evaluated one component of the ToxCast program, the use of primary human cell systems, by screening for chemicals that disrupt physiologically important pathways. Chemical-response signatures for 87 endpoints covering molecular functions relevant to toxic and therapeutic pathways were generated in eight cell systems for 641 environmental chemicals and 135 reference pharmaceuticals and failed drugs. Computational clustering of the profiling data provided insights into the polypharmacology and potential off-target effects for many chemicals that have limited or no toxicity information. The endpoints measured can be closely linked to *in vivo* outcomes, such as the upregulation of tissue factor in endothelial cell systems by compounds linked to the risk of thrombosis *in vivo*. Our results demonstrate that assaying complex biological pathways in primary human cells can identify potential chemical targets, toxicological liabilities and mechanisms useful for elucidating adverse outcome pathways.

Vast resources are devoted to understanding the toxicity potential in humans of pharmaceutical and environmental chemicals. However, traditional toxicity testing has major limitations as evidenced by the thousands of environmental chemicals lacking toxicity data and the high failure rate of investigational drugs due to adverse drug reactions[1,2]. These limitations arise from the costs of animal testing, societal concern over animal use and difficulties in extrapolating findings from animals to humans[3]. Nonanimal approaches that rely on *in vitro* assays to assess adverse effects of compounds on cells or targets have become standard practice in drug discovery, but drawbacks include insufficient numbers of validated targets and questions about the relevance and interpretation of assays[4,5]. Although the catalog of toxicity mechanisms can be expanded by computational models, and by methods to predict binding affinities[6,7] and structure-activity relationships[8], what is lacking is a way to directly link computational (*in silico*) targets and cellular (*in vitro*) responses with pathways of toxicity and adverse outcomes for relevant *in vivo* endpoints.

Here we describe how we used a panel of eight, complex, cell-culture systems, consisting of one or more primary human cell types, to detect and distinguish chemicals that act through a broad range of mechanisms relevant to human toxicity and pathological pathways. These cultures, called BioMAP Systems, are low-passage cells, which retain the intrinsic signaling patterns that enable them to respond physiologically to pharmacological agents. We based our selection of the eight cell systems on previously demonstrated sensitivity to specific drug mechanisms and pesticidal adverse effects[9–12]. The systems included primary endothelial cells (EC), peripheral blood mononuclear cells (PBMC), bronchial epithelial cells (BEC), dermal fibroblasts, keratinocytes and coronary artery smooth muscle cells in either mono- or co-culture conditions (**Table 1**). Concurrent activation of multiple signaling networks in each organ- or tissue-specific system, shown as "stimuli" in **Table 1**, generated activated, *in vitro* models that were maximally sensitive to perturbation by chemical exposures. A broad array of 87 endpoints including cell adhesion proteins, cytokines, matrix metalloproteases and cell surface receptors were measured by enzyme-linked immunosorbent assay (ELISA) to detect compound-induced changes in expression levels. These *in vitro* endpoints showed significant associations in predictive models of various toxicity phenotypes[13–15]. We challenged the cell culture systems with 776 diverse, unique, environmental and industrial chemicals (http://www.epa.gov/ncct/dsstox/sdf_toxcst.html)[16,17]; the toxicity information available for each ranged from none at all to animal guideline studies to effects measured in humans. Six pharmaceutical company partners contributed 135 failed drugs with accompanying preclinical and *in vivo* safety assessments, providing a unique and valuable asset toward developing alternative testing approaches. This work was conducted as a major component of the US Environmental Protection Agency's (EPA) ToxCast Program—a part of Tox21, the federal consortium that includes the US Food and Drug Administration and the National Institutes of Health—which seeks to develop more efficient approaches to predicting how chemicals may affect human health[18].

[1]National Center for Computational Toxicology, Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, North Carolina, USA. [2]BioSeek, a division of DiscoveRx, Inc., S. San Francisco, California, USA. Correspondence should be addressed to K.A.H. (houck.keith@epa.gov).

Received 3 June 2013; accepted 23 April 2014; published online 18 May 2014; doi:10.1038/nbt.2914

Table 1 Panel of 8 BioMAP systems used in this study

BioMAP system		3C	4H	LPS	SAg	BE3C	CASM3C	HDF3CGF	KF3CT
Primary human cell types		Venular endothelial cells	Venular endothelial cells	Peripheral blood mononuclear cells + endothelial cells	Peripheral blood mononuclear cells + endothelial cells	Bronchial epithelial cells	Coronary artery smooth muscle cells	Fibroblasts	Keratinocytes + fibroblasts
Stimuli		IL-1β + TNF-α + IFN-γ	IL-4 + histamine	TLR4	TCR	IL-1β + TNF-α + IFN-γ	IL-1β + TNF-α + IFN-γ	IL-1β + TNF-α + IFN-γ + EGF + bFGF + PDGF-BB	IL-1β + TNF-α + IFN-γ + TGF-β
Number of endpoints		13	7	11	10	11	14	12	9
Endpoint types	Acute inflammation	E-selectin, IL-8		E-selectin, IL-1α, IL-8, TNF-α, PGE2	IL-8	IL-1α	IL-8, IL-6, SAA	IL-8	IL-1α
	Chronic inflammation	VCAM-1, ICAM-1, MCP-1, MIG	VCAM-1, Eotaxin-3, MCP-1	VCAM-1, MCP-1	MCP-1, E-selectin, MIG	IP-10, MIG, HLA-DR	MCP-1, VCAM-1, MIG, HLA-DR	VCAM-1, IP-10, MIG	MCP-1, ICAM-1, IP-10
	Immune response	HLA-DR		CD40, M-CSF	CD-38, CD40, CD69, PBMC cytotox., T cell proliferation	HLA-DR	M-CSF	M-CSF	
	Tissue remodeling					uPAR, MMP-1, PAI-1, TGF-β1, SRB, tPA, uPA	uPAR	Collagen III, EGFR, MMP-1, PAI-1, fibroblast proliferation, SRB, TIMP-1	MMP-9, SRB, TIMP-2, uPA, TGF-β1
	Vascular biology	TM, TF, uPAR, EC proliferation, SRB, Vis	VEGFRII, uPAR, P-selectin, SRB	Tissue Factor, SRB	SRB		TM, TF, LDLR, SMC proliferation, SRB		
Disease/tissue relevance		Cardiovascular disease, chronic inflammation	Asthma, allergy, oncology, vascular biology	Cardiovascular disease, chronic inflammation	Autoimmune disease, chronic inflammation	COPD, respiratory, epithelial	Cardiovascular, inflammation, restenosis	Fibrosis, wound healing	Psoriasis, dermatitis, skin

BioMAP systems listed according to their short names comprise the cell types shown, cultured and activated with the indicated stimuli (added along with test compounds) for 24 h. For each system, the protein or mediator biomarker readouts listed (number of readouts is shown in parentheses) are measured by ELISA at 24 or 72 h. Biomarker endpoints measured were all cell-associated with the exception of TNFα and PGE2, which were measured in the supernatants. Pathway mechanisms detected in each system were assessed by testing highly selective, pathway-specific activators or inhibitors, as described[9].

Given the large size and complexity of the present data set, we provide an initial characterization of the data and assessment for relevant patterns. Owing to the large number of environmental compounds with unknown mechanisms or toxicities, an overall quantitative predictive assessment was not feasible; however, a number of known associations and previously unreported predictions emerged. Self-organizing maps (SOM) and hierarchical clustering were employed as unsupervised analyses. The data set was explored by supervised analyses using a reference database of compounds with known mechanisms, and classified using Support Vector Machine (SVM) models developed for 28 mechanism classes[9]. Examples substantiating the utility of this approach along with identified limitations are described below.

RESULTS

Overview of activities measured

We tested the 776-compound chemical library (800 coded samples, including blinded replicates) in concentration-response in the eight BioMAP systems with quantitative readouts for 7–14 protein biomarkers per system (**Table 1**), totaling 87 readouts per compound per concentration. In total, the complete data set contains 306,240 measurements (**Fig. 1a** and **Supplementary Table 1**; also available at http://epa.gov/ncct/toxcast/data.html), where 700 compounds were run at four test concentrations in duplicate, and a subset was run at four additional concentrations (**Fig. 1b**). Analysis by chemical use groups shows that pesticides and pharmaceuticals were the most active chemicals, whereas fragrances and colorants were least active. Overall, 22% of chemicals were overtly cytotoxic to one or more cell types at the top concentration tested, although the percentage of compounds ranged from 0% to 30%, depending on use group (**Supplementary Table 2**). Fibroblasts, endothelial cells and PBMCs were most frequently affected, although there was substantial diversity among the chemical groups.

The assays demonstrated excellent reproducibility across technical and biological replicates (**Supplementary Table 3**). Colchicine, included on every plate as a positive control, had Pearson correlation coefficients from 0.82–0.97, and the concordance for the 700 compounds run in duplicate was >95% across all assay endpoints. Principal component analysis of built-in test replicates (seven chemicals each, present as three independent samples and two chemicals each, present as six independent samples) showed tight clustering of each replicate set based on AC_{50} values across all assays (**Supplementary Fig. 1**).

Each BioMAP system (**Table 1**) contains one or more assay endpoints measuring general cytotoxicity that may indicate potential for toxicity *in vivo,* as well as confound interpretation of individual biomarker endpoints *in vitro*. For example, loss of specificity of an otherwise mechanism-based downregulation of a biomarker may be secondary to loss of cell function at concentrations approaching cytotoxicity. Thus, we separately evaluated compounds that were cytotoxic across cell types. Ten compounds were highly cytotoxic

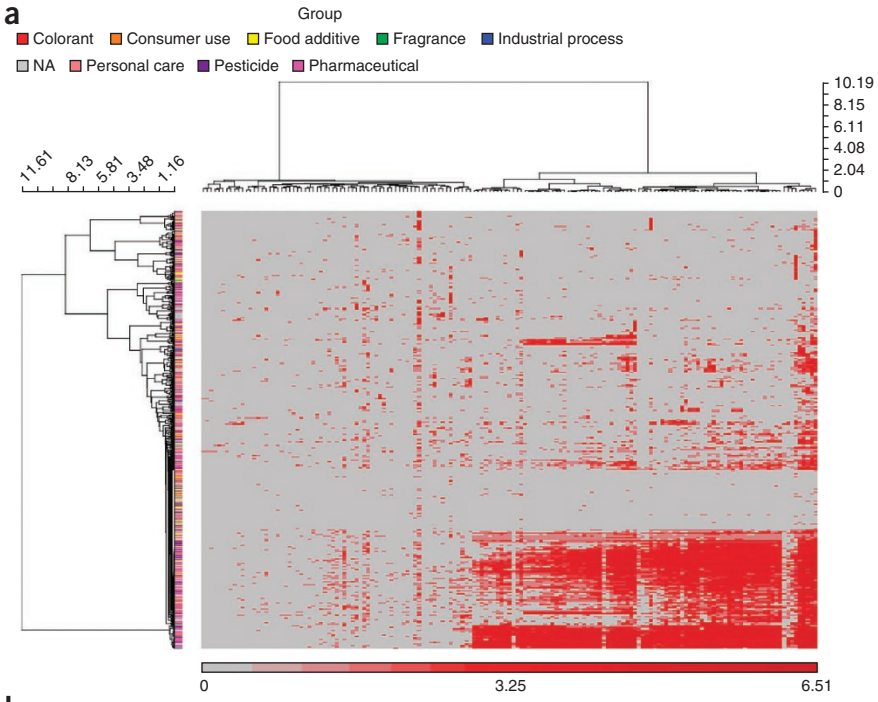

Figure 1 Overview of responses for all chemicals and endpoints. (**a**) Hierarchical clustering of ToxCast Phase II chemicals (rows) by activity in human primary cell systems (columns). Where information was available, chemicals are color-coded by use class (colorant, consumer use, food additive, fragrance, industrial process, personal care, pesticide, pharmaceutical or NA for not available). Activity (by log-transformed lowest effective concentrations) in 87 endpoints in two directions (down- and upregulation) across eight cell systems for 776 compounds. Clustering was by Pearson's dissimilarity and Ward's method. The left clade represents the majority of upregulated endpoints, whereas the right clade shows those that were downregulated. (**b**) Summary of activities by chemical group. Endpoint measurements that are outside a 95% confidence interval for vehicle controls were deemed active. Chemicals for which the number of active endpoints at the highest concentration is >5 were deemed active. Chemicals were deemed cytotoxic if the chemical at the highest concentration was cytotoxic to one or more cell types.

b

Chemical class	Number of samples tested	% Active	% Cytotoxic	% Active and noncytotoxic	% Inactive
ToxCast phase I	321	94	22	72	6
ToxCast phase II	800	73	22	51	27
Colorant	19	32	0	32	68
Consumer use	64	67	28	39	33
Food additive	92	60	9	51	40
Fragrance	3	33	0	33	67
Industrial process	35	46	9	37	54
NA	113	80	22	58	20
Personal care	136	68	18	49	32
Pesticide	88	82	30	52	18
Pharmaceutical	250	83	29	54	17

(>50% reduction in total protein levels) in the majority of cell systems at concentrations <15 μM (**Supplementary Table 4**), including several organometallics and three pharma compounds that failed preclinically: DPharma52, DPharma121 and DPharma123 (full names in **Supplementary Table 5**).

A large number of compounds exhibited cell type–specific responses, both in terms of cytotoxicity and functional activity; for example, a number of estrogenic compounds were selectively cytotoxic to endothelial cells in the 3C system (consisting of venular endothelial cells stimulated with three cytokines: interleukin (IL)-1β, tumor necrosis factor (TNF)α and interferon (INF)-γ) including estrogens, phytoestrogens and estrogenic pesticides (**Supplementary Fig. 2**). Other compounds with this pattern included flame retardants (tris(2-ethylhexyl) phosphate and triphenyl phosphate) and pharmaceutical compounds DPharma13 and DPharma 77. Another group of compounds including aromatic organic phenols and organic peroxides exhibited specific cytotoxicity in the HDF3CGF system, which consists of fibroblasts stimulated with six cytokines or growth factors. Ten of these 11 chemicals were activators of NRF-2, a transcription factor responsive to oxidative stress, in a reporter gene assay run in the ToxCast program (unpublished data). The chemical that was inactive in the NRF-2 assay was a known oxidant[19], suggesting the HDF3CGF system may be sensitive to oxidative stress-induced cytotoxicity. The surfactants perfluorooctanesulfonate (PFOS) and perfluorononanoic acid, and a pharma compound identified as a glycine transporter inhibitor showed cytotoxicity unique to BECs (BE3C). Notably, PFOS is known to affect lung development prenatally[20], and studies have shown preferential distribution of the chemical to the lung in pharmacokinetic studies[21]. It is intriguing that this cell model may possibly recapitulate some of the chemical's complex *in vivo* toxicity. Keratinocytes were preferentially affected by a cluster of 28 compounds including nicotine, shown to have toxicity mediated through nicotinic acetylcholine receptors expressed by keratinocytes[22]. Several clusters of compounds exhibited cell type–specific upregulation of inflammatory endpoints (**Supplementary Fig. 3**), such as 25 compounds affecting vascular endothelial cells in the 4H system (venular endothelial cells stimulated with IL-4 and histamine) including known angiogenic modulators 5HPP-33 (thalidomide analog[23]), lovastatin and simvastatin[24].

Among the 776 compounds tested, 63 compounds (8%) were inactive (no lowest effective concentration) across all the endpoints, including a number of pharmaceutical compounds (two donated failed drugs, DPharma8 and DPharma103, and marketed drugs erythromycin, tolazamide and 3-azido-3-deoxythymidine) with expected bioactivity that can be considered false negatives or indicate lack of appropriate bioassay systems.

Chemical clustering

To characterize the range of biological activities and patterns detected, we first analyzed the data set by unsupervised clustering of all endpoints for each compound/concentration pair. Responses at individual concentrations were used to discriminate potential effects of polypharmacology that are often seen at increased concentrations. The data were normalized by row (chemical, **Supplementary Table 6**) to reduce concentration effects and by column (assay, **Supplementary Table 7**) to highlight responses driven by specific proteins. Unsupervised

Table 2 Examples of clusters that emerged from the self-organizing map analysis

Norm. method [cluster(s)]	Cluster count	Common activity	Example compounds: known associations	Example compounds: novel associations
Chemical [1]	78	Analgesics	Aspirin Indomethacin Celecoxib Diclofenec Darbufelone Clove leaf oil Eugenol Isoeugenol	Propyl gallate Fluridone
Chemical [65]	31	Steroid hormone receptor modulators	Cyproterone acetate Norgestrel Progesterone 17-hydroxyprogesterone Mifepristone	Mirex Donated pharma: PPAR pan agonist A3 adenosine receptor antagonist
Chemical [57, 67]	52	AHR ligands	Hydroquinone 4-chloro-1,2-diaminobenzene 1,2-phenylenediamine Fenaminosulf	Color Index. Solvent Yellow 14
Chemical [48]	27	Estrogen receptor pathway modulators	Clomiphene citrate Tamoxifen citrate Fulvestrant Raloxifene hydrochloride Tamoxifen 4-hydroxytamoxifen	Cyclopamine Amiodarone hydrochloride Haloperidol Reserpine Donated pharma: NK1 receptor antagonist Bradykinin B1 receptor antagonist Lipid-lowering agent
Assay [46]	29	TNFα inhibition	All-trans retinoic acid Donated pharma: PDE inhibitors (8 compounds)	Terbuthylazine Donated pharma: $GABA_A1$ receptor antagonist
Assay [39]	31	SAA upregulation	Prednisone Dexamethasone Corticosterone Triamcinolone	Coumarin 4-octylphenol Cyclohexanol Pentaerythritol
Assay [90,100]	58	Potent cytotoxicants	Tributyltin methacrylate Tributyltin chloride Gentian violet Didecyldimethylammonium chloride Triclosan Phenylmercuric acetate	Octyl gallate 4-Nonylphenol 9-Phenanthrol Donated pharma: Factor Xa inhibitor CCK1R agonist Mast cell tryptase inhibitor

Examples shown here based on common mechanisms or observed *in vitro* activity. The normalization method and corresponding cluster number was by chemical (**Supplementary Table 5**) or by assay (**Supplementary Table 6**). The cluster count refers to the number of chemical/concentration pairs that appear in that cluster. Note: cluster numbers differing by 10, for example, 57 and 67, are adjacent in the 10 × 10 self-organizing map array and thus expected to be relatively closely related.

clustering was done using a self-organizing map (detailed in Online Methods)[25] approach in 10 × 10 arrays (**Supplementary Fig. 4**).

We found clusters containing chemicals with similar activity, whose mechanisms might be predictable, along with those that were not obvious (**Table 2**). Illustrative examples of clusters with known pharmacologic mechanisms include cluster one, containing pharmaceutical and natural analgesics and anesthetics as well as compounds not known to have analgesic activity (the antioxidant propyl gallate and the herbicide fluridone). Each of these chemicals was present at multiple concentrations in this cluster, suggesting polypharmacology was not a major confounding factor over the concentration range tested. Cluster 57 contained 44 chemical-concentration pairs consisting of 20 unique chemical structures of which several, for example, benz(a)anthracene, were polycyclic aromatic hydrocarbons, a class of chemicals known to activate the aryl hydrocarbon receptor (AHR). Data from a multiplexed, transcription-factor reporter gene assay, run on the same chemical library, showed that of the 20 unique structures in this cluster, 18 were positive for activation of AHR (unpublished data). These results demonstrate the utility of a cell systems approach where, even without explicitly including targets such as AHR, chemical perturbation of common critical pathways can be sensed. Cluster 46 was driven by selective inhibition of TNFα production in the lipopolysaccharide system and contained 11 donated pharma compounds, the herbicide terbuthylazine and all-trans retinoic acid (ATRA). Of the drugs whose targets were known, eight were identified as phosphodiesterase inhibitors (PDE3, -4 and/or -5). Inhibition of PDE4 is associated with cAMP elevation and was previously shown to suppress immune and inflammatory responses, including TNFα production[26]. Neighboring chemical clusters showed similar, but less potent, downregulation of TNFα. Cluster 39 was driven by specific upregulation of serum amyloid A (SAA) in smooth muscle cells, and included almost all test concentrations of the glucocorticoids prednisone, dexamethasone, corticosterone and triamcinolone. SAA is an acute-phase inflammatory response protein shown to be preferentially upregulated by glucocorticoids in muscle tissue[27].

Cluster 48 contained well-known, selective estrogen receptor modulators and antagonists. Other compounds found in this cluster, such as haloperidol and reserpine, were not specifically associated with activation of the estrogen receptor pathway. Another compound, the teratogen cyclopamine, acts through the sonic hedgehog pathway to downregulate estrogen receptor–α protein, and DPharma34, also known as SB-236057, has teratogenic effects through pathways similar to those of cyclopamine[28]. Inspection of the compound profiles in cluster 48 and comparison with estrogen receptor agonists (cluster 28) revealed tissue factor (TF) in the 3C system as a discriminating activity of particular interest (**Fig. 2**). TF levels in the 3C system were

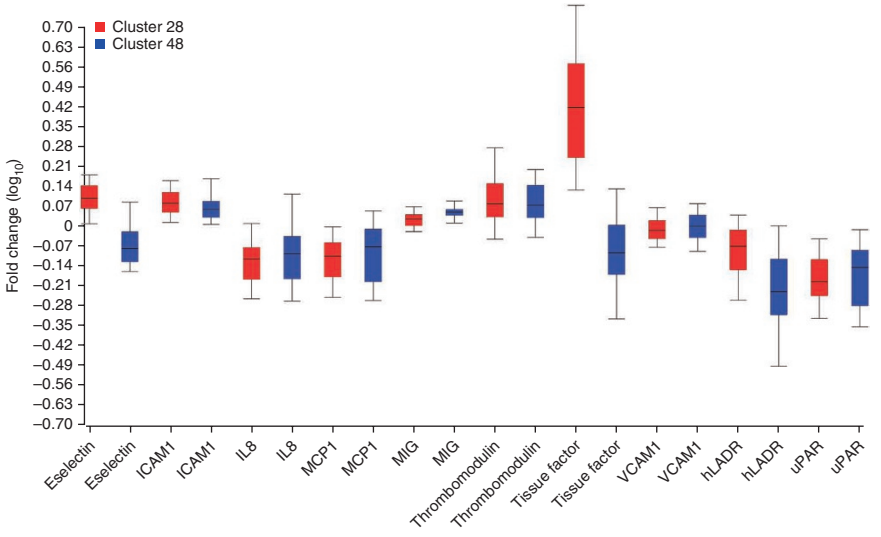

Figure 2 Comparison of endpoints from the 3C system for cluster 28 containing estrogen receptor (ER) agonists (blue) and cluster 48 containing estrogen receptor antagonists/selective estrogen receptor modulators (red). Boxes represent 95% of the values; the central line is the median; and the whiskers, the upper and lower 5%. The intraclass correlation coefficients, that is, the fraction of the variance of each variable that is explained by the difference between these clusters, are 0.79, 0.69, 0.35 and 0.28 for TF, E-selectin, hLADR and MIG, respectively (cluster 28: $n = 26$, cluster 48: $n = 27$).

preferentially increased by estrogen receptor antagonists, and decreased by estrogen receptor agonists. TF (coagulation factor III or thromboplastin) is an initiator of the extrinsic blood coagulation cascade and is a clinical risk factor for thrombosis[29]. In a previous study, among compounds from 28 mechanism classes, only compounds from two mechanism classes, mTOR inhibitors and AHR agonists were found to increase the level of TF in the 3C system[9]. Rapamycin, an mTOR inhibitor, has been shown to increase TF and promote arterial thrombosis *in vivo*[30], and cigarette smoke, associated with risk of cardiovascular disease and acute coronary thrombosis[31], causes exposure to polycyclic aromatic hydrocarbons with expected activation of AHR. In addition, thrombosis is associated with clinical use of the selective estrogen receptor modulators, tamoxifen (Nolvadex) and raloxifene (Evista)[32,33].

Pharmaceutical correlation analysis

To determine if we could classify compounds designed to be active in humans by mechanism of action, we first analyzed the BioMAP profiles of the 135 donated pharmaceuticals by pairwise correlation. We visualized these relationships in a function-homology map (**Fig. 3** and Online Methods)[34,35]. Compound pairs (Pearson's correlation coefficient (r) > 0.75) are shown as connected lines demonstrating that the pharmaceutical compounds could be sorted into groups (clusters in **Fig. 3**) by BioMAP profiling (Online Methods). In each of these clusters, there were multiple examples of drugs whose intended and classified mechanisms matched, and drugs whose classified mechanism may represent an unidentified off-target effect. Compounds were further analyzed by comparison to BioMAP profiles previously generated from reference chemicals with known mechanisms of action.

Representative BioMAP profile matches with $r > 0.75$ (**Fig. 3**) include examples of confirmatory classifications and previously unreported associations. We highlight five here. (i) Compound DPharma27 was identified as a CysLT1 antagonist based on the similarity of its profile to that of the reference compound montelukast ($r = 0.85$). This compound was developed as an LTB4 receptor antagonist, and failed clinical development owing to aneuploidy, possibly reflected by its BioMAP cytotoxicity in multiple cell types. (ii) Compound DPharma68 (40 μM) was identified as a PDE4 inhibitor based on profile similarity to ibudilast ($r = 0.82$), and was indeed developed as a PDE4 inhibitor, was modestly antiproliferative to endothelial cells, PBMCs, coronary artery smooth muscle cells and dermal fibroblasts, and showed strong inhibition of TNFα in the lipopolysaccharide system. Clinical failure involved emesis and linked vagal response, known adverse events associated with PDE4 inhibitors[36]. (iii) Compound DPharma2 was identified as an mTOR inhibitor based on profile similarity to rapamycin (4.44 μM, $r = 0.75$) and to AZD8055 (13.3 μM, $r = 0.77$, profile not shown). This compound was developed as a benzothiophene cell activation inhibitor, and was antiproliferative to endothelial cells, PBMCs, coronary artery smooth muscle cells and dermal fibroblasts, and inhibited uPAR and HLA-DR in endothelial cells. (iv) Compound DPharma86 (13.3 μM) was identified as a p38 MAPK inhibitor based on profile similarity to VX-745 ($r = 0.96$). This compound was developed as a MAP kinase inhibitor and inhibited TF and HLA-DR, strongly inhibited monocyte activation, and upregulated adhesion molecule VCAM-1 and chemokine IP-10 in dermal fibroblasts. Clinical failure was associated with central nervous system adverse events and acneiform rash and preclinical lethal effects in rabbit reproductive toxicity testing. (v) Two additional compounds (not shown), DPharma75 and DPharma101, developed as PDE4 inhibitors for treatment of asthma, matched the profile for ATRA at multiple concentrations with Pearson's correlations of 0.71–0.82. Rodent studies on these two compounds showed fetal effects, as would be expected from compounds acting like the developmental toxicant ATRA (http://actor.epa.gov/toxrefdb/faces/Home.jsp). The complete BioMAP profile similarity search results ($r > 0.7$) for the subset of donated pharmaceuticals is included as **Supplementary Table 8**.

Mechanism predictions

We next applied a support vector machine (SVM) learning algorithm to the BioMAP data to classify chemicals by likelihood of belonging to 28 predefined models of mechanistic classes trained on a reference compound database. The details of these models, their performance and example applications were described previously[9]. Mechanism class decision values (DV), a measure of in-class confidence, were calculated for each test concentration of every compound, again to separate polypharmacological effects. All compounds with high DV ($DV_{max} > 0.4$) for the respective mechanisms classes were systematically examined. We found 71 predictions consistent with known mechanisms and 65 potentially novel predictions (**Supplementary Table 9**). The complete set of SVM predictions is in **Supplementary Table 10**. To visualize the overall distribution of predicted compound activities, we clustered the SVM DVs using a 30 × 30 self-organizing map. **Figure 4** shows the distribution of mechanism class DVs for all compounds as a trellis self-organizing map plot. The in-class predictions for each mechanism are shown as separate plots illustrating both unique clusters of compounds corresponding to specific mechanism classes (e.g., glucocorticoid receptor (GR) agonists, H1 antagonists) as well as

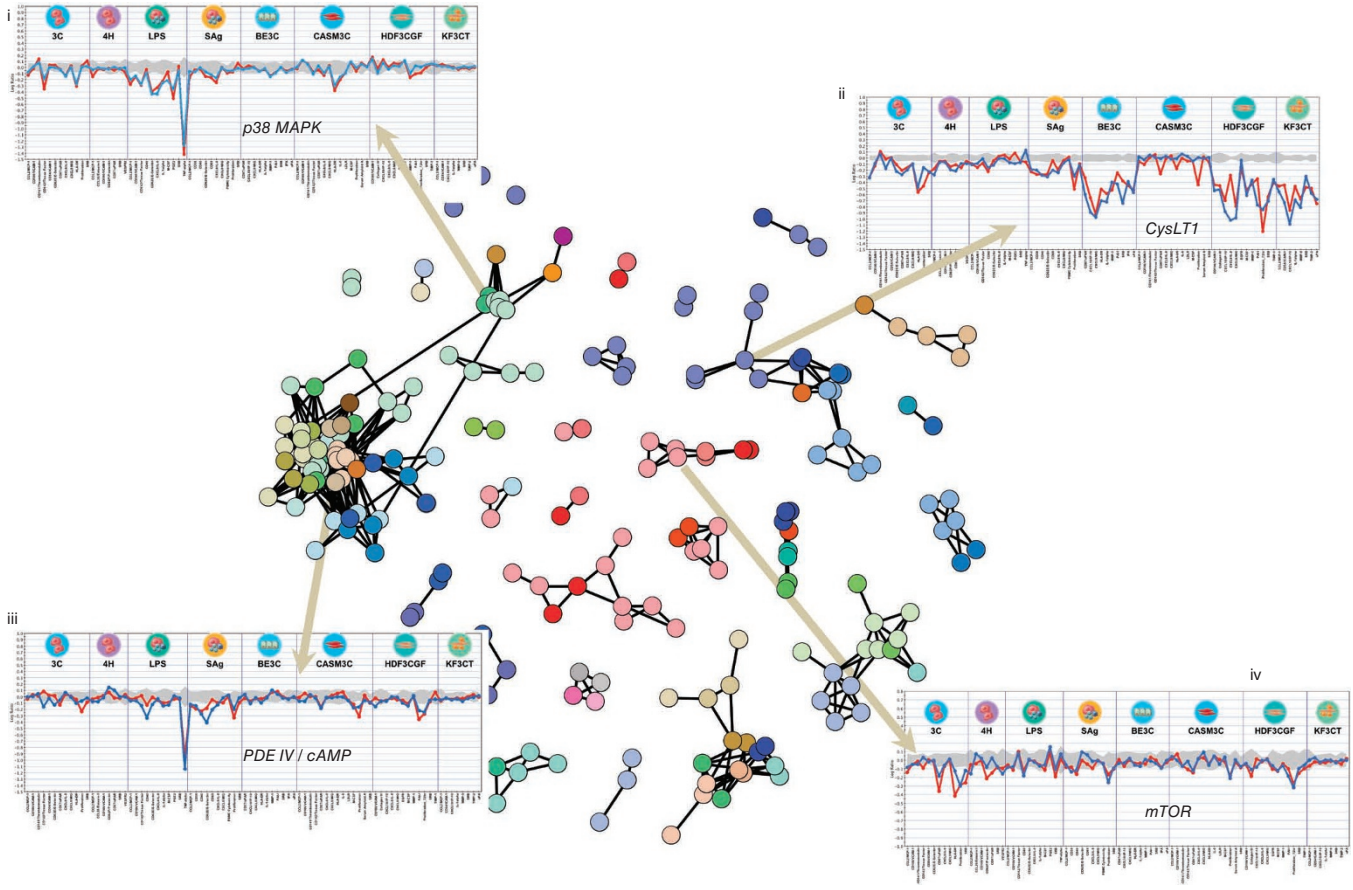

Figure 3 Function similarity map for 135 failed pharmaceutical compounds. Compound profiles in eight BioMAP systems were compared by pairwise correlation and subjected to nonlinear projection. Each circle represents a compound profile at a single concentration, colors (randomly assigned) represent different compounds with shading to indicate compound concentration (darkest shading indicating highest concentration). Lines are drawn between compound-concentration pairs with $r > 0.75$. Line graphs highlight examples of BioMAP profile similarities between donated pharmaceuticals (red) and reference compounds (blue). (i) DPharma86 (13.3 μM) is similar to the p38 MAPK inhibitor, VX-745 (3.33 μM), with a Pearson's correlation coefficient of 0.963. (ii) DPharma27 (40 μM) is similar to the CysLT1 antagonist, montelukast (10 μM), with a Pearson's correlation coefficient of 0.853. (iii) DPharma68 (40 μM) is similar to the PDE4 inhibitor, ibudilast (90 μM), with a Pearson's correlation coefficient of 0.821. (iv) DPharma2 (4.44 μM) is similar to the mTOR inhibitor, rapamycin (0.11 μM), with a Pearson's correlation coefficient of 0.75.

compound clusters that have high DVs for multiple mechanism classes (e.g., PDE4 inhibitors, retinoic acid receptor/retinoid X receptor (RAR/RXR) agonists, prostaglandin E receptor (EP) agonists), possibly reflecting pathway crosstalk.

The group of compounds predicted to be GR agonists included the corticosteroids dexamethasone, corticosterone, triamcinolone and prednisone. Other compounds in the GR agonist cluster included coumarin, previously shown to bind to steroid hormone receptors[37], and pentaerythritol, for which a mammalian enzyme, pentaerythritol tetranitrate reductase, has shown preferential binding to steroid substrates[38]. Five compounds had high DVs at multiple test concentrations unique to the histamine receptor H1 antagonist mechanism class with two of these, trelanserin and volinanserin, known as selective serotonin 5-HT2A antagonists; volinanserin and other serotonin modulators have been shown to control histamine release in mice[39]. Besonprodil is a *N*-methyl-D-aspartate (NMDA) receptor antagonist under development as a supplemental medication for Parkinson's disease[40], whose interaction with the histaminergic system has not been shown; however, another NMDA receptor antagonist, ketamine, has shown acute suppression of histamine release in rat limbic brain regions[41]. There were 11 compounds specifically predicted at multiple test concentrations to be AHR agonists, and this cluster showed a high degree of overlap with the aforementioned cluster of polycyclic aromatic hydrocarbons and AHR ligands that emerged during the unsupervised analysis.

Some compounds with specific known mechanisms also gave high DVs for additional classes, suggesting similarities or functional relationships. One group included p38 MAPK inhibitors, PDE4 inhibitors, IKK2 inhibitors, EP agonists and RAR/RXR agonists. There was concordance between predicted PI3K inhibitors, mTOR inhibitors, microtubule stabilizers and disruptors, and mitochondrial affecters, as well as predicted proteasome modulators, histone deacetylase inhibitors (HDAC) inhibitors and Hsp90 inhibitors. However, there were still small numbers of compounds that had high DVs unique to each mechanism. This was sometimes defined by specific concentrations. For example, colchicine had distinct DVs at multiple test concentrations distinguishing between microtubule disruption and microtubule stabilization. The reference compounds for HMG-CoA reductase inhibition, lovastatin and simvastatin, had high DVs for the corresponding pathway, as did the only other statin included in the library, pravastatin. The remaining compound clusters with positive DVs for this mechanism class overlapped substantially with clusters for microtubule stabilization and mTOR inhibition. Mechanism classes with consistently low DVs across the data set were JAK inhibitors, EGFR

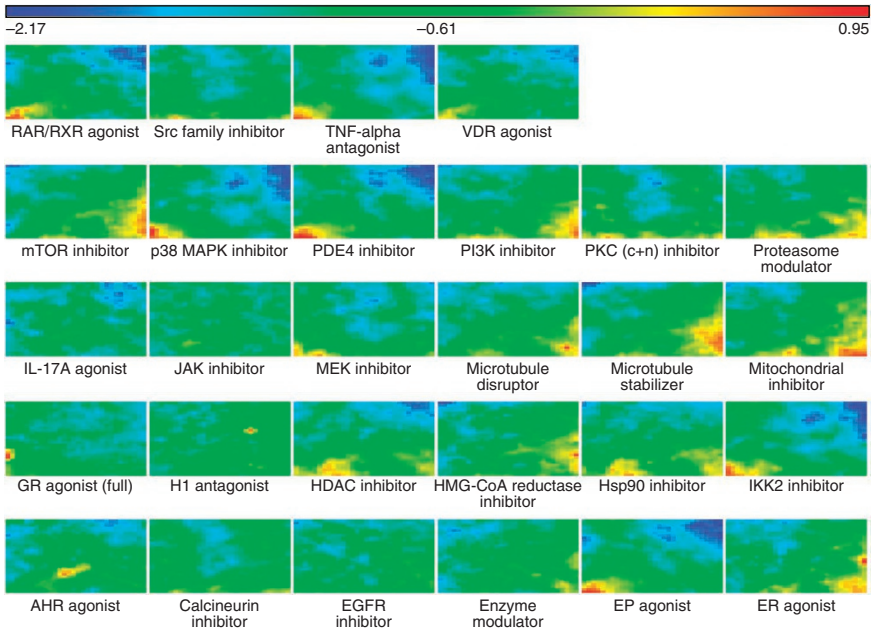

Figure 4 Distribution of mechanism class decision values (DV) for all test concentrations of all compounds shown as self-organizing maps in a trellis plot conditioned by mechanism class DV. Each plot shows a 30 × 30 grid of boxes, where each box represents a heterogeneously sized cluster of chemicals and/or concentrations. Color scale represents SVM DVs from blue (negative DV) to red (high DV), computed based on the weighted average of Pearson correlation coefficients to identified reference classification profiles (n = 3,600 chemical/concentration pairs). The relative probabilities for belonging to each mechanism class are shown as separate plots, where the clustering of compounds is consistent across all plots. Red areas represent compounds strongly predicted to belong to that class, and blue areas represent compounds strongly predicted not to belong to that class. Unique clusters of compounds corresponding to specific mechanism classes may be observed (e.g., GR agonists, H1 antagonists), as well as compound clusters that have high DVs for multiple mechanism classes (e.g., PDE4 inhibitors, RAR/RXR agonists, EP agonists).

inhibitors and calcineurin inhibitors, suggesting a lack of coverage of these targets within this chemical library.

There were 33 compounds with known mechanisms included in the 28 classes predicted by the SVM model (**Supplementary Table 11**). We reported the maximum DV and associated mechanism class for each chemical at any tested concentration, as well as the top DV for the intended mechanism class, in cases where it differed. Of these compounds, 78% (26) of 33 chemicals belonging to nine classes were correctly predicted (accuracy ranged from 50–100% depending on the class). The reference compound set was biased toward endocrine active compounds known to target the estrogen receptor (ER), and six of ten estrogen receptor agonists were correctly predicted by the SVM model. Notably, many estrogen receptor antagonists had nonzero DVs for the estrogen receptor agonist mechanism class, but their maximum DV corresponded to another class. Estrogen receptor reference compounds predicted to affect microtubule stabilization were 4-hydroxytamoxifen, 4-nonylphenol, clomiphene citrate, raloxifene hydrochloride, and tamoxifen, whereas, diethylstilbestrol and meso-hexestrol were predicted to affect mitochondria. These are both known potential side effects of estrogen receptor pathway modulation at superpharmacological concentrations[42,43].

DISCUSSION

This study demonstrates the promise of using panels of primary human cells in profiling bioactivity for a large and diverse set of chemicals with potential human exposure. In contrast to approaches focused on understanding the actions of a single chemical agent at the molecular and mechanistic level, this method harvests the collective knowledge embedded in reference chemicals with respect to their molecular targets, mechanism of action, and animal and human toxicity and applies it to characterizing the biological activity of large numbers of newly tested chemicals. Even using a limited set of primary human cell systems, we have shown the capability to recognize consistent patterns of bioactivity correlated with diverse drug actions and toxicities. Several statistical clustering methods consistently grouped known and unknown chemicals by similar bioactivity profiles. New chemicals falling into clusters with known activities suggest specific potential mechanisms of toxicity to be more carefully evaluated, greatly increasing the efficiency of toxicity testing by focusing resources for follow-up testing on bioactivities of highest concern. Nevertheless, the sheer volume of results from this study, together with the lack of comprehensive, publically accessible databases of the biological activity of these compounds, prevented us from exhaustively analyzing the results and determining the sensitivity and specificity of the predictions derived from the data.

Several findings illustrate the potential in this data set for elucidating mechanisms of action or toxicities recognized by profile similarities among drugs with common targets and similar or distinct clinical adverse events. For example, the donated pharmaceutical compound DPharma86 (correctly predicted to be a MAP kinase inhibitor) increased the inflammatory molecules VCAM-1 and IP-10 in dermal fibroblasts. These pro-inflammatory activities have been observed in several clinical compounds for which exposure has been associated with skin rash (MEK and p38 MAPK inhibitors)[44,45], and indeed this was an observed adverse outcome of this particular compound in clinical trials. Another intriguing finding is the discovery of a common feature between estrogen receptor pathway modulators, AHR activators and mTOR inhibitors, that is, increased TF in the 3C system, and its association with risk of thrombosis. Both of these examples illustrate the potential utility of this approach for elucidating toxicity mechanisms, defining adverse outcome pathways and building predictive models.

One notable advantage of this approach is the relatively close relationship of the endpoints measured (e.g., TF in EC-containing models, VCAM-1 in the dermal fibroblast model) to human adverse outcomes (thrombosis and skin rash, respectively), providing a means to gauge the plausibility of the findings and suggest possible biomarkers of exposure. The present study highlights only some of the most interesting findings (a number of other similar findings and examples are included in the **Supplementary Discussion**).

There are many challenges and limitations in using bioactivity profiles to predict mechanisms and adverse outcomes. One major issue is the lack of clinical information; what is available is limited, difficult to access and not in a computable form. Annotated public databases of clinical effects using standardized vocabulary combined with exposure and compound metabolism information would allow

development of human relevant models and likely widen the ability to uncover novel findings. In addition, it is necessary to perform the analysis at multiple testing concentrations as this is critical in handling the promiscuity and polypharmacology of chemicals including human drugs[7]. Moreover, some compounds that were expected to be biologically active, such as pharmaceuticals, did not affect any of the endpoints measured here, and other compounds with known mechanisms of action were not assigned to appropriate classes. For example, valproic acid is an HDAC inhibitor, but it did not have positive DVs for the corresponding mechanism classes or, in fact, for any of the classes examined here. Perhaps it was not tested at appropriate concentrations, as very high concentrations (mM) of valproic acid are required for HDAC inhibition relative to the 40 μM top concentration tested here. In other cases, we may have been unable to classify a compound by its mechanism of action because the current testing panel did not have appropriate cell systems. Although many of the activated signaling pathways used in these systems are conserved and used to varying degrees in most cell types, there are others that are specific to particular cells and tissues. Examples of missing and potentially useful systems include renal, adrenal, neural, intestinal, mammary and liver cells. Some of these systems might also supply a great deal of xenobiotic metabolism capacity that current systems may lack. Such biotransformation potential is likely to be crucial in building a complete predictive *in vitro* testing strategy.

This BioMAP primary human cell system platform is only one component of the Environmental Protection Agency's ToxCast program. The goal of ToxCast is to develop predictive toxicity models based on data from alternative testing methods. Other assay platforms in the program include biochemical assays targeting individual enzymes and receptors, cellular gene reporter assays for transcription factors and stress pathways, high-content imaging assays for cell health profiles, and assays targeting key cell signaling pathways and gene expression[17]. The use of primary human cells in the BioMap system is an important complement to the other assays that do not use primary cells. The same chemical library has been tested by all assay platforms and the results have been made publically available (http://actor.epa.gov). The large number of compounds tested and the rich detail about the molecular targets and cellular responses across a range of exposures make this a unique effort in toxicology. The compound libraries are also used in the US governmental Tox21 collaboration that has a goal of transforming traditional toxicity testing[46]. The Tox21 chemical library totals over 8,100 unique compounds and is being tested on an ongoing basis with an initial focus on nuclear receptors and stress pathways. Although these approaches are currently at the research stage, data are beginning to be applied toward chemical prioritization in support of the Endocrine Disruptor Screening Program[47].

This study has demonstrated the ability to use alternative methods combined with existing knowledge to classify compounds for potential mechanisms and to broaden our understanding of relevant doses and polypharmacology. This type of complex, *in vitro* assay panel based on human primary cells in carefully designed biological environments combined with innovative *in silico* analysis may serve as a first-tier alternative to animal testing using a human-relevant system for chemical screening and prioritization with applications to toxicity testing and drug discovery.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank our pharmaceutical industry partners Pfizer, GlaxoSmithKline, Roche, Merck, Sanofi and Astellas for donating 135 failed drugs and providing feedback on the manuscript, and W. Casey, L. Urban, C. Wood and K. Crofton for helpful suggestions on the manuscript. The EPA, through its Office of Research and Development, funded and managed the research described here. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the US Environmental Protection Agency.

AUTHOR CONTRIBUTIONS

R.J.K., D.J.D. and K.A.H. conceived and supervised the ToxCast project. A.M.R. oversaw chemical management. D.M.R., M.T.M. and R.S.J. designed and operated the data analysis workflow. N.C.K. and K.A.H. wrote the manuscript with editing by T.B.K., R.J.K., J.Y. and E.L.B. J.Y., M.P. and E.L.B. performed the experimental work. N.C.K., J.Y., K.A.H. and E.L.B. carried out the data analysis specific to this manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Arrowsmith, J. Trial watch, phase III and submission failures, 2007–2010. *Nat. Rev. Drug Discov.* **10**, 87 (2011).
2. Judson, R. *et al.* The toxicity data landscape for environmental chemicals. *Environ. Health Perspect.* **117**, 685–695 (2009).
3. National Research Council. *Toxicity Testing in the 21st Century: a Vision and a Strategy* (National Academies Press, Washington DC, 2007).
4. Bowes, J. *et al.* Reducing safety-related drug attrition: the use of *in vitro* pharmacological profiling. *Nat. Rev. Drug Discov.* **11**, 909–922 (2012).
5. Houck, K.A. & Kavlock, R.J. Understanding mechanisms of toxicity: insights from drug discovery research. *Toxicol. Appl. Pharmacol.* **227**, 163–178 (2008).
6. Chen, Y. & Shoichet, B.K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **5**, 358–364 (2009).
7. Kitchen, D.B., Decornez, H., Furr, J.R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **3**, 935–949 (2004).
8. Lounkine, E. *et al.* Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **486**, 361–367 (2012).
9. Berg, E., Yang, J. & Polokoff, M. Building predictive models for mechanism of action classification from phenotypic assay datasets. *J. Biomol. Screen.* **18**, 1260–1269 (2013).
10. Berg, E.L., Kunkel, E.J., Hytopoulos, E. & Plavec, I. Characterization of compound mechanisms and secondary activities by BioMAP analysis. *J. Pharmacol. Toxicol. Methods* **53**, 67–74 (2006).
11. Berg, E.L. *et al.* Chemical target and pathway toxicity mechanisms defined in primary human cell systems. *J. Pharmacol. Toxicol. Methods* **61**, 3–15 (2010).
12. Houck, K.A. *et al.* Profiling bioactivity of the ToxCast chemical library using BioMAP primary human cell systems. *J. Biomol. Screen.* **14**, 1054–1066 (2009).
13. Kleinstreuer, N.C. *et al.* Environmental impact on vascular development predicted by high-throughput screening. *Environ. Health Perspect.* **119**, 1596–1603 (2011).
14. Martin, M.T. *et al.* Predictive model of rat reproductive toxicity from ToxCast high throughput screening. *Biol. Reprod.* **85**, 327–339 (2011).
15. Sipes, N.S. *et al.* Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data. *Toxicol. Sci.* **124**, 109–127 (2011).
16. Dix, D.J. *et al.* The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* **95**, 5–12 (2007).
17. Kavlock, R. *et al.* Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* **25**, 1287–1302 (2012).
18. Tice, R.R., Austin, C.P., Kavlock, R.J. & Bucher, J.R. Improving the human hazard characterization of chemicals: a Tox21 update. *Environ. Health Perspect.* **121**, 756–765 (2013).
19. Ayala, A., Parrado, J., Bougria, M. & Machado, A. Effect of oxidative stress, produced by cumene hydroperoxide, on the various steps of protein synthesis: modifications of elongation factor-2. *J. Biol. Chem.* **271**, 23105–23110 (1996).
20. Lau, C. *et al.* Exposure to perfluorooctane sulfonate during pregnancy in rat and mouse. II: postnatal evaluation. *Toxicol. Sci.* **74**, 382–392 (2003).
21. Bogdanska, J. *et al.* Tissue distribution of (3)(5)S-labelled perfluorooctane sulfonate in adult mice after oral exposure to a low environmentally relevant dose or a high experimental dose. *Toxicology* **284**, 54–62 (2011).
22. Arredondo, J. *et al.* A receptor-mediated mechanism of nicotine toxicity in oral keratinocytes. *Lab. Invest.* **81**, 1653–1668 (2001).

23. Kleinstreuer, N. *et al.* A computational model predicting disruption of blood vessel development. *PLoS Comput. Biol.* **9**, e1002996 (2013).
24. Khaidakov, M. *et al.* Statins and angiogenesis: is it about connections? *Biochem. Biophys. Res. Commun.* **387**, 543–547 (2009).
25. Kohonen, T. Essentials of the self-organizing map. *Neural Netw.* **37**, 52–65 (2013).
26. McCann, F.E. *et al.* Apremilast, a novel PDE4 inhibitor, inhibits spontaneous production of tumour necrosis factor-alpha from human rheumatoid synovial cells and ameliorates experimental arthritis. *Arthritis Res. Ther.* **12**, R107 (2010).
27. Viguerie, N. *et al.* Multiple effects of a short-term dexamethasone treatment in human skeletal muscle and adipose tissue. *Physiol. Genomics* **44**, 141–151 (2012).
28. Augustine-Rauch, K.A. *et al.* Evidence for a molecular mechanism of teratogenicity of SB-236057, a 5–HT1B receptor inverse agonist that alters axial formation. *Birth Defects Res. A Clin. Mol. Teratol.* **70**, 789–807 (2004).
29. Owens, A.P. III & Mackman, N. Role of tissue factor in atherothrombosis. *Curr. Atheroscler. Rep.* **14**, 394–401 (2012).
30. Camici, G.G. *et al.* Rapamycin promotes arterial thrombosis in vivo: implications for everolimus and zotarolimus eluting stents. *Eur. Heart J.* **31**, 236–242 (2010).
31. Barua, R.S. & Ambrose, J.A. Mechanisms of coronary thrombosis in cigarette smoke exposure. *Arterioscler. Thromb. Vasc. Biol.* **33**, 1460–1467 (2013).
32. Cummings, S.R. *et al.* The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. Multiple outcomes of raloxifene evaluation. *J. Am. Med. Assoc.* **281**, 2189–2197 (1999).
33. Lipton, A., Harvey, H.A. & Hamilton, R.W. Venous thrombosis as a side effect of tamoxifen treatment. *Cancer Treat. Rep.* **68**, 887–889 (1984).
34. Kunkel, E.J. *et al.* An integrative biology approach for analysis of drug action in models of human vascular inflammation. *FASEB J.* **18**, 1279–1281 (2004).
35. Plavec, I. *et al.* Method for analyzing signaling networks in complex cellular systems. *Proc. Natl. Acad. Sci. USA* **101**, 1223–1228 (2004).
36. Robichaud, A. *et al.* Assessing the emetic potential of PDE4 inhibitors in rats. *Br. J. Pharmacol.* **135**, 113–118 (2002).
37. Kirkiacharian, S., Chidiack, H., Philibert, D., Van De Velde, P. & Bouchoux, F. Binding affinity to steroid hormone receptors and antiproliferative action on MCF-7 cells of coumarin derivatives and isoflavonoids. *Ann. Pharm. Fr.* **57**, 332–339 (1999).
38. Barna, T.M. *et al.* Crystal structure of pentaerythritol tetranitrate reductase: "flipped" binding geometries for steroid substrates in different redox states of the enzyme. *J. Mol. Biol.* **310**, 433–447 (2001).
39. Murotani, T., Ishizuka, T., Isogawa, Y., Karashima, M. & Yamatodani, A. Possible involvement of serotonin 5–HT2 receptor in the regulation of feeding behavior through the histaminergic system. *Neuropharmacology* **61**, 228–233 (2011).
40. Ouattara, B. *et al.* Changes of AMPA receptors in MPTP monkeys with levodopa-induced dyskinesias. *Neuroscience* **167**, 1160–1167 (2010).
41. Fell, M.J. *et al.* Activation of metabotropic glutamate (mGlu)2 receptors suppresses histamine release in limbic brain regions following acute ketamine challenge. *Neuropharmacology* **58**, 632–639 (2010).
42. Aizu-Yokota, E., Ichinoseki, K. & Sato, Y. Microtubule disruption induced by estradiol in estrogen receptor-positive and -negative human breast cancer cell lines. *Carcinogenesis* **15**, 1875–1879 (1994).
43. Thiede, A., Gellerich, F.N., Schonfeld, P. & Siemen, D. Complex effects of 17beta-estradiol on mitochondrial function. *Biochim. Biophys. Acta* **1817**, 1747–1753 (2012).
44. Cohen, S.B. *et al.* Evaluation of the efficacy and safety of pamapimod, a p38 MAP kinase inhibitor, in a double-blind, methotrexate-controlled study of patients with active rheumatoid arthritis. *Arthritis Rheum.* **60**, 335–344 (2009).
45. Ding, C. Drug evaluation: VX-702, a MAP kinase inhibitor for rheumatoid arthritis and acute coronary syndrome. *Curr. Opin. Investig. Drugs* **7**, 1020–1025 (2006).
46. Attene-Ramos, M.S. *et al.* The Tox21 robotic platform for the assessment of environmental chemicals - from vision to reality. *Drug Discov. Today* **18**, 716–723 (2013).
47. US Environmental Protection Agency. US Environmental Protection Agency Endocrine Disruptor Screening Program Comprehensive Management Plan. http://www.epa.gov/endo/pubs/edsp21_work_plan_summary%20_overview_final.pdf http://www.epa.gov/scipoly/oscpendo/pubs/EDSP_Comprehesive_Management%20Plan_%20021414_f.pdf(Jointly developed by the Office of Chemical Safety & Pollution Prevention and the Office of Water, EPA, 2014).

ONLINE METHODS

Cell culture. The panel of eight BioMAP systems employed are shown in **Table 1**. Preparation and culture of primary human endothelial cells (EC), PBMCs, neonatal foreskin fibroblasts (HDFn), BECs (Cell Applications, Inc., San Diego, CA), arterial smooth muscle cells (Lonza, Inc., Allendale, NJ) and keratinocytes (Cambrex, Inc., East Rutherford, NJ), as well as methods for the 3C, 4H, lipopolysaccharide, SAg, BE3C, HDFCGF, KF3CT and CASM3C systems were as previously described[11,12,34,48–50]. The following concentrations and/or amounts of agents were added to confluent microtiter plates to build the various systems: cytokines (IL-1β, 1 ng/ml; TNF-α, 5 ng/ml; IFN-γ, 20 ng/ml; IL-4, 5 ng/ml), activators (histamine, 10 μM; SAg, 20 ng/ml or lipopolysaccharide, 2 ng/ml), growth factors (TGF-β, 5 ng/ml; EGF, bFGF and PDGF-BB, 10 ng/ml) or PBMC (7.5×10^4 cells/well). All primary human cells used in this work were obtained under protocols that were reviewed by Institutional Review Board(s) (IRB) that operate in accordance with the requirement of EPA Regulation 40 CFR 26 and HHS Regulation 45 CFR 46 of the US Federal Government for the protection of human research subjects.

Compounds. Chemical information associated with the ToxCast library (i.e., chemical names, CASRN and substance description) was quality reviewed and structure-annotated within the US EPA's DSSTox project (http://www.epa.gov/ncct/dsstox/). The compound library is listed in **Supplementary Table 12** and a tabular listing and Structure Data Format (SDF) file of the complete ToxCast chemical library is available at: http://www.epa.gov/ncct/dsstox/sdf_toxcst.html. Analytical chemical analysis for the complete library is being conducted by OpAns (Durham, NC) and will be made publically available upon completion at the same URL. Compounds were tested in multiple batches (phase IIa, IIb, and IIc) in concentration-response in a single well per readout parameter. Phase IIa and IIb (700 compounds) were tested in duplicate at 40, 13.3, 4.4 and 1.48 μM. Due to the high level of reproducibility among the duplicate samples, the remaining phase IIc (100 compounds) set was tested in singleton at 40, 20, 10, 5, 2.5, 1.25, 0.625 and 0.313 μM. Out of 776 total compounds (800 with blinded replicates), 135 were donated pharmaceutical compounds from industry partners (Pfizer, GlaxoSmithKline, Roche, Merck, Aventis, Astellas), the majority of which exhibited preclinical or clinical toxicity and were not marketed drugs. Compounds were prepared in DMSO from 20 mM stock solutions, added 1 h before stimulation of the cells, and were present during the whole 24-h stimulation period. Final DMSO concentration was 0.2%. Colchicine, 1.1 μM, was included as a positive control. Compounds were tested in a blinded fashion and included two sets of sextuplet samples and seven sets of triplicates for quality control purposes.

Plate formats. Templates were prepared with seven compounds (four concentrations) per 96-well plate. One positive control (colchicine) and eight negative control wells (0.2% DMSO) were employed on each plate. Left and rightmost rows (A1-H1, A12-H12) were not employed for EPA compounds.

ELISA. The levels of readout parameters were measured by ELISA as described[11,12,34,48–51]. Briefly, microtiter plates are treated, blocked, and then incubated with primary antibodies or isotype control antibodies (0.01–0.5 μg/ml) for 1 h. Specific antibodies used for ELISA are listed in **Supplementary Table 13**. After washing, plates were incubated with a peroxidase-conjugated anti-mouse IgG secondary antibody or a biotin-conjugated anti-mouse IgG antibody for 1 h followed by streptavidin-horseradish peroxidase for 30 min. Plates were washed and developed with TMB substrate and the absorbance (OD) was read at 450 nm (subtracting the background absorbance at 650 nm). Quantification of TNF-α and PGE2 in the lipopolysaccharide system was done using commercially available kits according to the manufacturer's directions. Proliferation of PBMCs (T cells) was quantified by Alamar blue reduction and proliferation, and cytotoxicity of adherent cell types was quantified by sulforhodamine B (SRB) staining.

Other assessments. Overtly adverse effects of compounds on cells were determined by (i) measuring alterations in total protein using SRB staining, (ii) measuring the viability of PBMCs and (iii) microscopic visualization (3C system). SRB assay was performed by staining cells with 0.1% sulforhodamine B after fixation with 10% TCA, and reading wells at 560 nm. PBMC viability was assessed by adding Alamar blue to PBMC that had been cultured for 24 h in the presence of activators and compounds and measuring its reduction after 8 h. Samples were assessed visually according to the following scheme: 2.0 = cobblestone (nonactivated phenotype); 1.0 = activated (normal phenotype); 0.5 = lacy or sparse; 0.375 = rounded; 0.25 = sparse and granular; 0.1 = no cells in well. During this procedure, cells were also assessed for the presence of compound precipitates, and samples were flagged if precipitates are observed.

Data analysis. Measurement values for each parameter in a treated sample were divided by the mean value from eight DMSO control samples (from the same plate) to generate a ratio. All ratios were then $\log_{10}$ transformed. Visual categorical scores (see above) were similarly converted ($\log_{10}$ ratios of 0.3, 0.0, −0.3, −0.4, −0.6 and −1.0). Significance and hit prediction envelopes were calculated for historical controls (99% and 95%), and LECs were assigned as the minimum test concentration at which a significant response was observed (>99%). Concentration response plots were generated using an automated data workflow process in R (v2.13.0), and Hill functions were fit and half-maximal concentrations (AC_{50}) were calculated for responses exceeding a twofold change in either direction. Computer code for data processing is available in **Supplementary Software**. Overtly cytotoxic compounds were identified as generating profiles with one or more of the following readouts below the indicated thresholds: SRB < −0.3, PI or PBMC cytotoxicity < −0.3 or Visual score < −0.6 in one or more systems. A cytotoxicity filter was applied to the LEC and AC_{50} values to remove downregulation of protein targets due to overt cytotoxicity, primarily at the highest test concentrations. The complete set of results for the 776 chemicals for each of the 87 endpoints is included in **Supplementary Table 1**.

Correlation analysis. Bioactivity profile analysis was performed as described using Pearson's correlation, systems-weighted and real value Tanimoto metrics to compare nonovertly cytotoxic test compounds to a database of reference chemicals with known targets and modes of action[9,11,12]. A Pearson's correlation threshold of 0.7 was used to identify similar profiles, based on previous studies[10]. However, given that this cutoff was selected to control for the false-discovery rate (FDR), and FDR depends on dimensionality of the profile data, we have repeated the simulation to calculate the FDR for this new magnitude of multiple comparisons. Operating with a much larger BioMAP database than 2006, we used a slightly modified approach to generate the null distribution. Instead of permuting empirical profiles, random profiles were generated by sampling from a uniform distribution within each biomarker's range (min and max values). Both approaches take into account the differences in the amplitude of response of the different biomarkers. The biomarker ranges were developed from many experiments with different compound treatments in the BioMAP database, and all the biomarkers used in this paper have been profiled at least 7,000 times, many of them more than 10,000 times. Using this new null distribution, the FDR at a Pearson's cutoff of 0.7 for the 135 compounds would be 7.5% if using the same biomarkers as in Berg[10] . However, with a larger biomarker panel (eight BioMAP systems in this paper versus four systems in Berg[10]), the null and empirical distributions each become narrower, and the separation between empirical distribution and null distribution became much larger. Using this approach, we calculated the FDR at Pearson's cutoff of 0.7 for the 135 compounds, and found that the FDR is 0.2%, showing that this is a sufficiently conservative cutoff value.

SVM models. The 28 SVM models employed in the present study are fully described in reference[9] and summarized here. The 28 classes of mechanisms of action are as follow: AHR agonist, Hsp90 inhibitor, PI3K inhibitor, calcineurin inhibitor, IKK2 inhibitor, PKC (c+n) inhibitor, EGFR inhibitor, IL-17R agonist, proteasome inhibitor, EP agonist, JAK inhibitor, RAR/RXR agonist, estrogen receptor agonist, MEK inhibitor, Src family inhibitor, SR Ca^{2+} ATPase inhibitor, microtubule disruptor, TNF-α antagonist, GR agonist, microtubule stabilizer, vitamin D receptor agonist, H1 antagonist, mitochondrial inhibitor, mTOR inhibitor, HDAC inhibitor, PDE4 inhibitor, p38 MAPK inhibitor, HMG-CoA reductase inhibitor.

Selection of data for SVM models. For each mechanism class, where possible, multiple compounds from structurally distinct chemical classes

doi:10.1038/nbt.2914

were employed, and for all classes, profiles generated from compounds tested at more than one concentration were included. For the SVM models, 84 of 87 endpoints in the ToxCast data set were employed (visual scores (3C system) and TGFβ (HDF3CGF and KF3CT systems) measurements were not included). For each mechanism class, where possible, multiple compounds from structurally distinct chemical classes were employed, and for all classes, profiles generated from compounds tested at more than one concentration were included. All reference compounds, concentrations tested and data can be found in reference[9]. Compounds used for each model were selected based on literature information identifying them as selective for a specific target. BioMAP profile data (each profile consisting of data from a single compound at a single dose) for building mechanism models were selected based on their within-class consistency, with strongly toxic or weak profiles removed. Strongly toxic profiles are identified as profiles having values for three or more cell cytotoxicity endpoints (SRB and PBMC cytotoxicity) with $\log_{10}$ ratio < −0.3. For identification of weak profiles, the endpoint values (absolute $\log_{10}$ ratios) were digitized to a discrete number according to a lookup table, with each endpoint assigned a bit count from 0 to 18 depending on the magnitude of its $\log_{10}$ ratio value. The profile bit count represents the strength of a profile and is the sum of bits across all endpoints. For SVM model data, weak profiles were defined as those having 7 or fewer bits. After this initial filtering, profiles in each model were evaluated for within-class consistency. For this, profiles with Pearson correlation coefficients of at least 0.7 to at least one other profile in the same class were included. We have found that compounds that are highly target selective generate BioMAP profiles that are mathematically similar to each other over a wide range of concentrations, and we term compounds exhibiting this feature 'dose resistant'. Profile data used to build the models are included in Berg[9]. Due to the availability of selective compounds and our criteria above, different numbers of compounds and concentrations were selected for each class. As a result, some classes, such as the AHR agonist, contain few profiles (1 compound/4 doses), whereas other classes include larger numbers. A list of the mechanism classes, compounds and numbers of profiles that were used for generating SVM models as well as a table that contains the average class profile for each mechanism class can be found in reference [9]. These profiles were calculated by averaging the values for each biomarker endpoint for all profiles selected to build each SVM model.

Building SVM models. Predictive models for 28 mechanism classes were built using a two-class approach with SVM. SVM was selected as this method gave the best performance among a number of machine-learning algorithms that were tested, including Lasso, Random Forest, Gradient Boosting Method and Linear Discriminant Analysis (data not shown).

The R SVM package e1071 (http://www.r-project.org/) was used to build the SVM models for each mechanism class. Based on performance testing using cross-validation with reference and other data, the key variables for building SVM models were determined to be (i) row (profile) normalization of both model and test data with no column (marker) normalization; (ii) selection of linear kernel; and (iii) cost set to 500, resulting in a small increase in positive predictive value and a decrease in sensitivity. Other parameters were set to default values.

There are several different ways to employ SVM to build predictive models in our case. These include: "one class-versus-the rest of the classes"; "one class-versus-zero (or null) class"; or a "multiclass-classification mode", where for k classes, $k(k-1)/2$ number of "one-versus-one" binary classifiers are trained, and the appropriate class is assigned by a voting scheme. When evaluating the approach of "one class-versus-the rest of the classes," testing against external data sets revealed that a greater number of models performed very poorly. We did not test the multiclass-classification mode, given that the number of classes is large, $k = 28$ in our case, and building $k(k-1)/2$ number of "one-versus-one" binary classifiers quickly becomes unwieldy. In addition, the number of classes would be continuously growing as we generate data on more compounds and add additional classes to our set. Thus, in the current study, the mode of "one class-versus-zero (or null)" was selected as the most appropriate method. For the "null" class, we randomly generated a set of weak profiles using control data ($\log_{10}$ ratio values within the 95% significance envelope of the controls). In cases where profiles generated from the same compound but at different concentrations give different mechanism classes, separate SVM models were built as "one class-versus-the other class (or classes)" and profiles re-tested. The few number of missing values in the reference data set were filled in using the KNNimpute method from the R package imputation.

Criteria for assigning mechanism classes to profiles. For assigning mechanism classes to compound profiles at each concentration tested, individual profiles were tested against each of the 28 class models. The resulting DV against each model was calculated as described in reference [9] with DV reflecting the distance from the hyperplane separating the two classes (in this case, the selected mechanism class versus the null class). Support vectors of the positive class (subset of the positive class profiles) have DVs near 1. Any DV > 0 indicates class membership; however, as DV increases, confidence in class membership will increase.

For each test profile, however, there may be more than one class that gives a DV > 0. In these cases, the class with the highest DV becomes the predicted class. The performance of this scheme ("largest DV = assigned class") was tested using cross-validation on reference profiles, where 10% of the profiles ("test profiles") from each class was set aside, the remaining 90% of profiles from each class was used to build models. Then the set aside "test profiles" were evaluated and assigned according to the "largest DV = assigned class" scheme. From the results, measures of performance: Positive Predictive Value (PPV = TP/(TP+FP), TP = true positive, FP = false positive) and sensitivity (= TP/(TP+FN), FN = false negative) were calculated.

As described above, the process of generating the "null" class, randomly generated as a set of weak profiles using control data, introduces some variance to the DVs obtained each time a prediction is run. The s.d. of multiple runs was found to be 0.005. Thus in cases where multiple classes give DV > 0, we required that the difference between the highest DV and second highest DV be > 0.03, which represents six s.d. for class assignment.

Modified nonlinear mapping (NLM). The modified NLM technique employed in the present study has been described previously[10,11,48]. Methods are repeated here for convenience. The function similarity map uses the results of pairwise correlation analysis to project the "proximity" of related profiles from multidimensional space to two dimensions. The two-dimensional (2D) projection coordinates were generated by applying a modified nonlinear mapping technique, using a modified stress function by Clark[52]. A gradient descent minimization method was used to minimize the modified stress function, starting from a set of initial positions (e.g., from principal components analysis). Distances between compounds are representative of their similarities, and lines are drawn between compounds whose profiles are sufficiently similar, with metrics that are above the selected thresholds.

Self-organizing map (SOM). The Kohonen SOM is an unsupervised learning approach that allows for visualization of patterns in the data by mapping the underlying topology. This technique employs self-organizing neural networks to reduce the dimensionality of the data so that it may be plotted in 2D space. SOMs are based on a competitive learning system implemented by lateral inhibition connections where neurons are competing for spatial locations, resulting in an emergent structure that optimizes the similarity of neighboring neurons. The process of self-organization can be summarized as follows. First, all the connection weights between neurons are initialized with small random values, and then the neurons compute their respective values of a discriminant function. The neuron with the smallest value of the discriminant function is the "winner," and determines the spatial location of a topological neighborhood of excited neurons that may cooperate with one another. The excited neurons decrease their individual values of the discriminant function through adjustment of the associated connection weights, such that the response of the winning neuron to the subsequent application of a similar input pattern is enhanced. These steps are repeated and the feature map is refined until convergence is reached, where the map no longer changes and provides an accurate statistical quantification of the input space. Parameter values must be carefully chosen to reflect the dimensionality of the data; typically the number of nodes reflects the number of input samples (here, a 10 × 10 grid was chosen), and the number of training iterations is at least

500 times the number of neurons in the network, here set at 100,000 iterations[25]. The SOM analysis was performed in Partek Genomics Suite 6.6.

48. Kunkel, E.J. *et al.* Rapid structure-activity and selectivity analysis of kinase inhibitors by BioMAP analysis in complex human primary cell-based models. *Assay Drug Dev. Technol.* **2**, 431–442 (2004).
49. Williams, O. *et al.* Discovery of dual inhibitors of the immune cell PI3Ks p110delta and p110gamma: a prototype for new anti-inflammatory drugs. *Chem. Biol.* **17**, 123–134 (2010).
50. Berg, E.L., Kunkel, E.J. & Hytopoulos, E. Biological complexity and drug discovery: a practical systems biology approach. *Syst. Biol. (Stevenage)* **152**, 201–206 (2005).
51. Melrose, J., Tsurushita, N., Liu, G. & Berg, E.L. IFN-gamma inhibits activation-induced expression of E- and P-selectin on endothelial cells. *J. Immunol.* **161**, 2457–2464 (1998).
52. Clark, R.D., Patterson, D.E., Soltanshahi, F., Blake, J.F. & Matthew, J.B. Visualizing substructural fingerprints. *J. Mol. Graphics Model.* **18**, 404–411 (2000).

doi:10.1038/nbt.2914