Kohonen, Pekka; Parkkinen, Juuso A.; Willighagen, Egon L.; Ceder, Rebecca; Wennerberg, Krister; Kaski, Samuel; Grafström, Roland Č.

A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury

Published in:
Nature Communications

DOI:
10.1038/ncomms15932

Published: 03/07/2017

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Kohonen, P., Parkkinen, J. A., Willighagen, E. L., Ceder, R., Wennerberg, K., Kaski, S., & Grafström, R. C. (2017). A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury. Nature Communications, 8, 1-15. [15932]. https://doi.org/10.1038/ncomms15932
A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury

Pekka Kohonen1,*, Juuso A. Parkkinen2,*, Egon L. Willighagen1,3, Rebecca Ceder1, Krister Wennerberg4, Samuel Kaski2,5 & Roland C. Grafström1

Predicting unanticipated harmful effects of chemicals and drug molecules is a difficult and costly task. Here we utilize a ‘big data compacting and data fusion’—concept to capture diverse adverse outcomes on cellular and organismal levels. The approach generates from transcriptomics data set a ‘predictive toxicogenomics space’ (PTGS) tool composed of 1,331 genes distributed over 14 overlapping cytotoxicity-related gene space components. Involving \( \sim 2.5 \times 10^8 \) data points and 1,300 compounds to construct and validate the PTGS, the tool serves to: explain dose-dependent cytotoxicity effects, provide a virtual cytotoxicity probability estimate intrinsic to omics data, predict chemically-induced pathological states in liver resulting from repeated dosing of rats, and furthermore, predict human drug-induced liver injury (DILI) from hepatocyte experiments. Analysing 68 DILI-annotated drugs, the PTGS tool outperforms and complements existing tests, leading to a hereto-unseen level of DILI prediction accuracy.
Products and compounds to be placed in products undergo safety testing to variable levels of depth and complexity. Even highly regulated and rigorous testing practices unfortunately still fail in detecting inherent toxicological properties, of which some effects become evident only after exposure to a marketed product. For the drug developing pharmaceutical industry, severe drug-induced liver injury (DILI) remains an enormous problem, as its occurrence is often not predicted.

Systems biology-based assays relying upon quantitative mechanistic information are increasingly envisaged as cornerstones of future safety evaluation of drugs and chemicals. Accordingly, various modelling approaches have analysed ‘omics data sets to generate biomarker signatures or to characterize mechanisms of toxicity at a system-wide level, but suffer from the high dimensionality of omics data relative to sample number as well as problems in scaling across experimental systems (for example, hepatocyte cultures to liver) or species (for example, rat to human). Complicating predictive biomarker discovery, different genes and gene families will unlikely have similar dose response, so dose-dependent transitions will influence the classification of (toxic) modes-of-action. Addressing the issue, the adverse outcome pathways (AOP) concept was designed to facilitate the use of modern human-specific in vitro models to understand toxicity and disease pathways at multiple levels of biological organization.

Embracing the complete chain of events from the first compound-induced molecular cellular changes to influence adversity on cellular, organ, individual and even population levels, the AOP concept has been incorporated into in vitro tests for an accepted replacement of animal experiments. Large-scale consortia and projects, for example, Tox21, ToxCast and SEURAT/EUToxRisk, address the issue by complementing the traditional structure-based analysis with high-throughput in vitro assays that capture generalizable cytotoxicity-related responses (Supplementary Fig. 2).

The PTGS was defined with probabilistic component modelling of the combined CMap and NCI-60 data, as the minimally sized component gene space that captured dose-dependent cytotoxicity within the complete data set. Modelling of the CMap transcriptomics response was done in a two-step semi-supervised manner: performing unsupervised component modelling on the whole CMap data set and subsequently using the component models and the NCI-60 cytotoxicity data to build supervised models. Gene sets that represent the components were then derived and applied as a basis for predictive scoring: Fig. 1 and Supplementary Fig. 1 depict the overall analysis and validation strategies that generated the PTGS. The protocol extracted and reduced the number of data points, compounds and genes. Positive concentration-dependent data indicated which CMap measurement instances had been produced at a concentration inducing at least 50% growth inhibition, and therefore reflected a potentially cytotoxic response (Supplementary Fig. 2).

The modelling approach decomposed the entire pre-processed CMap data, consisting of 3062 instances (an instance represents a chemical treatment of one cell line), to 100 partially overlapping and non-orthogonal components (Fig. 2a). Superimposing the NCI-60 data enabled integrating 222 CMap compounds and 492 instances, measured over a ~10^6-fold potency range relative to GI50. This crossover data set permitted the selection of an optimally sized set of the 14 most cytotoxicity-associated components, which defined the PTGS (Fig. 2bc). With area under the ROC curve (AUC) peaking at 40 components, these fourteen components give 95% of the highest AUC value for predicting whether cytotoxicity was above the GI50-level. Most of the components were proportionally active in all cell lines, suggesting that they capture generalizable cytotoxicity-related responses (Supplementary Fig. 3a). Hierarchical clustering of the PTGS revealed clustering of the components into one group comprising a majority of the components, another less distinct cluster (including E and K), and one outlier component (L), demonstrating that most of the components had overlapping gene activities (Supplementary Fig. 3b).

Defining a cytotoxicity scoring concept from the PTGS. A PTGS-based scoring concept was defined based on the premise that activation of any of the PTGS components indicated dose-dependent cytotoxicity (Fig. 2, Supplementary Figs 4 and 5 and Supplementary Data 1). The cytotoxicity effects of the compounds correlated with the transcriptional variation (Pearson correlation is 0.69; P-value < 2.2 × 10^-16, Fig. 2cd). The 14 components overall responded over a wide dose-range and, as expected, primarily became active at or above the GI50-dose (Fig. 2dg). The PTGS therefore covered instances with varying numbers of differentially expressed genes and toxicities.
Some CMap instances represented toxicities above the TGI level (Fig. 2e,f); such instances tended to have many differentially expressed genes (Fig. 2f) and highly active components A–C, D and F–H (Supplementary Figs 4 and 5a). On the other hand, instances belonging to the smaller cluster and components E, K, I and M tended to be active at around the GI50 growth-inhibitory level and displayed smaller numbers of differentially expressed genes (Fig. 2c,e; Supplementary Fig. 5b). A low number of instances that reflected cell-killing doses, that is, LC50, (Fig. 2g; Supplementary Fig. 5c) were also covered by the PTGS and components A–C in particular. A PTGS scoring concept for ranking compounds for probability of cytotoxicity was thereafter

**Figure 1 | Generating the Predictive Toxicogenomics Space (PTGS) concept.** (a) The probabilistic component modelling leading to the PTGS scoring concept utilized latent Dirichlet allocation. This unsupervised method uncovers common themes that describe collections of profiles, seeking associations between compound treatments ('instances') and differential expression of gene sets, leading to data reduction and discovery of components that can be used to quantitatively classify new gene expression profiles. (b) Probabilistic modelling of transcriptomics and cytotoxicity data from exposed cells was used to identify specific component models representing mechanistic aspects of the responses and genes activated by the treatments. Scores derived either from the models or the gene set encapsulated by the PTGS serve to predict a variety of types of dose-dependent cytotoxicity effects; the analysis steps are presented in detail in Supplementary Fig. 1. Validation of the PTGS scoring concept encompassed: bioinformatics-driven assessment of the component-associated genes relative to genes known as cytotoxicity-related, generation of cellular cytotoxicity screening data for comparison of the omics-based PTGS relative to quantitative structure-activity relationships (QSAR) analysis, and finally, assessment of the in vitro to in vivo extrapolation applicability of the PTGS in two ways against the Open ‘Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system’ (TG-GATEs), that is, for prediction of histopathology of rats subjected to repeat dose-toxicity studies, and for prediction of human drug-induced liver injury from human and rat hepatocytes. Numbers of compounds assessed within each omics data set used to establish the PTGS and to validate the concept are indicated.
Figure 2 | Generating the PTGS and establishing the cytotoxicity-scoring concept. (a) Selecting the number of probabilistic components to retrieve as many biologically significant associations with as few components as possible. (b) Selecting an optimal size of the PTGS based on cytotoxicity-predictive performance relative to the NCI-60 data. (c) The 14 PTGS components (labelled) ranked based on their probability-weighted mean concentration-dependent cytotoxicity values (that is, log10 CMap–log10 GI50 concentration) versus the number of associated genes. (d) Correlation of the number of differentially expressed genes with the concentration-dependent cytotoxicity. Colour and size indicate amount of transcriptional variation explained by the PTGS that is, the component-based score (n = 492). (e,f) Instances with a small number of differentially expressed genes tend to have cytotoxicity below the TGI level (blue oval), whereas (g) compounds profiled at cell-killing doses (> LC50) show greater differences (green circle). (h) Analysis of component-based PTGS scores versus concentration-dependent cytotoxicity was used to determine (i) a cut-off, plotted here against the proportion of instances above the GI50-level. Dashed red line indicates the threshold at the GI50-level and the dashed black line the cut-off at 0.12 when ~ 50% of CMap instances are above GI50. (j) The gene-based scoring, based on the proportion of active PTGS-related genes, was evaluated similarly. (k) The cut-off was set at 25% (cf. Supplementary Figs 4 and 5, for data see Supplementary Data 1).
defined, that is, being the sum of the contributions of the 14 components relative to the other 86 components; for the calculation formula, see Materials and Methods (Fig. 2h). The score served to predict whether an instance was measured at a concentration above GI50, employing a decision-threshold designed to maximize sensitivity versus specificity (Fig. 2i).

The gene alterations corresponding to the PTGS components were next assessed and applied to cytotoxicity scoring (Fig. 2c,j and k). Overall 1331 genes associated to PTGS (listed in Supplementary Data 2). In analogy to the components, the PTGS-associated genes exhibited a dose–response relationship that could be used to predict similar accuracy whether an instance was measured at a level above GI50. The composite of these results confirmed that the PTGS scoring constitutes a tool for predicting cytotoxicity over a wide range of concentrations and corresponding gene alterations.

PTGS captures diverse cytopathological changes. A number of different approaches were taken to validate the functionality of the PTGS (Figs 1b and 3, Supplementary Fig. 6 and Supplementary Data 3–8). The PTGS gene lists were enriched in a variety of basic biological and metabolic processes associated to growth inhibition, cellular cytotoxicity and stress response pathways, as well as to pathological effects in liver, kidney and heart; this analysis is plotted as an eye diagram with reference to highly associated instances (drug/cell line-pairs) (Fig. 3a).

The pathological effects included changes typically associated with adverse drug reactions and those seen in repeated-dose toxicity studies of laboratory animals, for example, hepatic fibrosis4,12,16,36. Components A–C enriched most strongly for toxicology studies of laboratory animals, for example, hepatic fibrosis4,12,16,36. Components A–C enriched most strongly for liver necrosis/cell death, whereas E and K enriched for liver cell proliferation, among other organ effects. The analysis also indicated enrichment for receptor-mediated toxicity, for example, Aryl Hydrocarbon Receptor Signaling and LXR/RXR Activation.

The genomic structure complexity of PTGS involved overall 1331 genes (716 up-regulated and 835 down-regulated, meaning that around 200 genes had up-regulation in one or several components, and down-regulation in others), 101 molecular networks and 97 transcriptional regulators (Supplementary Data 3). Regarding the respective components, the proportional network complexity varied extensively, and was only partially related to the gene numbers (Fig. 3b). Components towards the higher toxicity end exhibited mostly up-regulated genes, whereas the lower toxicity end components showed mostly, or even entirely, down-regulated genes (Supplementary Data 3). Component C contained many genes associated to many variables, whereas components G and B associated to many variables from fewer gene numbers. The number of upstream regulators also varied extensively among the components. Commonly involved transcriptional regulators, spanning three or more components, are shown in Fig. 3c. Totally 19 of these regulators were found in the 1331 gene set that constituted PTGS (Supplementary Data 5). Components such as G enriched for regulatory factor gene signatures (for example, TP53, NFKB1A), inflammation-related gene ontology categories as well as stress from DNA damage and reactive oxygen (Supplementary Fig. 6). Components E and K enriched for cell cycle and cell division related categories, for example, S phase of mitotic cell cycle, as well as related regulators including MYC, CDKN2A and E2F1. Of all the regulators, P53, EP300 and CDKN2A were associated with the largest numbers of components (Fig. 3c). The functional associations of the components based on gene-level analysis thus reflected the component-level clustering based on CMap instances; indicating that many aspects of the probabilistic model are preserved within the gene lists (Supplementary Fig. 3b).

Comparison to the transcriptional regulators addressed in the comprehensive ToxCast project21 indicated coverage of 14 of the 35 regulators identified in the PTGS (Supplementary Data 5). The transcriptional regulators non-examined in ToxCast were distributed evenly across the 14 components of the PTGS, indicating potentially a different coverage of cytototoxic mechanisms relative to the ToxCast assays; the PTGS genes overall matched 22% of the genes annotated to ToxCast, see Materials and Methods.

PTGS-based grouping outperforms structure-based grouping. The components variably enriched for particular structural and functional classes among the CMap compounds, for example, A–C were enriched for protein synthesis inhibitors and cardenolide glycosides (Fig. 3a and Supplementary Data 8). Grouping of diverse classes of compounds to specific components supported applicability of the PTGS tool as such to connectivity mapping and compound grouping. Cell culture experiments were therefore designed to challenge the fact that structure basis is so far the one accepted means of grouping compounds for avoidance for toxicity testing in regulatory contexts22,24. Cytotoxicity screening of 38 CMap test compounds, for which such data are not available in the NCI-60 data, and a set of 16 NCI-60-assessed controls demonstrated a wide range of cytotoxicity effects, and moreover verified the comparability of the chosen cytotoxicity assay relative to the NCI-60 assay (Fig. 4a–c). Both gene set-based and ‘component-based’ scores predicted the cytotoxicity of the non-NCI-60-assessed compounds with high sensitivity and specificity (Fig. 4d), and consistently outperformed predictions generated from quantitative structure-activity relationships (QSAR) analysis. This result was obtained applying either the structures for 201 of the 222 training compounds (448 of the 492 instances with cytotoxicity data), or the 35 structures of the 38 validation compounds (85 of the 91 instances) (Fig. 4, Supplementary Fig. 7 and Supplementary Data 9).

PTGS predicts dose-dependent liver toxicity. The applicability of the PTGS scoring concept was next assessed in relation to non-dividing normal hepatocytes and rat liver pathology data in the TG-GATEs toxigenomics database (Figs 1b and 5a–h and Supplementary Data 10–15). Applying component-based scoring, human hepatocyte experiments generated increased scores with concentration at both 8 and 24 h exposures (Fig. 5a,b). Machine learning-based inference with TG-GATEs 28-day study data was then used and resulted in the selection of components G, H, I and N, as they had the highest liver toxicity predictive ability (see Materials and Methods; Supplementary Data 13–14). Capture of dose–response applying gene-based scoring with these components was verified in human and rat hepatocytes (Fig. 5c,d). Scoring using these components predicted diverse pathological changes in 45 combinations of pathological findings and severity grade, covering 1689 distinct treatments with 143 compounds in the rat liver 28-day repeated dosing data set, and thus constituted a functional DILI score (Fig. 5e,f). All 17 types of pathological effects were captured at high sensitivity and specificity, including severity grades (Fig. 5e–g; Supplementary Data 15). The endpoints included liver necrosis/cell death, ground glass appearance, fibrosis, hyperplasia/hyperproliferation (swelling), cholestasis (degeneration, fatty) and further pathologies (Fig. 5e,f), as well as liver pathologies as aggregate endpoints according to grade and even ‘death’, the latter being the one organism-level endpoint scored. These results exemplified effective PTGS-based extrapolations from cells to organ-level, as well as between species.
Figure 3 | Validation of the PTGS using gene set enrichment analysis. (a) ‘Eye diagram’ showing the associations between the genes associated with the 14 PTGS components (middle, colour) and the top 5 CMap instances (left) and overrepresented toxicological functions (right). Line widths indicate association strengths. The components have been sorted according to similarity, as shown in Supplementary Fig. 3b; data in Supplementary Data 4. (b) Biological and toxicological complexity of the PTGS components defined as the proportion of results (above a set statistical threshold) in each analysis category ascribed to the component gene set. Numbers above bars denote the numbers of genes in each component. Details of the data are found in Supplementary Data 3–7. (c) Frequency plot of the upstream regulator enrichments for the PTGS components depicting multiple transcriptional regulators associated with stress responses, inflammation and with cell division. For data and further related analyses, see Supplementary Fig. 6b and Supplementary Data 5.
PTGS broadly predicts human drug-induced liver injury. The hypothesis was thereafter tested that toxicogenomics changes measured in vitro can be used to predict DILI potential in human patients (Figs 1b and 5i, Supplementary Fig. 9 and Supplementary Data 16–18). For in vitro-based prediction we used the in vivo phenotype ‘presence of toxicity’ that is, the presence of pathological findings in the animal liver, to tune the predictions (see Material and Methods). A threshold for the magnitude of the score was set at a level above which at least 50% of instances showed pathological changes (Fig. 5g,h). Basis for predicting clinical exposure levels of concern is shown and explained in Supplementary Fig. 9a and b, including legend. Two withdrawn
Figure 5 | Validation of the PTGS using in vitro and in vivo profiles from the TG-GATEs toxicogenomics database. The increase with dose in the proportion of treatments exceeding the virtual GI50-level (dashed line) in human hepatocytes measured at (a) 8 h (n = 388) and (b) 24 h (n = 394) and in (c) human (n = 388) or (d) rat (n = 419) hepatocytes measured at 8 h, using either the component-based (a,b) or the gene-based (c,d) methods (Supplementary Data 10,11). The PTGS DILI score (for analyses see Supplementary Data 12–15, Supplementary Fig. 8), defined as the score given by the most sensitive component from among G, H, I and N, (e,f) predicts the severity grade (denoted by colour) and covers 17 different types of histopathological changes observed in repeated dose treatments of rats for up to 28 days. (g) Separation between positive and negative classes increases with the severity of histopathological changes from present to severe; n = 463, 448, 282, 116 and 30 of 1689 total. (h) Defining a threshold for the score above which more than 50% of the observations have histopathological changes present (dashed line). (i) The ability of PTGS to predict clinical exposure levels raising DILI concerns was tested and compared to other in vitro assays. Numbers of matching compounds with rat hepatocyte data are indicated inside red bars. PTGS, by itself, outperforms the other approaches, and in combination with other hepatocellular-based assays achieved a positive predictive ability of 72-86% without a loss of specificity (further details in Supplementary Fig. 9 and Supplementary Data 16-18).
drugs (nimesulide and benzbromarone) and one drug with a good safety profile (aspirin) illustrate the calculations. Associated to idiosyncratic DILI, nimesulide and benzbromarone are metabolically converted drugs for which the mechanism of action/toxicity is not known precisely and likely to vary between patients. The PTGS DILI score is activated in a dose-dependent manner by both compounds, in hepatocytes (Supplementary Fig. 9b) and rats (Supplementary Data 12). Interestingly, aspirin at the highest doses, also caused liver injury and activates the DILI score, reaffirming that dosing and exposure needs to be taken into account when assessing compound toxicities.

A literature search of the TG-GATEs data thereafter enabled annotating 68 compounds with their therapeutic Cmax values and results on liver toxicity. The gathered information implied that the data set in several instances reflected therapeutic doses also below the Cmax values (Supplementary Data 16). The PTGS DILI score was then applied to derive a safety margin of exposure relative to the therapeutic Cmax concentration (see Materials and Methods). The DILI potential of the annotated agents was found to be predictable to a level of 100% specificity and 71% sensitivity with rat hepatocyte data (Supplementary Fig. 9c and Supplementary Data 17). Differently, the similar analysis with human hepatocytes indicated 100% specificity and 58% sensitivity. Interestingly, while the rat hepatocyte data performed best overall, perhaps owing to low sample variation, the human data performed better and exceeded the rat performance, by 73% versus 71%, in predicting the most clinically troubling withdrawn and boxed warning labelled toxicities in the Liver Toxicity Knowledge Base. The analyses provided the similar level of prediction with the subset of drugs labelled ‘most DILI-concerning’. Compared to other in vitro methods applied to predict DILI, PTGS provided better predictive performance, and moreover, provided further improved prediction levels in combination with the methods (Fig. 5i, Supplementary Data 18).

**Discussion**

This study represents a large-scale data analysis aimed at addressing broadly human health and safety of chemical compounds, including drug molecules. Coupling of omics data to the prediction of dose-dependent induction of cytotoxicity effects resulted in the first ever description of a PTGS. Representing a comprehensively validated construction, it captures a wide range of dose-dependent cytotoxicity effects, and therefore serves to improve prediction of hepatocellular toxicity and liver pathologies in humans and rats relative to existing methods.

The data fusion underlying the PTGS tool involved extensive probabilistic modelling-driven transformation, compacting and selection of the data points, instances, compounds, components and genes (summarized for overview in Table 1). The level of reduction was to between 1 and 10% of the input data; for example, the CMap was reduced, transformed and decomposed to 0.7% of the original data size, and altogether, 22% of the gene expression alterations, that is, 1331 versus 6064 genes (11% of all measured transcripts), connected to cytotoxicity-related transcriptomic changes. As around 25% of the CMap gene expression profiles likely reflect cytotoxicity above G150 (cf. Figs 2c and 4c), the PTGS is based on, and covers, a significant portion of the CMap gene and sample dimensionality. Giving further support to this assumption, the 14 components included those of the overall 100 original components with the most extensive gene expression changes (cf. Fig. 2c). The PTGS calculation methods are most likely equally applicable to both microarray and RNA-seq gene expression data. Because of the ability of RNA-seq to detect alterations more sensitively than microarrays, it may detect activation of PTGS at lower doses, an issue that would be testable in sufficiently large and matching data sets. Further studies could also consider the PTGS approach and scoring concept using proteomics and metabolomics data.

Overall, the described 'big data-driven' analysis enabled: (1) a virtual cellular cytotoxicity probability estimate intrinsic to omics-data, (2) calculation of toxic exposure thresholds for compound effects, (3) grouping of compounds into mechanistically similar classes, (4) assessment of the cytotoxicity of CMap profiles, with implications for using the database and gene expression profiles generally for mode of action studies, (5) coverage of adverse outcome-coupled toxicity effects involving a multitude of transcriptional regulators, (6) prediction of known measured liver toxicity and pathology effects in the TG-GATEs, including a ‘severity-grade response’, from data obtained in cultured cells (for example, rat/human hepatocytes) and laboratory animals (for example, in rats) and, finally, (7) prediction of exposure levels raising concern for human DILI from hepatocyte experiments. The latter analysis includes opportunity for improved preclinical ab initio prediction of safety margin for novel drug molecules, while serving in a complementary manner to raise the prediction level of existing evaluation tests (range 14–38%; cf. Fig. 5i), including a commercially available test. An ab initio testing of a previously non-tested compound under the PTGS concept would generate a probability score for both cytotoxicity and liver pathology. Under a qualified, preclinical efficacy drug testing protocol, a range of human-relevant concentrations would be derived that could be assessed with PTGS to then include risk-prediction of DILI to this analysis. The overall results would constitute a qualitative and quantitative hepatotoxicity/DILI measure, including coverage of mild to overt effects. The DILI prediction scoring could likely be further refined from standardizing drug concentrations relative to the therapeutic Cmax more precisely and by incorporating further negative control compounds. Furthermore, future connectivity mapping-based testing with PTGS components to predict in vivo outcomes from in vitro hepatocyte toxicogenomics data would likely indicate further the applicability of PTGS in relation to specific pathological states. Applying the concept to capture further organ toxicities is an even further interesting task, agreeing with that the bioinformatics assessment indicated component association to a diversity of heart and kidney conditions (cf. Fig. 3a; Supplementary Data 4).

Being the focus of the current study, DILI is multifactorial, sometimes receptor-mediated or occurs in response to gross stress. Idiosyncratic DILI occurs unpredictably, with variable length latency and sometimes without dose-dependency. Interestingly, the PTGS classified idiosyncratic DILI-drugs in dose-dependent manners, for example, nimesulide (cf. Supplementary Fig. 9b). The predictive components (G,H,N,I) might therefore quantitatively evaluate a relatively broader complexity of DILI-inducing mechanisms than existing tests. Interestingly, these components associated to lower cytotoxicity in the CMap training data set (cf. Fig. 2c, Supplementary Data 3), implicating that relatively milder, rather than severe, cellular toxicity effects might better reflect at least certain DILI mechanisms. We hypothesize overall that the current work could serve to stimulate the integration of component models in future DILI studies, and generally, scoring concepts into AOP-based risk assessment strategies. For example, the PTGS component gene sets are enriched in liver fibrosis-related gene signatures and detect hepatocellular damage markers in the fibrotic mechanism. Thus, PTGS could be used to ‘biomark’ key events detailed in the corresponding AOP.
important role of implying broad applicability of the PTGS.

Thus, the demonstration of this mostly expected outcome fills the

Overall applied 84M data points and 1217 compounds to generate the PTGS (cf. Table 1), and assessed 250 × 10⁶ data points overall, including the TG-GATEs data. Variously from 140 to 170 compounds were assessed to validate the scoring concept. Being a small but important part of the current study, the toxicogenomics-based scoring outperformed the QSAR-based toxicity predictions (cf. Fig. 4d). Regulatory agencies such as the European Chemicals Agency and the United States Environmental Protection Agency are increasingly advocating for the inclusion of transcriptomics data and new approach methodologies in chemicals risk evaluation1,2,11,18,24. Thus, the demonstration of this mostly expected outcome fills the important role of implying broad applicability of the PTGS concept also outside of drug discovery studies. Challenging to traditional means of optimizing biological testing practices and coupled mechanistic reasoning, the CMap-derived PTGS establishes that even tumour-derived cellular models with known aberrant metabolism and differentiation capacity can be used to capture mechanisms that predict in vivo dose-dependent liver toxicity in a cross-organism manner. The rich variety of agents assessed in the CMap, including direct acting cytotoxic cancer drugs, may potentially underlie the capturing of cytotoxicity/pathology of agents requiring metabolism to exert their effects. Although complex in overall structure and function, PTGS is naturally suited for analysis in high-throughput transcriptomics assays, for example, the Tox21 platform10,33. We emphasize finally the full adherence of our study and the PTGS concept to replacing animal testing protocols with quantitative systems toxicology and human cell culture-based experiments, arguing overall for broad and opportune applicability of the PTGS concept in diverse future safety testing practices.

Table 1 | Key features of data processing that generated the Predictive Toxicogenomics Space (PTGS).

| Data items | Number | Percentage |
|-----------|--------|------------|
| Data Points | | |
| Entire Data set (CMap) | 84 M | 100 |
| Data set after pre-processing of the most abundant platform (A) | 34 M | 41 |
| Data set mapped to MSigDB-C2 gene sets (B1) | 8 M | 9.5 |
| Data set mapped to the component model (B2) | 0.6 M | 0.7 |
| PTGS scores calculated from the data set (D3) | 3,062 | 0.004 |
| Instances | | |
| Entire Data set (CMap) | 6,100 | 100 |
| Instances after selecting one array platform, pre-processing and averaging (A) | 3,062 | 50 |
| Instances in the crossover data set with toxicity data (C1) | 492 | 8 |
| Instances with toxicity above GI₅₀ (C2) | 121 | 2 |
| Compounds | | |
| Entire Dataset (CMap) | 1,309 | 100 |
| Compounds after pre-processing of the most abundant platform (A) | 1,217 | 93 |
| Compounds with toxicity data (C1) | 222 | 38 |
| Compounds with toxicity above GI₅₀ (C2) | 68 | 5 |
| Components | | |
| Full component model (B2) | 100 | 100 |
| PTGS components (D3) | 14 | 14 |
| DILI predictive components (E5) | 4 | 4 |
| Genes | | |
| Genes mapped to Ensembl IDs in the CMap | 11,948 | 100 |
| HG-U133A series (A1) | 11,350 | 95 |
| Genes after pre-processing of the most abundant platform (A2) | 6,064 | 51 |
| Genes responding to chemical perturbations (A4) | 1,331 | 11 |
| PTGS associated genes (D5) | 299 | 2.5 |

*Steps in data reduction and analysis, letters refer to detailed explanations in Supplementary Fig. 1.
†M = 1 million data points.
‡Percentages calculated from the first item in the category.

Methods

Pre-processing of the connectivity map data set. To decrease the low-intensity noise in the data the Connectivity Map (CMap) raw data25, the CEL-files (downloaded from http://www.broadinstitute.org/cmap; and E-GEOD-5258 for build01) were robust multi-array normalized with R/Bioconductor-package aroma.affymetrix and mapped to Ensembl gene identifiers (custom CDF version 12, http://brainarray.ncbi.nlm.nih.gov/Brainarray/Database/Chroma/) CDF downloadable. Results from the most abundant microarray platform (HT-HG-U133A) were used, containing measurements for the three cell lines MCF7, PC3 and HL60. To further reduce the noise in the expression data the 5% of the genes displaying the highest variance in the control measurements were removed48. Differential expression was then computed as the log₂-ratio between the drug treatments and respective control measurements. The CMap measurements had been made in batches. In the case of multiple negative controls per batch, adapting established procedures, a more robust control was formed by calculating a mean of the control measurements after first removing, as an outlier, the control with the highest (Euclidean) distance to the other controls. To balance between the varying sample sizes for different compounds, the instance for each compound and cell line with the strongest effect, measured as the highest (Euclidean norm of) response, was selected for further analysis. A total of 18 compounds in the data set had more than one (and mostly two) concentrations. To balance between the varying sample sizes for different compounds, the instance for each compound and cell line with the strongest effect, measured as the highest (Euclidean norm of) response, was selected for further analysis. The resulting gene expression data consisted of 3062 treatment instances (compound and cell line pair) and profiles for 1217 distinct compounds in the three cell lines (MCF7, PC3 and HL60, with 1203, 1131 and 728 instances per cell lines, respectively). For further details see Table 1.

Probabilistic component modelling. It was assumed that compound treatments may activate multiple response patterns, each of which may be shared by several compounds49. These patterns were identified with probabilistic modelling that decomposes the chemical-induced transcriptional variation into components of interrelated activity. Biological prior knowledge was brought into the analysis while also reducing the data dimensionality with Gene Set Enrichment Analysis (GSEA)50. GSEA was computed (Java software version 2–2.05, http://www.broadinstitute.org/gsea) using 1321 distinct C2-curated genes sets25 from the Molecular Signature Database (http://www.broadinstitute.org/gsea/msigdb). The false discovery rate q value (FDRq), which GSEA produces to represent the significance of the enrichment, was converted into a conservative integer values with the transformation min((−log(FDRq)), 1, 0), separately for the positively and negatively activated genes in the gene sets, resulting in activation counts for 3062 instances over 2442 gene sets. The latent Dirichlet allocation (LDA) model was then used to identify transcriptional response patterns from the gene set activation count data. Each resulting component associates probabilistically a subset of the treatments with a subset of the gene sets. Each component thus represents a specific chemical-induced response pattern, interpretable based on the associated gene sets. To select the number of components, an external validation set describing the functional similarity of the drugs based on their known protein targets and ATC (Anatomical Therapeutic Chemical, http://www.whocc.no/atc_ddd_index/) codes was used48. Drug target information was obtained from ChEMBL (http://www.ebi.ac.uk/chembl/), DrugBank (http://www.drugbank.ca/), DUD (http://dud.docking.org/) and ZINC (http://zinc.docking.org/). In addition targets and ATC codes for the CMap compounds were extracted from publicly available sources26. Drugs sharing four-level ATC codes were treated as functionally similar for the purposes of this analysis. In total, 4427 associations between 821 CMap compounds and 796 targets
or ATC codes were used. The component count which maximized the performance in retrieving (that is, predicting) drugs sharing these annotations was chosen from the set of 20, 50, 100, 150 and 200. The posterior distribution of the model parameters was computed with collapsed Gibbs sampling. For the hyperparameters controlling the sparsity of the model, gamma hyperpriors were applied with fixed parameters and their posterior was estimated with Metropolis sampling.

**Toxicological profiles from NCI-60.** Toxicological profile data were downloaded from the NCI-60 DTP human tumour cell line screen web site (http://dtp.ncl.nih.gov/docs/cancer/cancer_data.html)\(^{30}\). The data set has three reported drug response values: GI50 (50% growth inhibition), TGI and LC50 (50% lethality concentration) for 59 different cell lines. These values have been inferred from measurements covering typically five concentration values, most common range being from 10 nM to 100 μM (or from ~8 – 4 log10 M). The NCI-60 and CMap instances were matched based on the compound names. In addition, alvespumpecin and tanespymycin, named 17-DMAG and 17-AAG in NCI-60, respectively, were added manually. The three drug response values were extracted from NCI-60 data for in total 222 CMap compounds and 492 cross-over measurement instances on the three cell lines (MCF7, PC3 and HL60), with 197, 179 and 116 instances per cell line, respectively, averaging over multiple measurements when available. The resulting NCI-60 data are provided in Supplementary Data 1.

**Concentration-dependent cytotoxicity.** Concentration-dependent cytotoxicity was defined as the difference of the logarithmic CMap concentration and GI50 values, that is, log10(CMap concentration) – log10(GI50). Cellular growth inhibition above the GI50-level was used as a cut-off to classify the 492 cross-over instances as either cytotoxic (n = 121) or non-cytotoxic (n = 371), as shown in Fig. 2.

**Defining the Predictive Toxicogenomics Space (PTGS).** As the CMap generally includes one concentration assessment (10 μM), dose-dependent cytotoxicity was modelled across compounds under the a-priori assumption that compound-induced transcriptomic responses are subject to the compounds’ intrinsic potency to cause cytotoxicity (for additional details see Supplementary Fig. 2). The 100 components produced by the probabilistic model covered the full space of transcriptional responses caused by the 3062 CMap measurement instances. Associations of the components to cytotoxicity were sought by evaluating their ability to predict the concentration-dependent cytotoxicity for the cross-over instances. The concentration-dependent cytotoxicity values have the highest density around GI50, making the data set ideal for predicting relatively low concentrations of cytotoxicity. Thus, a classification model was trained to identify whether an instance had been measured above the GI50-level. The 100 LDA-components were first ranked based on their probability-weighted mean concentration-dependent cytotoxicity values over the 492 training instances. The mean cytotoxicity values were computed as

\[
\pi_{\text{TOX}} = \sum_i P(i | z) \cdot \pi_{z | x},
\]

where \(\pi_{z | x} \) is the concentration-dependent cytotoxicity in relation to GI50 and where the normalized probabilities \(P(i | z)\) for the training instances \(i \) to belong to component \(z\) were computed as

\[
P(i | z) = \frac{P(z | i)}{\sum_{k} P(z | k)}.
\]

Starting with components with the highest associated cytotoxicity and using the sum of the component probabilities to calculate a predictive score, the 100 components were progressively included in the model. The cumulative concentration-dependent cytotoxicity classification performance over the test instances was evaluated, providing area under the ROC curve values (AUC) for each component count (Fig. 2B). A perfect model would have an AUC-ROC of 1 and an AUC-ROC of 0.5 indicates a random classifier. To focus on the components with the highest relevance to cytotoxicity, the number of components was chosen where the AUC value reached 95% of the highest value, resulting in a trade-off between interpretability and the highest predictive performance. Cytotoxicity-predictive performance is expected to decline with a large number of components, as non-relevant components are included, see Fig. 2a,c. The resulting top 14 components and their posterior was estimated with Metropolis sampling.

**Characterization of the PTGS by enrichment analyses.** Biological interpretations of the PTGS were enabled by the enrichment analysis of the component gene sets using Ingenuity Pathway Analysis (IPA, application version 220127, content version 16542223; build: 430520M and 31813880 content version for networks) and Gene Ontology (GO) enrichment analysis (R package topGO, version 16542223; build: 430520M and 31813880 content version for networks) and Gene Ontology (GO) enrichment analysis (R package topGO, version 2.12.0)\(^{35}\). The results were visualized with eye diagrams\(^{36}\). For visual interpretation, the results were thresholded at \( P\text{-value} \leq 0.01 \) and at least three genes were required to be annotated to each GO category, IPA toxList or IPA regulator. IPA upstream regulator analysis results were further filtered to include only regulators that were enriched both in the overall gene set of 199 genes (PTGS Core) as well as in any of the components, and additionally connected to other regulators via a mechanistic network\(^{35}\) to give further evidence of a genuine regulatory relationship. Furthermore, since the core set did not cover all biological functions, highly overrepresented regulators (\( P\text{-value} < 10^{-7} \)) that occurred in at least one third of the 14 components were added to the eye diagram. To compare upstream regulator analysis results with the ToxCast assay information\(^{37}\), information on genes associated with the ToxCast assays was obtained from the ToxCastDB (http://actor.epa.gov/actor/faces/ToxCastDB/GenesAssocAssays.jsp) and matched with Ingenuity upstream regulators on the basis of the gene symbol. Association of cytotoxicity to biological complexity was calculated for each type of analysis by

\[
\text{analysis}_{\text{TOX}} = \sum (\pi_{z | x} \times z_{\text{BC}}),
\]

where \(z_{\text{BC}}\) describes the biological complexity of a components, computed as

\[
\text{node}(z_{\text{node}})/(n_{\text{total}}(z_{\text{node}}) - z_{\text{node}}),
\]

where \(z_{\text{node}}\) denotes each individual component and \(Z\) all components and \(n_{\text{total}}\) is the number of statistically significant results produced by an analysis, as detailed in Supplementary Data 3.

**Hierarchical clustering and principal components analyses.** Grouping of the components was assessed by hierarchical clustering of CMap instances mapped to the PTGS components with ggdendro (v. 0.1–20), using default settings.

**Cell culture and in vitro cytotoxicity predictions.** To validate the predictive performance of the PTGS, a set of CMap instances that were not included in the NCI-60 data set were assessed for cytotoxicity (Fig. 4a). CMap cell lines MCF7 (ATCC CRL-11051), PC3 (ATCC CRL-1435) and HL60 were ordered from American Type Culture Collection (LCM Promocem AB) and maintained at 37°C with 5% CO2 in a humidified incubator according to

\[
\frac{Z}{X}\text{ or }\frac{Y}{X}\text{.}
\]
provider’s instructions. As described previously, the cell lines were grown in larger volume to make assay ready cells, tested for mycoplasma using PCR-based test kit and frozen in several ampules. Before screening, the cell number was titrated to volume to make assay ready cells, tested for mycoplasma using PCR-based test kit provider’s instructions. As described previously, the cell lines were grown in larger

Component selection for analysis of liver pathology. To identify and study the components most central to liver toxicity and to demonstrate the applicability of the PTGS component-based method to assess risk of agent-induced (for example, chemical compounds, drugs) liver toxicity, predictive modelling was undertaken. To begin, PTGS model-derived components were computed as in the “Defining the Predictive Toxicogenomics Space (PTGS)” section: Broad Institute GSEA tool was run on differential expression (fold-change based) results using the R/Bioconductor limma version 3.20.9 and, as for the CMap data, the output was quantized. To update older symbols, the gene symbols were mapped to Ensembl gene identifiers for human and rat, using the multi-symbol checker tool (http://www.genenames.org/cgi-bin/symbol_checker). Based on the estimated component distributions, the individual component predictive value of the PTGS scores were computed and used for toxicity prediction. For an example, see code at Zenodo (DOI:10.5281/zenodo.570115).

Subsequently, in order to study which individual components are predictive of liver pathologies, 24 elastic net regularized regression models, one for each finding (19) and for each severity grade (5), were fitted with the 14 component probabilities as input (X-variables) and the dichotomized pathologic findings as the output (y-variable); and trained using repeated (10 times) three-fold cross-validations. The weighted scoring of pathological findings was employed. The findings were then dichotomized using a score at least 3 for present and minimal grade (1), at least 4 for the other three and at least 5 for the severe grade (5). For robust results, only findings with more than 15 positive instances were included in the analysis. Receiver operator curves (ROC) were computed for each model using the R pROC version 1.7.2 (ref. 61). Significance for the AUCs for the classifiers was estimated using two-tailed univariate Wilcoxon rank-sum statistics in R between the effected and non-effected groups. For robust results, only findings with more than 15 positive instances were included in the analysis.

The endpoints that were assessed for histopathological changes, while 3 of those were profiled with arrays. All findings were processed and later selected for analysis based on sample number. The type of pathological change (for example, fibrosis) and its severity grade were combined, and are here defined as endpoints. Pathology endpoints were assessed cumulatively in the order given, and the pathological change was classified as toxic67. This difference is justified because the concentration-
carried out using the lasso penalized score test, termed lassocscore, employing the lambda values derived earlier using cross-validation27,28. Nested multiple testing procedures from the R package structSSI were used29, employing the adaptive Group Benjamini-Hochberg Procedure with the ‘tst’ (two step test) method and model identity, that is, the pathological finding endpoint/endgrade as the nesting variable or group index. A q < 0.05 for both component and over-all model significance was used as a dual threshold. Based on these analyses, components were selected for scoring hepatic injury (Supplementary Data 3 and 13).

Analysis of liver pathology using the gene-based method. To test the predictive ability of the component-based and the gene-based methods the rat and mouse data sets were analysed using either all components or the ones which were selected earlier as being the most liver pathology predictive (that is, G, H, I and N). Gene set activities and P-values were computed with the ROAST method using 9999 rotations and the Benjamini-Hochberg procedure as the component summary statistic. In addition to the full PTGS, drug-induced liver injury (DILI) predictive scores were defined as:

Component-based DILI score = sup(probcg, probhs, probc, probhs)

Gene-based DILI score = max(0.025, 0.005, 0.05, 0.025)

Gene-based DILI P-value = min(Pc, Ph, Pn, Pn)

Where the prob-prefix refers to the component probability i.e. P(1). As per the ROAST function, the %act is the percentage of genes which are at least marginally differentially expressed at |z| > sqrt(2) where z as denotes a z-score according to limma analysis and P.

To evaluate the gene-based scores with AUC analysis, the proportion of active genes was used for scoring. The findings were dichotomized using a score at least three for present and minimal grades and at least two for the other three; only findings with at least 10 positive instances were included in the analysis (range for n: 16–463 for gene-based and 11–444 for component-based). Significance for the AUCs was computed using two-tailed univariate Wilcoxon rank-sum statistics in R between the effected and non-effected groups and multiple testing corrected using the Benjamini-Hochberg Procedure with the ‘tst’ (two step test) method.

To evaluate the performance of the PTGS and DILI scores, a threshold for predicting DILI was chosen. To establish a threshold for the PTGS score, findings with at least 10 positive instances were included in the analysis (range for n: 463 for the ‘present’ endpoint), and the 50% level (about two-fold enrichment of findings) was used as the decision threshold (score > 0.3) in parallel with the significance level of q < 0.05.

Predicting human drug-induced liver injury. Human and rat hepatocyte data from the Open TG-GATEs database was analysed in combination with Cmax values (maximal total blood concentration) from literature to predict clinical exposure levels of the chemotherapeutic compounds (Supplementary Data 14). Withdrawn drugs and other labelling associated with drug-induced liver injury concern were also obtained from the Liver Toxicity Knowledge Base33. Assay withdrawals and other labelling associated with drug-induced liver injury were predicted as DILI positive. Comparative and combinatorial analyses permitted the analysis of 11 negative controls and 54 compounds annotated as DILI positive in rat hepatocytes when PTGS becomes active)–log10(human blood therapeutic concentration). To derive a threshold for predicting DILI, negative control findings with at least 10 positive instances were included in the analysis (range for n: 1689 overall and n = 463 for the ‘present’ endpoint), and the 50% level (about two-fold enrichment of findings) was used as the decision threshold (score > 0.3) in parallel with the significance level of q < 0.05.

Various R packages were used for data pre-processing and transformations68, tidygr (v. 0.4.1), stats:reshape (R 2.15.3–3.2.3), plyr (v. 1.8.4), dplyr (v. 0.4.3), magrittr (v. 1.5), reshape (v. 0.8.5) and reshape2 (v. 1.4.1). Statistics analysis utilized R base functions, stats:stats4 (R 2.15.3–3.2.3), MASS (v. 7.3–45), and (v. 1.3), structSSI (v. 1.1.1)29, vcov (v. 1.4–1), glmnet (v. 2.0–5)3, q value (v. 2.2.2), lassocscore (v. 0.6)7, caret (v. 6.0–7.0), ISLR (v. 1.0) made (v. 1.40)37 and plot (v. 3.1.3). A framework (v. 1.4.3) and BioPacd (v. 1.4.3) libraries, as well as the pander (v. 1.5.3), knitr (v. 1.2), ggdendro (v. 0.4.1), RCoolBrewer (v. 1.1–2), ggplot2 (v. 2.0.0–1), ggrepel (v. 0.5), grid (R 2.15.3–3.2.3) and gridExtra (v. 2.2.1). EyelidDiagrams were produced with the custom software (2011–2012)38. BioConductor72 packages were utilized for bioinformatics analyses: BioCGenerics (v. 2.3.0), BioGenerics (v. 0.16.1), aroma.affymetrix (v. 1.2.0)46, limma (v. 3.26.9)52, simpleaffy (v. 2.46.0), affy (v. 1.48.0), topGO (v. 2.12.0)44 and GO.db (v. 2.9.0). Microsoft Excel (various versions) was used for browsing and editing of tables.

Code availability. Code for R/Bioconductor47,51,52 packages is available at http://bioconductor.org. Custom R code and methods to calculate component-based PTGS scores is archived via the CERN Openaire online service Zenodo (DOI: 10.5281/zenodo.570115).

Data availability. Freely available data were used in the project throughout. Data sources included the Connectivity Map (CMap)25, NCI-60 DTP human tumour cell line screen database50, the Molecular Signatures Database (MSigDB)30, the Open TG-GATEs toxicogenomics database46, the Liver Toxicity Knowledge Base53, and Cmax and DILI potential-related information extracted from various sources as Supplementary Data 16. Validation data generated in the study is available at Figshare (DOI: 10.6084/m9.figshare.4954583). All other data listed on this page are available on reasonable request.

References

1. Collins, F. S., Gray, G. M. & Bucher, J. R. Toxicology. Transforming environmental health protection. Science 319, 906–907 (2008).
2. Hamburg, M. A. Advancing regulatory science. Science 331, 987 (2011).
3. Willyard, C. Foretelling toxicity: FDA researchers work to predict risk of liver injury from drugs. Nat. Med. 22, 450–451 (2016).
4. Olson, H. et al. Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul. Toxicol. Pharmacol. 32, 56–67 (2000).
5. Hussani, S. H. & Farrington, E. A. Idiosyncratic drug-induced liver injury: an update on the 2007 overview. Expert Opin. Drug Saf. 13, 67–81 (2014).
6. Atienza, F. A. et al. Key challenges and opportunities associated with the use of in vitro models to detect human DILI: integrated risk assessment and mitigation plans. BioMed Res. Int. 2016, 9737920 (2016).
7. Bai, J. P. & Abernethy, D. R. Systems pharmacology to predict drug toxicity: integration across levels of biological organization. Annu. Rev. Pharmacol. Toxicol. 53, 451–473 (2013).
8. Strurla, S. J. et al. Systems toxicology: from basic research to risk assessment. Chem. Res. Toxicol. 27, 314–329 (2014).
9. Krewski, D. et al. A framework for the next generation of risk science. Environ. Health Perspect. 122, 796–805 (2014).
10. Andersen, M. E., McMullen, P. D. & Krewski, D. Developing tools for defining and establishing pathways of toxicity. Arch. Toxicol. 89, 809–812 (2015).
11. Craftrion, R. C. & Toward the replacement of animal experiments through the bioinformatics-driven analysis of ‘omics’ data from human cell cultures. Altern. Lab Anim. 43, 325–332 (2015).
12. Natsoulis, G. et al. The liver pharmacological and xenobiotic gene response repertoire. Mol. Syst. Biol. 4, 175 (2008).
13. Chen, M., Zhang, M., Borlak, J. & Tong, W. A decade of toxicogenomic research and its contribution to toxicological science. Toxicol. Sci. 130, 217–228 (2012).
14. Gusenleitner, D. et al. Genomic models of short-term exposure accurately predict long-term chemical carcinogenicity and identify putative mechanisms of action. PLoS ONE 9, e102579 (2014).
15. Kirshenbaum, L. M. et al. Cancer biology, toxicology and alternative methods development go hand-in-hand. Basic Clin. Pharmacol. Toxicol. 115, 50–58 (2014).
16. Igarashi, Y. et al. Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Res. 45, D921–D927 (2015).
17. Holsapple, M. P. & Wallace, K. D. Response dose considerations in risk assessment—an overview of recent ILSI activities. Toxicol. Lett. 180, 85–92 (2008).
18. Thomas, R. S. et al. Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. Toxicol. Sci. 136, 4–18 (2013).
19. Lynchey, G. et al. Lessons from toxicology: developing a 21st-century paradigm for medical research. Environ Health Perspect. 123, A268–A272 (2015).
20. Vinken, M. Adverse outcome pathways and drug-induced liver injury testing. Chem. Res. Toxicol. 28, 1391–1397 (2015).

Statistical and bioinformatics analyses. Nonparametric statistics were extensively used for between-group comparisons28, and for sample numbers below 5 variance-adjusted parametric tests were used52. Analyses were performed using the R statistical programming language, v. 2.15.3–3.2.3 (http://www.r-project.org/).
21. Kleinstreuer, N. C. et al. Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. Nat. Biotechnol. 32, 583–591 (2014).
22. Dayton, G. et al. SEURAT: safety evaluation ultimately replacing animal testing—recommendations for future research in the field of predictive toxicology. Arch. Toxicol. 89, 15–23 (2015).
23. Huang, R. et al. Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. Nat. Commun. 7, 10425 (2016).
24. ECHA Read-Across Assessment Framework (RAAF). European Chemicals Agency; Helsinki, Finland. http://echa.europa.eu/documents/10162/13628/raaf_en.pdf (2017).
25. Lamb, J. et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313, 1929–1935 (2006).
26. Iorio, F. et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc. Natl Acad. Sci. USA 107, 14621–14626 (2010).
27. Iskander, M. et al. Characterization of drug-induced transcriptional modules towards drug repositioning and functional understanding. Mol. Syst. Biol. 9, 662 (2013).
28. Parkkinen, J. A. & Kaski, S. Probabilistic drug connectivity mapping. BMC Bioinf. 15, 113 (2014).
29. De Abreu, K. N. et al. Grouping 34 chemicals based on mode of action using connectivity mapping. Toxicol. Sci. 151, 447–461 (2016).
30. Shoemaker, R. H. The NCI60 human tumour cell line anticancer drug screen. Nat. Rev. Cancer 6, 813–823 (2006).
31. Muller, P. Y. & Milton, M. N. The determination and interpretation of the therapeutic index in drug development. Nat. Rev. Drug Discov. 11, 751–761 (2012).
32. Shah, F. et al. Setting clinical exposure levels of concern for Drug-Induced Liver Injury (DILI) using mechanistic in vitro assays. Toxicol. Sci. 147, 500–514 (2015).
33. Chen, M. et al. The liver toxicity knowledge base: a systems approach to a complex end point. Clin. Pharmacol. Ther. 93, 409–412 (2013).
34. Bell, S. M., Angrish, M. M., Wood, C. E. & Edwards, S. W. Integrating publicly available data to generate computationally predicted adverse outcome pathways. Nat. Biotechnol. 32, 1600–1607 (2014).
35. Caldas, J., Gehlenborg, N., Faisal, A., Brazma, A. & Kaski, S. Probabilistic rule-based visualization of biologically relevant microarray experiments. Bioinformatics 25, 1145–1153 (2009).
36. Culhane, A. C., Thioulouse, J., Perriére, G. & Higgins, D. G. MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics 21, 2789–2790 (2005).
37. Ilin, K. et al. High throughput cell-based screening of 4,910 known drugs and drug-like small molecules identifies disulfiram as an inhibitor of prostate cancer cell growth. Clin Cancer Res. 15, 6070–6078 (2009).
38. Gautam, P. et al. Identification of selective cytotoxic and synthetic lethal drug responses in triple negative breast cancer cells. Mol. Cancer 15, 34 (2016).
39. Hubley, S. B., Cummings, S. R., Browner, W. S., Grady, D. & Newman, T. B. Designing Clinical Research: an Epidemiologic Approach 4th ed. (Lippincott Williams & Wilkins, 2013) Appendix 6C.
40. Robin, X. et al. pROC: an open-source package for R and S-+ to analyze and compare ROC curves. BMC Bioinf. 12, 77 (2011).
41. Willighagen, E. L., Wehrens, R. & Buydens, L. M. C. Molecular chemometrics. Mol. Syst. Biol. 8, 447–461 (2012).
42. Wehrens, R. in Chemometrics with R: Multivariate Data Analysis in the Natural Sciences and Life Sciences (Springer, 2011).
43. Todeschini, R. & Consonni, V. in Methods and Principles in Medicinal Chemistry (eds. Mannhold, R., Kubinyi, H. & Timmermann, H.) 11 (Wiley-VCH, 2000).
44. Willighagen, E. L. et al. The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching. J. Cheminform. 9, 33 (2017).
45. Lee, A. C., Shelden, K., Rosania, G. R. & Crippen, G. M. Data mining the NCI60 to predict generalized cytotoxicity. J. Chem. Inf. Model. 48, 1379–1388 (2008).
46. Dixon, J., Grier, S., Cleverd, D. & Obermayer, K. A new summarization method for afymmetrix probe level data. Bioinformatics 22, 943–949 (2006).
47. Wickham, H. Tidy data. J. Stat. Softw. 59, 1–23 (2014).
48. Sankaran, K. & Holmes, S. structureSS: Simultaneous and selective inference for grouped or hierarchically structured data. J. Stat. Softw. 59, 1–21 (2014).
49. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22 (2010).
50. Voorman, A., Shojaie, A. & Witten, D. Inference in high dimensions with the penalized score test. Preprint at https://arxiv.org/abs/1401.2678 (2014).
51. Taylor, J. & Tibshirani, R. J. Statistical learning and selective inference. Proc. Natl Acad. Sci. USA 112, 7629–7634 (2015).
52. Aleo, M. D. et al. Human drug-induced liver injury severity is highly associated with dual inhibition of liver mitochondrial function and bile salt export pump. Hepatology 60, 1015–1022 (2014).
53. Gustafsson, F., Foster, A. J., Sarda, S., Bridgland-Taylor, M. H. & Kenna, J. G. A correlation between the in vitro drug toxicity of drug cell lines that express human P450s and their propensity to cause liver injury in humans. Toxicol. Sci. 137, 189–211 (2014).
54. Greene, N. et al. Developing structure-activity relationships for the prediction of hepatotoxicity. Chem. Res. Toxicol. 23, 1215–1222 (2010).
55. Regev, A. Drug-induced liver injury and drug development: industry perspective. Semin. Liver Dis. 34, 227–239 (2014).
56. Rochon, J., Gondan, M. & Kieser, M. To test or not to test: Preliminary assessment of normality when comparing two independent samples. BMC Med. Res. Methodol. 1, 81 (2012).
57. Wickham, H. ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag, 2016).

Acknowledgements
We want to thank Ida Lundenschmidt and the High Throughput Biomedicine unit at FIMM for technical support to cellular high-throughput screening assays. I.A.P. and S.K. acknowledge support from FP7-Theme HEALTH-2010-Alternative-Tasting through SEURAT/ToxBank and Cosmetics Europe under Grant Agreement no. 267042 EU H2020 projects NMMP30-2015 calIBRARE (no 682239) and NanoReg2 (no. 646221), the Swedish Research Council, Swedish Vinnova/EUROSTARS E9698-ToxHQ CRO, Swedish Cancer and Allergy Fund, the Swedish Fund for Research without Animal...
Experiments, Finnish Foundation’s Post-doc research grant award to P.K., and Karolinska Institutet. K.W. acknowledges support from the Jane and Aatos Erkko Foundation.

**Author contributions**
J.A.P., R.C.G., S.K., K.W., P.K. designed the study. J.A.P. and K.W. processed the NCI-60 data. J.A.P. and S.K. designed the probabilistic models. K.W. performed the in vitro measurements. J.A.P., P.K., E.L.W. and R.C. performed the data analysis. P.K., R.C. and R.C.G. analysed the toxicological results. P.K. and R.C.G. designed and P.K. carried out the DILI prediction work. All authors contributed to writing.

**Additional information**
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

**Competing interests:** J. Parkinen, P. Kohonen, S. Kaski, R.C. Grafström declare ‘personal financial interest’ for being equal contributors to a patent application. All other authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

---

How to cite this article: Kohonen, P. et al. A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury. *Nat. Commun.* **8**, 15932 doi: 10.1038/ncomms15932 (2017).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2017